NOTE IMPORTANTE : Veuillez prendre note que depuis la publication de cette version du livre blanc, COVI n'utilise plus la technologie GPS. Plus de détails à venir dans une prochaine version.

Livre blanc de COVI – Version 1.0

Hannah Alsdurf[1], Yoshua Bengio[2;3], Tristan Deleu[2;3], Prateek Gupta[2;4;5], Daphne Ippolito[6], Richard Janda[7], Max Jarvie[8], Tyler Kolody[7], Sekoul Krastev[9], Tegan Maharaj[2;3], Robert Obryk, Dan Pilat[9], Valérie Pisano[2], Benjamin Prud'homme[2], Meng Qu[2;10], Nasim Rahaman[2;11], Irina Rish[2;3], Jean-François Rousseau[12], Abhinav Sharma[7], Brooke Struck[9], Jian Tang[2,10], Martin Weiss[2;3], Yun William Yu[13]

Résumé

La pandémie de SARS-CoV-2 (COVID-19) a entraîné une forte pression sur les soins de santé et les services de santé publique du monde entier. La fonction de traçage des contacts est un outil essentiel pour les responsables de la santé publique et les membres de communautés locales afin de changer le cours de la pandémie de COVID-19. Bien que le traçage manuel standard des personnes exposées à la COVID-19 soit pour l'instant la norme en vigueur, il présente des défis importants qui limitent la capacité des autorités en santé publique à minimiser les infections au niveau communautaire. Le traçage de contacts personnalisé pair à pair qui utilise une application mobile a le potentiel de modifier le paradigme concernant la propagation de la COVID-19 dans les communautés. Bien que certains pays aient déployé des systèmes de traçage centralisés par GPS ou Bluetooth, des systèmes décentralisés protégeant davantage la vie privée offrent à peu près les mêmes avantages sans la concentration des données entre les mains des gouvernements ou de sociétés à but lucratif.

De plus, des méthodes d'apprentissage automatique peuvent être utilisées pour contourner certaines des limites du traçage numérique standard en incorporant de nombreux indicateurs (y compris les conditions médicales, l'autodéclaration des symptômes et de nombreuses rencontres avec des personnes à différents niveaux de risque, pour des durées et des distances différentes) et leur incertitude en une estimation plus graduée et plus précise du risque d'infection et de contagion. Le risque estimé peut être utilisé pour fournir une prédiction des risques, des recommandations personnalisées et des informations pertinentes à l'utilisateur et le mettre en contact avec les services de santé. Enfin, les données non identificatoires concernant ces risques peuvent alimenter des modèles épidémiologiques détaillés formés conjointement avec l'apprentissage automatique prédictif, et ces modèles peuvent capter statistiquement les interactions et l'importance des différents facteurs impliqués dans la transmission de la maladie. Ils peuvent également être utilisés pour suivre, évaluer et optimiser les différents politiques de santé et scénarios de confinement/déconfinement en fonction d'indicateurs médicaux et de productivité économiques.

Toutefois, une telle stratégie basée sur des applications mobiles et l'apprentissage automatique devrait atténuer de manière proactive les risques en matière d'éthique et de protection de la vie privée qui pourraient avoir des répercussions importantes sur la société (non seulement les impacts sur la santé, mais aussi les impacts tels que la stigmatisation et l'abus d'utilisation de données personnelles). Nous présentons ici un aperçu de la raison d'être, de la conception, des considérations d'ordre éthique et de la stratégie de protection de la vie privée de COVI, une application mobile publique de dépistage pair à pair de la COVID-19 et de l'évaluation des risques développée au Canada.

[1] Université d'Ottawa, [2]Mila, [3]Université de Montréal, [4]The Alan Turing Institute, [5]University of Oxford, [6]University of Pennsylvania, [7]Université McGill, [8]Borden Ladner Gervais LLP, [9]The Decision Lab, [10]HEC Montréal, [11]Institut Max Planck, [12]Libéo, [13]University of Toronto.

Auteur général correspondant : richard.janda@mcgill.ca
Auteur correspondant pour la santé publique : abhinav.sharma@mcgill.ca
Auteur correspondant pour la protection de la vie privée : ywyu@math.toronto.edu
Auteur correspondant pour l'apprentissage automatique : yoshua.bengio@mila.quebec
Auteur correspondant pour le point de vue de l'utilisateur : brooke@thedecisionlab.com
Auteur correspondant pour la mise en œuvre technique : jean-francois.rousseau@libeo.com

Table des matières

1 Aperçu

1.1 Introduction

L'épidémie de SARS-CoV-2 (COVID-19) est la pandémie majeure depuis un siècle et est actuellement la plus grande crise mondiale depuis la Seconde Guerre mondiale [1]. La pandémie a créé une urgence sanitaire mondiale qui a eu un impact considérable sur tous les aspects de la vie moderne et a mis à rude épreuve les systèmes de santé, les économies et les institutions politiques [2, 3, 4]. L'évolution rapide de notre compréhension de la COVID-19 a remis en question les décisions concernant les stratégies utilisées pour contenir et prévenir la propagation du virus [5, 6].

Le traçage manuel des contacts, qui tente d'identifier et d'isoler les personnes à haut risque de contracter la COVID-19, est la principale stratégie utilisée par les services de santé publique pour dépister de manière sélective le virus et ainsi réduire sa propagation [7, 8]. Au cours des derniers mois, des solutions de traçage automatique des contacts ont été proposées pour relever plusieurs des défis posés par la méthode manuelle [9, 10, 11, 12, 13, 14, 15]. Elles visent à réduire le coût et la charge de travail des professionnels de la santé, le biais de rappel des contacts antérieurs, l'incapacité à identifier les personnes rencontrées faisant partie du grand public (par exemple un employé de supermarché) et le délai entre l'identification d'une personne ayant été infectée et le suivi manuel pour alerter les personnes à haut risque de contagion. De plus, alors que la méthode manuelle de traçage peut être très efficace aux premiers stades d'une épidémie potentielle, son efficacité diminue à mesure que l'infection se répand et que le nombre de cas dépistés nécessaires augmente [16, 8]. Enfin, elle exige que les personnes se sentent à l'aise de divulguer des informations potentiellement sensibles et identificatoires au sujet de leur cercle social à des représentants du gouvernement [17].

Le respect de la vie privée est une préoccupation importante, tant pour la méthode de traçage manuelle, que pour les solutions basées sur des données [18, 19, 20, 11]. La fonction de traçage des contacts, telle qu'elle est généralement conçue, exige que les personnes fournissent un certain nombre d'informations personnelles aux autorités publiques [21]. Avec la méthode de traçage manuelle, des patients diagnostiqués essaient de se rappeler toutes leurs rencontres physiques des deux dernières semaines et de les signaler à un responsable de la santé publique (avec leur nom et numéro de téléphone), responsable qui tente ensuite de contacter chaque personne figurant sur la liste et de leur demander d'entreprendre un exercice de rappel similaire. Le traçage automatique des contacts, qui s'appuie sur la technologie Bluetooth ou sur l'utilisation du GPS des téléphones des utilisateurs pour détecter quand ils ont été en contact les uns avec les autres, peut également être mis en œuvre de manière centralisée. La divulgation d'informations personnelles aux autorités publiques prend la forme d'une base de données centralisée qui repère les contacts parmi des individus identifiables [14, 10].

Bien que les solutions technologiques puissent amplifier l'impact du traçage automatique des contacts, si elles ne sont pas mises en œuvre correctement, elles peuvent également présenter des risques importants pour les citoyens [22, 20], notamment la perte de libertés civiles, l'érosion de la vie privée et la surveillance par le gouvernement, des entreprises ou des particuliers [23]. La différence entre un système de traçage numérique centralisé et décentralisé a récemment fait l'objet d'un débat animé, allant de la question de savoir si cette distinction est vraiment significative [24] à une critique ciblée des implications des systèmes centralisés sur la vie privée [25, 26]. Ces préoccupations peuvent limiter considérablement l'efficacité de ces applications, car les citoyens qui ne font pas confiance à une application seraient peu susceptibles de l'utiliser ou de l'utiliser correctement. Les sociétés démocratiques doivent donc adopter avec l'utilisation des technologies numériques pour le traçage des contacts une approche qui protège la vie privée afin de renforcer la confiance du public dans ces applications [22, 20, 27]. La nécessité d'un traçage des contacts décentralisé a suscité une forte augmentation des propositions de stratégies de traçage automatique des contacts préservant la vie privée [13,

11, 12, 15, 28]. Les approches décentralisées notifient automatiquement les personnes de contacts récents à haut risque sans confier aucune information pouvant identifier la personne à une autorité gouvernementale centralisée.

Au-delà de la fonction la plus élémentaire du traçage des contacts, qui consiste à informer les personnes lorsqu'elles ont été en contact avec une personne infectée, une application de santé publique efficace peut offrir un large éventail de possibilités pour aider ses utilisateurs à prendre des décisions concrètes [29] et pour contribuer à l'élaboration de meilleurs modèles épidémiologiques et de meilleures politiques publiques [30].

Alors que la plupart des propositions de traçage automatique des contacts ne peuvent fournir qu'une notification binaire (c'est-à-dire qu'une personne a été exposée à quelqu'un dont le test de dépistage de la COVID-19 était positif ou non), une gradation des risques à plusieurs niveaux plus réaliste, où le risque est basé sur des facteurs supplémentaires tels que les symptômes, les comorbidités, et la profession, permet des suggestions nuancées et une meilleure compréhension du risque individuel et collectif. Alors que ces informations sensibles supplémentaires doivent par défaut être conservées en toute sécurité sur le téléphone de chaque utilisateur, en donnant aux utilisateurs de l'application la possibilité de partager volontairement leurs données non identificatoires pour la conception des modèles épidémiologiques et d'évaluation des risques, cela leur donne la possibilité de contribuer de manière significative à la lutte contre la COVID-19 et à la qualité du prédicteur utilisé dans leur téléphone.

Dans ce livre blanc, nous décrivons la conception de COVI, une application mobile développée pour le Canada, axée sur la protection de la vie privée et basée sur la détection de proximité par Bluetooth. L'application COVI vise à atteindre les objectifs suivants : 1) réduire et contrôler le nombre d'infections en responsabilisant les citoyens, au sens le plus large du terme, à se protéger et protéger les autres en suivant des recommandations ciblées en fonction de leur risque d'infection ; 2) extraire des informations cruciales pour alimenter et orienter une approche de la politique de santé publique fondée sur les données concernant le confinement durant la pandémie et la planification de déconfinement ; et 3) établir de solides méthodes de protection de la vie privée en utilisant une stratégie décentralisée pour tenir les informations personnelles des utilisateurs à l'écart des autres personnes, des entreprises et des organisations gouvernementales. Ce livre blanc présente la philosophie et l'approche générales de COVI et démontre comment, en combinant la recherche en santé publique, en épidémiologie, en vie privée, en apprentissage automatique, en éthique et en psychologie, COVI vise à atténuer les risques tout en permettant aux citoyens de baser leur réaction personnelle à la crise de la COVID-19 sur des décisions fondées sur des données probantes.

1.2 Principaux objectifs de COVI

1.2.1 Réduire la propagation de la COVID-19

L'objectif principal de COVI est de réduire la propagation de la COVID-19, afin de : 1) de réduire le taux de morbidité et la mortalité associée à l'infection par la COVID-19 ; et 2) réduire la charge de la COVID-19 sur les systèmes de santé. Pour atteindre cet objectif, la principale stratégie de COVI est d'informer les individus de leur risque d'infection afin qu'ils puissent agir de manière responsable pour se protéger et protéger les autres et limiter la propagation du virus.

COVI s'appuie sur des probabilités de niveaux de risque (comme détaillé ci-dessous), plutôt que sur les résultats binaires d'une infection confirmée (ou non), pour attribuer aux utilisateurs de l'application des niveaux de risque d'infection par la COVID-19. Cette probabilité de risque est modifiée et mise à jour en fonction de deux catégories d'information principales : le profil individuel des utilisateurs et leur profil d'interaction. Les profils individuels sont établis à partir des informations saisies par l'utilisateur, notamment les données démographiques, les comorbidités avec la COVID-19, la profession de la personne et la présence de nouveaux symptômes. Le profil

d'interaction est enregistré lorsque deux ou plusieurs utilisateurs sont à proximité pendant un certain temps. La probabilité de transmission dépend de détails tels que la distance qui les sépare, le temps écoulé lors d'une interaction, les contacts antérieurs d'un individu (en particulier avec la même personne) et l'utilisation de masques ou d'autres dispositifs de distanciation physique. La combinaison de ces informations génère une distribution de probabilité personnalisée, dérivée de l'apprentissage automatique, concernant le fait que l'utilisateur soit infecté, quand l'infection aurait eu lieu, et du potentiel de contagion sur différents jours après l'infection.

COVI permet ensuite aux citoyens d'adopter une approche progressive et proactive de la gestion du risque d'infection pour eux-mêmes et pour les autres. Le niveau de risque individuel spécifique ne sera pas affiché, car la réaction à l'obtention d'un nombre spécifique lié au risque peut être hétérogène et susciter un comportement contre-productif [31]. Au contraire, si le niveau de risque de l'utilisateur augmente, des messages fondés sur des données probantes [32] au préalablement approuvés par les autorités de santé publique [33] seront fournis par le biais de l'application. Ces messages recommanderont diverses mesures pouvant être prises par l'utilisateur pour réduire le risque d'infection par la COVID-19.

Bien que l'application puisse grandement contribuer à l'efficacité et à la précision du dépistage, elle ne peut pas (et n'a pas pour but) de remplacer un travailleur de la santé, en particulier dans les cas difficiles où un jugement professionnel est nécessaire. Au contraire, l'application COVI se veut complémentaire à la méthode de traçage manuel et représente une stratégie d'optimisation des ressources limitées des institutions de santé publique. Lorsque COVI calcule un risque élevé pour une personne, celle-ci peut être orientée vers les services de santé publique locaux pour y subir un test pour la COVID-19, suivi, avec son consentement, d'un dépistage anonyme des personnes avec qui elle est entrée en contact.

1.2.2 Contribuer à une approche fondée sur les données pour la politique de santé publique de la COVID-19

Outre le traçage des contacts et une connaissance plus précoce du risque d'infection pour un individu, basé sur le risque de propagation de la COVID-19 parmi les utilisateurs de l'application, COVI vise à mieux comprendre les facteurs contribuant à la transmission de la maladie, ce qui permet d'éclairer les politiques de santé publique sur le confinement et le déconfinement grâce à des modèles épidémiologiques adaptés aux données recueillies sur les téléphones. L'application recueillera des données démographiques de base et des données de comorbidité concernant des variables qui ont été identifiées comme étant des pronostics pour contracter la COVID-19 et développer un risque accru des effets indésirables d'une infection. Les personnes peuvent ajouter en temps réel l'apparition de symptômes ainsi que les symptômes spécifiques qu'ils ressentent. De plus, le profil des symptômes peut être mis à jour au fil du temps, créant ainsi un profil d'évolution des symptômes. Ces détails sont souvent difficiles à obtenir dans des études épidémiologiques en raison du biais de rappel [34]. Ces informations seront intégrées dans les modèles d'apprentissage automatique et les modèles épidémiologiques (voir ci-dessous) afin d'obtenir une compréhension plus pointue de la dynamique de la transmission virale. Il a été modélisé que l'infection peut commencer environ 2,5 jours avant l'apparition des symptômes avec un pic de virulence la veille de l'apparition des symptômes [35, 36]. L'excrétion virale semble plafonner pendant 5 à 7 jours, puis décliner. Bien que ces détails soient essentiels pour comprendre le profil viral général chez un individu infecté, la capacité à comprendre comment ces informations se traduisent en transmission virale réelle est limitée [37]. En faisant correspondre l'apparition des symptômes au niveau individuel, le traçage des contacts, le profil individuel, et les détails du profil d'interaction (comme décrit dans la section 1.2.1) avec les données sur la santé autodéclarées et officiellement vérifiées des diagnostics de la COVID-19, COVI fournira des informations essentielles pour améliorer notre compréhension de la transmission virale et de la dynamique de la propagation du virus au sein de la population des utilisateurs de l'application.

L'analyse de ces données peut permettre une compréhension plus précise de la progression de la pandémie, ce qui est essentiel à l'élaboration d'une politique de santé publique efficace [38]. Les modèles d'apprentissage

automatique épidémiologique formés à partir de ces données peuvent être utilisés pour évaluer et optimiser les politiques de santé publique en simulant leur effet futur sur la propagation de la maladie en fonction de la réponse des citoyens aux recommandations de santé publique, parfois directement envoyées par l'application et sélectionnées en fonction du profil de l'utilisateur et du niveau de risque de l'individu. À mesure que la pandémie progresse, des stratégies de déconfinement efficaces gagnent en importance pour aider une population à passer d'une distanciation sociale stricte à une distanciation plus indulgente. En comprenant comment les différents profils d'utilisateurs ont une influence sur le risque d'infection, des approches plus ciblées de déconfinement peuvent être proposées. Les modèles d'apprentissage automatique et épidémiologiques résultent de l'agrégation des données individuelles et anonymisées et peuvent être fournis aux agences de santé publique locales et régionales. Cela peut servir de base à des stratégies régionales spécifiques visant à réduire la distance sociale. COVI peut en outre fournir une rétroaction en temps réel et localisé de la transmission virale une fois que ces changements de politique ont eu lieu. Cela sera essentiel alors que les gouvernements explorent prudemment les stratégies de déconfinement.

1.2.3 Protéger la vie privée et maintenir la confiance du public

Malgré les avantages du traçage numérique des contacts, il suscite des inquiétudes légitimes quant au respect de la vie privée et aux libertés civiles des individus dans les sociétés démocratiques. Afin d'atteindre les objectifs énoncés aux points 1.2.1 et 1.2.2, il est primordial de protéger efficacement la vie privée afin de maintenir la confiance du public dans la technologie et d'atténuer les risques d'atteinte aux droits de la personne et à la démocratie.

COVI est conçu pour minimiser la collecte, l'utilisation et la divulgation de renseignements personnels et maximiser les possibilités pour les utilisateurs de donner leur consentement, tout en remplissant ses objectifs fondamentaux visant à sauver plus de vies tout en préservant la vie privée et la dignité. Deux raisons principales expliquent l'importance de la vie privée. Premièrement, nous considérons la vie privée comme une forme de bien public, ayant une valeur pour le bon fonctionnement des systèmes politiques démocratiques [39, 40] et de la société civile en général [40, 41]. L'adoption massive de systèmes de traçage des contacts qui n'intègrent pas suffisamment le respect de la vie privée dans la conception du système peut constituer un véritable danger pour ces institutions. Deuxièmement, les individus accordent tous une certaine importance à leur vie privée, et, étant donné que le traçage des contacts dépend fortement du nombre d'utilisateurs - il faut qu'une masse critique d'utilisateurs ait installé l'application pour qu'elle soit la plus efficace possible - la protection de la vie privée devient donc un critère déterminant pour que les utilisateurs l'adoptent [42, 19].

Il a été largement démontré que les systèmes de traçage risquent de porter atteinte aux valeurs fondamentales de la vie privée, tant des citoyens qui soupçonnent une intrusion de l'État que de ceux qui craignent l'humiliation potentielle, voire les représailles, qu'entraînerait l'exposition d'informations personnelles à des tiers [18, 43, 44, 20]. Les systèmes de traçage automatique pouvant avoir une portée bien plus grande que les efforts de traçage manuel, l'effet potentiel sur la vie privée des individus est proportionnellement important. Toutefois, malgré l'impact potentiel significatif sur la vie privée, la capacité de collecter des données sur la propagation de l'épidémie à un rythme beaucoup plus rapide que celui qui est possible avec la méthode de traçage manuel peut s'avérer une solution clé pour la combattre. Les gouvernements des États démocratiques se trouvent donc confrontés à un grave dilemme : pour traverser la pandémie, ces gouvernements doivent-ils choisir entre la préservation de la vie ou les principes fondamentaux des droits de la personne ?

Heureusement, il s'agit d'un faux dilemme. Si tout système de traçage des contacts implique des compromis, certains de ces compromis seront plus acceptables pour les sociétés démocratiques que d'autres. La méthodologie de traçage des contacts de COVI n'est pas une extension des efforts de l'autorité centrale de l'État pour localiser et informer les individus. Le système COVI est également conçu pour éviter autant que possible que des tiers puissent connaître le niveau de risque ou le statut d'infection d'une personne, tout en échangeant

suffisamment d'informations pour influencer positivement le comportement individuel et permettre à ces autres utilisateurs de se protéger et de protéger les autres.

Le coût qu'implique l'approche COVI est principalement supporté par l'État, qui dispose de moins d'informations pour exercer un contrôle direct sur la pandémie. Toutefois, cette absence de contrôle direct ne doit pas nécessairement se traduire par un préjudice pour la communauté. L'utilisation de l'application COVI à la place de méthodes plus invasives de traçage des contacts peut toujours bénéficier de l'objectif premier du traçage des contacts, qui est d'isoler et de tester les personnes présentant un risque plus élevé d'infection, à condition que les recommandations faites par l'application soient suivies par la grande majorité des participants. En outre, les États peuvent toujours bénéficier des données agrégées que le système collecte, ce qui facilitera l'élaboration de politiques. En conséquence, le compromis présenté par COVI se traduit par un impact négatif minimal sur la société civile des États démocratiques.

De plus, COVI demande le consentement pour la collecte, l'utilisation et la divulgation des informations nécessaires à la fonction de l'application ou pour optimiser son efficacité. Afin de promouvoir l'autonomie des individus dans le cadre du consentement COVI, nous proposons également des options lorsque cela est possible. Pour plus d'informations sur la formule de consentement proposée, veuillez consulter la section 2.

L'approche de COVI en matière de protection de la vie privée implique une stratégie à plusieurs facettes qui fait de la protection de la vie privée un élément central de la fonctionnalité de l'application. La principale approche de COVI pour protéger la vie privée est de décentraliser la communication des risques entre les utilisateurs [45, 46, 13]. De plus, les informations sensibles concernant le traçage numérique des contacts (leurs réseaux de contacts) sont par défaut uniquement stockées sur leur téléphone. Cela se fait par le biais de protocoles cryptographiques pour communiquer en toute sécurité entre les téléphones et entre le serveur hébergeant le modèle d'apprentissage automatique et le modèle épidémiologique. Enfin, les données pseudonymisées et agrégées sur les profils individuels et d'interactions des utilisateurs, compilées avec leur consentement, sont gérées par une organisation indépendante à but non lucratif dont la seule mission est de soutenir les utilisateurs dans leur lutte contre l'épidémie de COVID-19, tout en protégeant leur santé, leur dignité et leur vie privée. Ces données sécurisées seront utilisées pour former les modèles d'apprentissage automatique prédictifs et épidémiologiques. Les données recueillies ne pourront jamais être utilisées à des fins commerciales ni vendues à des entreprises privées et seront toutes supprimées dès la fin de la pandémie. De plus, les informations stockées sur les appareils des utilisateurs seront purgées sur une base continue, généralement tous les 30 jours (voir section 2.6.4). Elles ne peuvent pas être utilisées à des fins de surveillance ou pour imposer une quarantaine, et le gouvernement n'a pas accès aux données au-delà des données agrégées partagées avec les autorités sanitaires pour éclairer la prise de décision.

Cette approche - qui met le pouvoir entre les mains des citoyens - est le complément naturel du modèle de protection de la vie privée [47]. Les données confèrent un pouvoir (c'est pourquoi les préoccupations relatives à la protection de la vie privée sont si importantes) ; par conséquent, le citoyen pourrait donner du pouvoir à un gouvernement, par exemple, en lui fournissant des données. En gardant le contrôle de leurs données, les citoyens eux-mêmes ont le pouvoir et la responsabilité de prendre des mesures pour faire face à la crise.

L'application sert à organiser l'information de manière à aider les individus à prendre des décisions en connaissance de cause, afin qu'ils puissent utiliser leur pouvoir de manière efficace et assumer les responsabilités qu'ils prennent pour influencer l'issue de la crise. Les mesures prises restent entre les mains des personnes qui choisissent d'installer COVI : les citoyens décident de la quantité de données qu'ils souhaitent envoyer et de la manière dont ils répondent aux recommandations que l'application contient.

Toutefois, il convient de mentionner que l'utilisation par COVI d'un protocole plus sophistiqué de transmission de messages liés aux risques, à l'inverse du plus traditionnel traçage de contacts binaire, présente des risques supplémentaires pour la vie privée. Nous pensons que ces risques sont acceptables face à l'important défi de

santé publique que représente la pandémie et à l'avantage apporté par l'apprentissage automatique en termes de réduction du facteur de reproduction du virus (le nombre de nouvelles personnes infectées par personne infectée), voir la section 4.5 ; nous avons donc élaboré un protocole conçu pour minimiser les risques d'atteinte à la vie privée qu'ils introduisent. Dès le début du projet, nous avons travaillé avec le Commissariat à la protection de la vie privée du Canada à la révision de notre protocole et avons appliqué les principes énoncés le 7 mai 2020 dans la Déclaration commune des commissaires fédéraux, provinciaux et territoriaux à la protection de la vie privée intitulée « Appuyer la santé publique et bâtir la confiance des Canadiens : principes de protection de la vie privée et des renseignements personnels pour les applications de traçage des contacts et autres applications similaires » [48]. En outre, nous avons cherché à respecter les principes fondamentaux du « Privacy by Design » (protection de la vie privée dès la conception) [49]. Pour une explication détaillée du protocole et de sa surface d'attaque, veuillez consulter la section 2.3.

1.2.4 Protéger les droits de la personne

Le droit à la vie privée est protégé par la Charte canadienne des droits et libertés ainsi que par la Charte québécoise des droits et libertés de la personne. Cependant, les technologies, y compris l'application COVI, peuvent avoir un impact sur d'autres droits et libertés. COVI a adopté une approche fondée sur les droits de la personne lors de la conception et de la mise en œuvre de l'application [50]. Nous avons également pensé à faire en sorte que l'application ne respecte pas seulement « passivement » les droits de la personne (par exemple en veillant à ce que l'algorithme ne reproduise pas de discrimination ou de préjugés), mais nous avons également mis en place un ensemble de mesures pour garantir que les droits et libertés des Canadiens soient activement protégés à tout moment.

À titre d'exemple, le projet COVI a pris les mesures suivantes :

- COVI sera disponible dans plusieurs langues autres que le français et l'anglais, y compris les langues autochtones.
- COVI recueillera volontairement, et sur la base du consentement des utilisateurs, des données supplémentaires pour aider à promouvoir notre compréhension sur la façon dont la pandémie et les mesures de santé publique associées ont un impact sur différentes communautés ou populations.
- La gouvernance de COVI, en tant qu'organisme indépendant à but non lucratif, est structurée de manière à garantir que les membres de communautés marginalisées ainsi que les peuples autochtones sont impliqués dans la prise de toutes les décisions importantes liées à l'application et aux données qui lui sont associées. En outre, l'un des mandats principaux informant la gouvernance de COVI consistera à garantir l'inclusion, la diversité et l'équité à tout moment.
- COVI et sa gouvernance sont indépendants des gouvernements et ne permettront à aucune de ses données d'être utilisée à des fins de surveillance, de répression ou à d'autres fins connexes.

1.3 Aperçu de l'application

1.3.1 L'interface utilisateur

L'interface de COVI pour les utilisateurs fonctionne comme suit. Lors du téléchargement, les utilisateurs reçoivent un aperçu du fonctionnement de l'application et des conséquences de l'utilisation de COVI sur la protection de la vie privée (voir section 2). Après vérification de l'âge de l'utilisateur, celui-ci est invité à remplir un bref questionnaire sur les données démographiques et les conditions médicales préalables pour initialiser

l'application. Par défaut, toutes ces données sont conservées sur le téléphone. Une fois son installation terminée, l'utilisateur arrive sur l'écran d'accueil.

L'écran comporte quatre éléments principaux :

- La fonction de conseils personnalisés qui aide les utilisateurs à prendre quotidiennement des décisions en temps réel concernant leurs activités en fonction de leur niveau de risque personnel (actions hors ligne).

- La fonction « actions à prendre » qui invite les utilisateurs à saisir des informations supplémentaires ou à les actualiser pour mieux adapter leur profil de risque (actions intégrées à l'application).

- Une fonction de sondage et de visualisation des données qui permet aux utilisateurs d'exprimer ce qui est important pour eux et de voir comment la crise se déroule.

- Un bouton « partager » afin que l'utilisateur puisse promouvoir l'adoption de l'application COVI parmi ses amis, sa famille, ses collègues, etc.

Au-delà de ces éléments figurant sur l'écran d'affichage, COVI envoie également des notifications à l'utilisateur, soit pour mettre à jour ses informations dans l'application (actions de faible priorité), soit lorsque des recommandations urgentes sont mises à jour (actions de haute priorité). Lorsque les utilisateurs n'utilisent pas activement l'application, celle-ci fonctionne en arrière-plan, échangeant les niveaux de risque (de manière protégée par cryptographie) avec les autres utilisateurs de l'application qu'ils rencontrent. Si un utilisateur fait un test de dépistage pour la COVID-19, il pourra, dans les versions antérieures, signaler lui-même les résultats du test. Dans les prochaines versions, les utilisateurs pourront récupérer les résultats de leur test directement dans l'application COVI. Lorsqu'ils saisiront (ou recevront) un résultat positif, il leur sera demandé de donner leur accord pour qu'il soit communiqué (en raison de leur niveau de risque élevé) de manière confidentielle aux utilisateurs qu'ils ont récemment rencontrés.

1.3.2 Processus d'arrière-plan

La fonctionnalité de l'application COVI destinée aux utilisateurs est alimentée par des processus en arrière-plan qui prédisent le niveau de risque de l'utilisateur, enregistrent les rencontres faites et récupèrent les informations sur le niveau de risque pour les contacts récents. L'application COVI est essentiellement axée sur le traçage des contacts et la notification de potentielle exposition au virus en combinant tous les indices disponibles. Comme de nombreuses autres propositions de traçage automatique des contacts [45, 10, 11, 12, 13], les téléphones utilisent les informations provenant de la technologie Bluetooth pour déterminer les contacts. Nous testons actuellement plusieurs plateformes Bluetooth différentes, notamment le protocole TCN [45], les API (interface de programmation d'applications) de notification d'exposition de Google et Apple [13] et un nouveau système développé par le NHS [51], pouvant déterminer les occurrences de contacts (voir 2.3).

Contrairement à de nombreux autres efforts de traçage numérique des contacts, au lieu de simplement retracer les expositions binaires aux cas de COVID-19 ayant reçu un diagnostic positif, COVI utilise l'apprentissage automatique pour calculer localement les niveaux de risque scalaires qui estiment quand un utilisateur peut avoir été infecté et quel peut être son niveau de contagion sur différents jours, dans un délai récent. Bien que les cas diagnostiqués de COVID-19 représentent un risque maximal, les personnes qui n'ont pas été diagnostiquées peuvent quand même présenter un risque non nul, c'est-à-dire qu'elles peuvent être asymptomatiques, mais contagieuses. Lorsque ces niveaux de risque sont sensiblement modifiés pour un jour particulier dans le passé, ils sont alors envoyés à tous les contacts de ce jour, ce qui permet au réseau d'applications COVI de recalculer les niveaux de risque de chacun de manière décentralisée, améliorant ainsi la précision globale des prévisions.

Bien que les niveaux de risque jouent un rôle majeur en arrière-plan, ils ne sont jamais explicitement démontrés à l'utilisateur.

Notez que bien que nous utilisions la technologie Bluetooth pour identifier la proximité avec d'autres utilisateurs, nous utiliserons le GPS local (transformé de manière appropriée pour obtenir le k-anonymat) comme une caractéristique indirecte pour le prédicteur de risque et pour la modélisation épidémiologique. Pour les volontaires qui consentent à fournir leurs données pour la recherche, des données de localisation à faible résolution spatiale seront également envoyées au serveur d'apprentissage automatique de COVI et cumulées entre les utilisateurs afin de créer des cartes thermiques, sans les associer à des utilisateurs individuels. Il suffit qu'une petite fraction de la population choisisse de participer pour produire suffisamment de données d'entraînement pour le prédicteur.

1.4 Utilisation de l'apprentissage automatique

Les informations générées par le traçage automatique des contacts à l'état brut ne sont pas très exploitables par les utilisateurs. Une personne qui a eu un seul contact de courte durée avec une personne infectée doit-elle prendre la quarantaine auto-imposée avec autant de sérieux qu'une personne qui a eu plusieurs contacts de longue durée (par exemple sur un lieu de travail) ? Si une personne a signalé ressentir de nombreux symptômes de la COVID-19, mais n'a pas encore pu se faire tester, ses contacts doivent-ils quand même être avertis ? Dans le cadre du dépistage manuel, un professionnel de la santé utilise son jugement professionnel pour faire des recommandations aux contacts, ce qui prend du temps et nécessite une expertise importante. COVI vise à tirer parti de l'apprentissage automatique pour optimiser et automatiser l'intégration des indicateurs concernant la possibilité qu'une personne soit infectée, et utiliser les niveaux de risque gradués qui en résultent pour formuler des recommandations appropriées et envoyer des alertes appropriées aux autres utilisateurs afin qu'ils puissent mettre à jour leur propre évaluation des risques. Les utilisateurs de COVI peuvent choisir d'envoyer leurs données (voir section 2.4) au serveur sécurisé d'apprentissage automatique de COVI, où elles sont utilisées pour former deux modèles distincts, mais complémentaires, le prédicteur de risque et le simulateur épidémiologique.

1.4.1 Prédicteur de risque

Au lieu de présenter aux utilisateurs des données brutes sur l'occurrence de contacts, COVI calcule à l'interne un ensemble de niveaux de risque sur les deux dernières semaines. Les niveaux de risque personnels calculés sont basés sur une combinaison de symptômes signalés par les utilisateurs, d'informations démographiques et d'informations sur l'interaction avec d'autres contacts, y compris l'estimation du niveau de contagion (niveaux de risque) des personnes rencontrées. Ils sont ensuite utilisés à deux fins. Premièrement, le niveau de risque actuel est intégré dans les recommandations personnalisées que l'application fait aux utilisateurs (voir ci-dessous). Plus précisément, un modèle d'apprentissage automatique prédit la probabilité qu'une personne ait été infectée et le degré de contagion de cette personne dans un passé récent et aujourd'hui. L'estimation du niveau de contagion des jours passés est cruciale pour informer l'application des autres utilisateurs rencontrés par le passé afin qu'ils puissent recalculer leur propre niveau de risque. Par exemple, imaginez qu'Alice et Bob ont passé beaucoup de temps ensemble il y a 3 jours, et qu'en raison des nouvelles informations disponibles, l'appareil de Bob estime qu'il est probablement infecté et qu'il a probablement été très contagieux au cours des 4 derniers jours. Le téléphone de Bob enverrait alors un message à l'appareil d'Alice pour qu'il actualise son niveau de contagion d'il y a trois jours. En disposant de ces informations actualisées sur le risque à un moment où elle pourrait elle-même devenir contagieuse, mais avant l'apparition des symptômes, COVI permet d'alerter Alice de façon précoce avant qu'elle-même contribue à la propagation du virus. COVI renforcerait sur l'application d'Alice des messages suggérant qu'elle s'isole davantage et minimise les contacts. Si Alice réagit comme la plupart des personnes qui se rendent compte qu'elles peuvent être infectées, elle agira de manière responsable et réduira considérablement la propagation silencieuse du virus qui se serait produite autrement. D'un point de vue de

probabilité, le prédicteur de risque prend les données observées au cours des deux dernières semaines et prédit la distribution de probabilité des variables non observées passées (comme le fait d'avoir été infecté lors d'une rencontre particulière, ou le degré de contagion sur différents jours dans le passé).

1.4.2 Simulateur épidémiologique

Les données pseudonymisées volontaires peuvent également être utilisées pour adapter un modèle épidémiologique au niveau individuel qui saisit le flux stochastique des événements dans le temps, à travers des événements asynchrones correspondant à des mouvements de personnes, des rencontres entre personnes, des occurrences médicales (comme être infecté, avoir une charge virale particulière ou certains symptômes pertinents) et des comportements (comme porter un masque au travail, passer plus ou moins de temps dans différentes catégories d'endroits comme les magasins, les bureaux, les hôpitaux ou les parcs). Le modèle épidémiologique comprend des connaissances préalables sur les aspects pertinents de la vie des personnes (comme les déplacements et les comportements tels que le port d'un masque) et est structuré autour de nombreuses probabilités conditionnelles pour les événements ci-dessus qui modifient l'état du système. Ces probabilités conditionnelles sont paramétrées et ces paramètres peuvent être estimés, à l'aide de méthodes décrites dans la section 4, en profitant du prédicteur de risque pour échantillonner les variables non observées telles que le fait d'être infecté et le degré de contagion. Ces modèles épidémiologiques peuvent ensuite être intégrés dans un simulateur qui peut être utilisé par les responsables de la santé publique de plusieurs manières : pour cartographier géographiquement l'évolution de la maladie (par exemple, les zones où les gens sont infectés plus rapidement, avant que l'augmentation du nombre de cas ne soit visible dans les hôpitaux), pour comprendre les choix des citoyens (par exemple, où les gens sont-ils plus ou moins bons à suivre et respecter les recommandations) et pour mieux définir les facteurs qui comptent pour la contagion et la manière dont ils interagissent. Ces modèles épidémiologiques peuvent également être utilisés pour simuler l'évolution des foyers de contagion dans différents scénarios hypothétiques, et pour optimiser les politiques publiques par rapport à des objectifs tels que la réduction du nombre d'hospitalisations dues à la maladie ou le maintien du taux de reproduction de base quotidien R_t en dessous de 1.

Voir les sections 3 et 4 pour plus de détails sur le prédicteur de risque et le modèle épidémiologique et sur la façon dont ils peuvent être formés avec des méthodes telles que l'inférence variationnelle.

1.5 L'expérience utilisateur

Si la mise en œuvre technologique présentée ici représente un outil viable pour réduire les infections à la COVID-19, le succès du traçage numérique est fortement lié à l'adhésion des citoyens, à leur participation au partage des données et à l'utilisation continue de l'application. Dans la même veine, en raison des risques inhérents associés au partage de données privées par les citoyens et à la réception d'informations de santé publique, COVI a été conçu pour s'aligner sur les intérêts des utilisateurs. La science psychologique nous a montré que la meilleure façon d'y parvenir n'est pas de recourir à la coercition, mais de susciter les préférences en constante évolution des utilisateurs. La conception de COVI sera donc largement axée sur la création de mécanismes de mesure (sondages intégrés à l'application, groupes de discussion, etc.) qui nous permettront de mieux comprendre les utilisateurs et d'ajuster les paramètres et les environnements informationnels de manière à répondre à leurs intérêts exprimés (et non supposés). Des approches fondées sur des données probantes provenant de divers sous-domaines scientifiques sont utilisées pour atteindre ces objectifs de plusieurs manières, ce qui donne lieu aux principes fondamentaux suivants qui guident la feuille de route de COVI.

- Les préférences des utilisateurs déterminent l'expérience d'un bout à l'autre. Alors que les recommandations des gouvernements sur les réponses personnelles appropriées à avoir lors de la crise évoluent et deviennent plus graduelles et dépendantes de la situation, il est important de comprendre l'évolution des préférences des utilisateurs en matière de risques. COVI utilise divers outils de mesure

- au niveau de la population, pendant le processus d'installation et tout au long du cycle de vie de l'application - afin de connaître ces préférences. Il est important de noter que des décennies de recherche ont montré qu'il est insuffisant de demander aux gens ce qu'ils préfèrent - les préférences doivent être sollicitées, validées et mises à jour régulièrement [52, 53].

- En même temps, nous comprenons que l'engagement initial, l'interaction régulière et l'utilisation continue sont essentiels à l'impact de COVI sur les résultats sanitaires au niveau de la population [8]. Ainsi, la prise en compte effective des préférences des utilisateurs doit être combinée à une conception informationnelle et visuelle attrayante. Nous y parvenons en nous appuyant sur des listes de contrôle ergonomiques, les meilleures pratiques en matière d'expérience utilisateur et des vérifications constantes de la facilité d'utilisation. De plus, l'accent est mis sur la création de mécanismes de mesure de l'engagement qui permettent de tester et de réitérer constamment les variantes. Cette combinaison d'un suivi étroit des préférences des utilisateurs et d'une expérience utilisateur efficace génère une interface dynamique qui s'adapte à chaque utilisateur, ce qui permet à leurs interactions continues d'être à la fois stimulantes et engageantes.

- La compréhension de l'utilisateur est priorisée et vérifiée plutôt que supposée. La psychologie de la divulgation nous dit qu'il y a un monde de différence entre divulguer en théorie et communiquer efficacement [54]. Dans un effort pour donner à COVI une transparence totale, l'application (1) affiche en permanence et de façon bien visible un lien vers les méthodes de protection de la vie privée utilisées par COVI, ainsi que vers l'accord de confidentialité lui-même et (2) utilise des tests d'utilisateurs pour s'assurer que ces déclarations ne sont pas seulement présentes de facto, mais qu'elles se reflètent également dans la conscience des utilisateurs.

- La capacité des utilisateurs à se protéger et à protéger les autres est maximisée. Une communication efficace avec le public concernant l'évolution de la situation et les mesures à prendre en cas de crise est essentielle, en particulier lorsque la réaction du public est le principal moteur d'amélioration de la situation. Les informations doivent être communiquées clairement et les recommandations doivent émaner d'une autorité. Toutefois, comme pour toute communication, le sens se trouve dans la réponse. Dans cette optique, nous avons combiné une approche basée sur (1) des preuves basées sur le domaine de la communication de crise [55], (2) une collecte continue de données primaires examinant les réactions des Canadiens aux variations des messages et (3) des données d'utilisateurs examinant le lien entre les messages et la probabilité de réduire les comportements à risque. Nous utilisons ces outils pour nous assurer que l'application permet aux utilisateurs d'agir de la manière qu'ils jugent appropriée.

- Le bien-être psychosocial de l'utilisateur est favorisé. Étant donné la nature sensible de la communication contenue dans l'application COVI, il est important de surveiller les réactions des utilisateurs et de s'assurer que nous ne créons pas de tension excessive. Si l'absence totale de stress est une réaction inappropriée à une crise, le fait de créer une urgence qui ne peut pas faire l'objet d'une action peut provoquer une méfiance et une fatigue qui entravent la mise en œuvre des recommandations. En outre, une telle tension psychologique peut avoir des effets plus graves, allant de l'augmentation des niveaux d'anxiété à l'augmentation des cas de violence domestique [56]. Il faut donc veiller tout particulièrement à ce que les données et les recommandations contenues dans COVI soient soigneusement élaborées pour fournir les informations nécessaires qui limitent la tension psychologique.

- L'inclusion des utilisateurs reflète la diversité de leurs besoins. La crise de la COVID-19 présente des pressions et des risques variés pour différents segments de la population [57]. Malheureusement, les groupes marginalisés [58] sont à la fois les plus susceptibles d'être touchés et les moins susceptibles d'avoir accès et d'utiliser un outil tel que COVI. C'est pourquoi COVI utilise les meilleures pratiques en

matière d'accessibilité dès le départ et tire parti des modèles d'accès, de diversité et d'inclusion pour identifier les éventuelles lacunes le plus tôt possible dans le cycle de vie du produit afin de les combler le plus rapidement possible.

- La dynamique des genres dans la réponse des Canadiens à la COVID-19 est une question qui présente un intérêt particulier dans le contexte de l'inclusion [59]. En incluant des structures d'analyse des genres dans nos tests de comportement des utilisateurs en cours, nous acquérons continuellement une meilleure compréhension des risques associés à la dynamique des genres (par exemple, le risque que la technologie de traçage soit utilisée comme moyen de limiter la liberté de mouvement d'un partenaire) ainsi que des circonstances opportunes (par exemple, le développement de modèles sur la pénétration de COVI dans les ménages tout en tenant compte des questions des genres).

Voir la section 5 pour plus de détails sur les aspects psychologiques et l'interface utilisateur de l'application.

1.6 Comparaison avec d'autres approches

Il existe de nombreuses approches différentes pour le traçage des contacts [12, 10, 28, 13, 20, 60, 11, 61]. Bien qu'une taxonomie complète des méthodes de traçage des contacts dépasse le cadre de ce livre blanc, nous pensons qu'il est important de discuter de certains des principaux choix de conception que nous avons faits lors de l'élaboration de COVI, car ils sont liés à certaines décisions importantes.

1.6.1 Manuel ou automatique

Une considération importante est le degré auquel un humain doit être impliqué. Rappelons que la méthode classique de traçage des contacts est entièrement manuelle et implique qu'une personne retrace les contacts faits et demande au patient de se rappeler tous ses contacts et tous les lieux visités au cours des deux dernières semaines. L'approche SafePaths/SafePlaces du MIT PrivateKit [20] et l'application TraceTogether [10] de Singapour complètent la méthode manuelle standard de traçage des contacts, en fournissant des informations supplémentaires pour faciliter le traçage manuel des contacts, tout en conservant le discernement et le contact humain. Cela implique un travail important de la part des autorités de santé publique, mais permet également de porter un jugement professionnel minutieux sur la gravité des contacts.

D'autre part, les approches entièrement automatiques [60, 28] ont l'avantage de nécessiter beaucoup moins de travail de la part des autorités de santé publique, mais peuvent également ne pas bénéficier des avantages d'un jugement professionnel. Les approches entièrement automatiques peuvent également être plus vulnérables à des attaques sur le système par des entités malveillantes, car il n'y a pas de discernement humain intégré à chaque étape pour le protéger.

En raison de la surcharge des autorités de santé publique canadiennes, nous avons choisi de faire en sorte que COVI se rapproche beaucoup plus de l'extrémité entièrement automatique du spectre, tout en conservant un point de contact avec les autorités de santé publique en fournissant aux utilisateurs à haut risque ou infectés, des recommandations pour qu'ils se soumettent à des tests. De cette façon, COVI peut être complémentaire au dépistage manuel tout en ayant la possibilité d'avoir un impact positif significatif en soi, sans qu'il soit nécessaire d'avoir une intervention humaine au moment où cela compte, c'est-à-dire lorsque des alertes précoces sont propagées par le réseau de contacts. Enfin, la suppression d'une intervention humaine réduit le risque de violation de la vie privée et d'utilisation abusive des données personnelles des utilisateurs par une autorité gouvernementale.

1.6.2 Types de messages de risque

La grande majorité des applications de traçage des contacts envoient des notifications d'exposition binaires, ou tout au plus une notification à deux niveaux où les diagnostics symptomatiques et les tests cliniques sont différenciés. Comme l'application de traçage des contacts du NHS [62], COVI cherche à envoyer des messages de risque à plusieurs niveaux. La nouvelle interface de programmation d'application (API) de notification d'exposition proposée par Google et Apple, et le protocole TCN comprennent également une certaine prise en charge des messages de risque non binaires [13, 45].

Bien que le traçage binaire des contacts révèle moins d'informations sur les utilisateurs (notamment, les utilisateurs qui ne sont pas diagnostiqués ne révèlent aucune information), l'envoi de messages de risque à plusieurs niveaux permet de faire des recommandations personnalisées précoces et précises aux individus. Nous pensons que la valeur ajoutée qu'offre COVI en matière de niveaux de risque personnel et de recommandations plus précises vaut la peine de faire un compromis en demandant aux utilisateurs d'envoyer plus d'informations à leurs contacts. Ces informations supplémentaires se traduisent par des recommandations personnalisées plus précises et peuvent avoir un impact significatif sur la capacité de COVI à donner aux utilisateurs les connaissances dont ils ont besoin pour se protéger et protéger les autres, en particulier dans la phase présymptomatique de la maladie. L'utilisation de l'apprentissage automatique pour intégrer des indices complexes qui, autrement, nécessiteraient l'intuition humaine, atténue l'absence d'intervention humaine directe dans une application de traçage des contacts entièrement automatique. En envoyant des niveaux de risque à plusieurs niveaux et en permettant des prévisions de risque précises, COVI peut automatiquement donner des recommandations plus pertinentes et effectuer un triage des occurrences d'exposition potentielles. Bien entendu, les utilisateurs recevront des recommandations pour contacter un professionnel de la santé le cas échéant, et en complétant le dépistage binaire par des messages de risque, COVI comble les lacunes du traçage automatisé.

1.6.3 Centralisation des données

Les types d'attaques et d'adversaires que l'on cherche à contrecarrer jouent un rôle central dans la conception des applications de traçage des contacts. Le volume d'informations sensibles potentielles, allant des contacts sociaux aux antécédents médicaux en passant par les historiques de localisation, sur un large sous-ensemble de citoyens soulève à juste titre des questions sur les abus [19, 11, 12]. Ainsi, le degré de confiance des citoyens dans les autorités centrales devrait être un élément central dans la conception d'une application de traçage des contacts.

Une façon simpliste de concevoir une application de traçage des contacts consiste à télécharger la trajectoire complète et les coordonnées de tous les utilisateurs vers une autorité centrale, qui effectue les recherches. Le gouvernement israélien semble avoir proposé une telle approche [63]. Il est évident que cela n'a rien de privé par rapport aux autorités gouvernementales, mais cela présente l'avantage de permettre au gouvernement de disposer de données détaillées sur lesquelles il peut prendre des décisions de santé publique, en plus de lui permettre d'exercer un jugement professionnel pour le traçage manuel des contacts. De plus, comme toutes les données sont directement détenues par l'autorité centrale, en l'absence de violation de données (bien que celles-ci soient d'une fréquence inquiétante [64]), les données de chacun sont protégées des autres personnes.

Cependant, de nombreux résidents d'autres pays sont moins disposés à transmettre toutes leurs données à une seule autorité centrale. Surtout si l'installation d'une application est sur une base volontaire, il devient alors nécessaire de concevoir des applications qui offrent de meilleures garanties en matière de respect de la vie privée. De nombreuses applications ont donc adopté une approche partiellement centralisée dans le cadre du dépistage binaire, où seuls les utilisateurs diagnostiqués téléchargent leurs données de contact/traçage vers un serveur central. Le serveur central peut alors utiliser ces données pour informer les utilisateurs qu'ils ont peut-

être été exposés. La méthode manuelle standard de traçage des contacts entre dans cette catégorie, ainsi que l'application TraceTogether [10]. Cette approche partiellement centralisée fournit à l'autorité centrale des données sur tous les utilisateurs diagnostiqués et leurs contacts, et constitue un compromis qui semble gagner du terrain dans certains cercles, comme l'initiative de traçage de proximité paneuropéenne, PEPP-PT (Pan-European Privacy Preserving Proximity Tracing) [14]. Malheureusement, cette approche partiellement centralisée est incompatible avec l'envoi de messages de risque non binaires de COVI, car les utilisateurs non diagnostiqués envoient des messages de risque tout comme les utilisateurs infectés. Toute autorité centrale obtiendrait donc des données sur la quasi-totalité des utilisateurs de l'application.

Plus loin encore sur le spectre, on trouve le groupe d'approches totalement décentralisées, qui tentent d'empêcher toute autorité de disposer des informations complètes de contact/traçage. Bien que la plupart de ces approches impliquent l'envoi de données via un serveur central, le serveur ne reçoit pas de données non cryptées/identifiantes dans les approches décentralisées. COVI, et beaucoup d'autres propositions récentes, y compris DP-3T [28], covid-watch [61], la coalition TCN [45], MIT PACT [12], Washington PACT [11], et la nouvelle API de Google et Apple [13] tentent de répondre à cette norme en utilisant une variété de technologies différentes. Les approches totalement décentralisées pour le traçage des contacts binaire sont sans doute plus faciles à protéger, bien qu'il existe encore des limites inhérentes à la protection de la vie privée (section 2.2).

En raison des limitations techniques liées à la largeur de bande pour l'envoi de nombreux messages de risque, COVI utilise actuellement une stratégie différente (section 2.3) tout en évaluant la possibilité de passer soit à l'API de Google et Apple dans l'intérêt de l'interopérabilité et de l'utilisation de normes ouvertes.

2 Protection de la vie privée et consentement

Comme indiqué ci-dessus à la section 1.2.3, COVI est conçu pour recueillir, utiliser et communiquer le moins de renseignements personnels possible et offrir aux utilisateurs autant de possibilités de donner leur consentement tout en répondant aux objectifs visés. Ce faisant, COVI se conforme aux exigences des lois canadiennes sur la protection des renseignements personnels dans le secteur privé. À plusieurs égards, COVI vise bien au-delà de l'effort minimal nécessaire pour satisfaire aux exigences de ces lois sur la protection de la vie privée, qui permettent généralement de trouver un équilibre flexible et pragmatique entre les droits des personnes et les besoins des organisations qui utilisent des renseignements personnels pour fournir leurs services. COVI ne cherche pas à tirer profit de ce que la loi permet ; il cherche plutôt à fournir ses services tout en respectant la vie privée dans le sens strict qu'il offre à l'individu un contrôle maximum sur les informations qui le concernent [65, 66].

Au Canada, la collecte, l'utilisation et la divulgation de renseignements personnels dans le secteur privé sont régies par la Loi sur la protection des renseignements personnels et les documents électroniques et par toutes les lois provinciales jugées essentiellement similaires à celle-ci. Bien qu'il existe des différences entre les lois fédérales et provinciales, ces lois visent généralement à garantir que les organisations qui recueillent, utilisent et divulguent des renseignements personnels, entre autres, (i) sont tenues responsables des renseignements personnels dont elles ont la charge, (ii) sont transparentes quant à leurs pratiques en matière de protection de la vie privée et aux fins pour lesquelles les renseignements personnels sont recueillis, (iii) respectent le principe de minimisation des données, (iv) appliquent des mesures de sécurité appropriées et (v) recherchent un consentement valable sous une forme adaptée aux circonstances et à la nature des renseignements personnels.

Lorsque l'on examine le modèle de protection de la vie privée de COVI, il est important de garder à l'esprit que les lois sur la protection de la vie privée - y compris les lois canadiennes - sont généralement formulées en partant du principe qu'un certain niveau de confiance devra être accordé à une organisation qui agira en tant que gardien ou régulateur des renseignements personnels. Ces lois visent à garantir que cette confiance est bien fondée en prévoyant des règles régissant des considérations telles que celles soulevées au paragraphe

précédent, et, en dotant les autorités de réglementation du pouvoir d'enquêter sur des plaintes, de publier des rapports publics, ainsi que dans certaines circonstances, d'imposer des sanctions monétaires.

En ce qui concerne ses fonctions essentielles, le modèle de protection de la vie privée de COVI a été conçu pour éliminer, dans la mesure du possible, la nécessité de ce type de confiance. Comme le montre la discussion à la section 2.3, les mesures prises pour chiffrer et brouiller à la fois le contenu et l'acheminement des données qui traversent le système de messagerie, ainsi que la décentralisation ou la fédération délibérée du contrôle des divers éléments de ce système, visent collectivement à éliminer la nécessité d'accorder une confiance importante aux organisations qui agissent en tant que fournisseurs de services dans le système ou aux autres utilisateurs (dans certaines limites). Bien qu'il ne soit pas entièrement réalisable dans les faits, l'objectif de cette structure est de rendre insignifiantes pour quiconque, les informations circulant dans le système, sauf pour le destinataire visé, et de faire obstacle aux tiers qui cherchent à compromettre cet objectif par l'espionnage, la coercition ou la corruption. Les lois sur la protection de la vie privée telles qu'interprétées par les autorités de réglementation n'obligent généralement pas aux organisations d'aller jusqu'à ces extrêmes pour respecter les exigences liées aux garanties de sécurité et à la minimisation des données.

Cela étant dit, comme la confiance des utilisateurs doit être accordée au système dans son ensemble, COVI cherche également à maximiser sa responsabilité et sa transparence en appliquant un modèle de consentement explicite pour garantir un consentement valable. Afin de garantir cette responsabilité, COVI et son écosystème d'information seront placés sous la supervision d'une entité indépendante constituée à cet effet, qui agira en tant qu'administrateur du système COVI et assumera la responsabilité de son fonctionnement ainsi que de l'optimisation continue du modèle de protection de la vie privée. En ce qui concerne la transparence, en plus de la politique habituelle de protection de la vie privée, COVI publiera des infographies accessibles sur son site web, le présent livre blanc, et publiera le code source pour inspection dans le cadre d'un modèle de licence à code source ouvert. Afin d'obtenir un consentement valable, COVI demande un consentement séparé pour les différents éléments d'information personnelle et rend le consentement facultatif dans la mesure du possible, comme expliqué plus en détail dans la section 2.1.

2.1 Consentement

COVI demande le consentement pour la collecte, l'utilisation ou la divulgation de différents éléments de renseignements personnels à différentes étapes de l'expérience utilisateur. Bien que la forme de consentement demandée soit toujours explicite, elle est présentée comme une condition de service (c'est-à-dire, requise) ou facultative selon la nature des informations en question et les fins auxquelles elles seront utilisées. Dans ce qui suit, nous décrivons le consentement demandé à chacun de ces moments.

2.1.1 Consentement à l'utilisation des données pour les fonctions essentielles

Lors de l'installation et du démarrage de l'application, les utilisateurs de COVI sont invités à consentir expressément à ce que COVI puisse recueillir, utiliser et divulguer le minimum d'informations nécessaires pour effectuer le traçage des contacts, calculer le risque d'infection et échanger des messages de risque. Lors de la mise en œuvre initiale, le consentement recueilli sera le suivant :

- Le consentement des utilisateurs de l'application sera obtenu pour la collecte, l'utilisation et la divulgation des informations suivantes, comme condition d'utilisation du service, au moyen d'un langage de consentement et d'une politique de confidentialité :
    - *Historique de géolocalisation par GPS (seules les positions floues sont conservées, au niveau des aires de diffusion de Statistique Canada, dans le but de prédire le risque lié au lieu et de modéliser l'évolution géographique de la maladie)*

- Identification aléatoire des contacts générée par la demande
- Les niveaux de risque actuels de l'utilisateur

L'application ne fonctionnera pas correctement sans ces informations, c'est à dire que le consentement de l'utilisateur est une condition d'utilisation de COVI. Il est toutefois important de noter qu'en ce qui concerne la divulgation de ces informations, le protocole de confidentialité cherche à réduire au maximum leur contenu informationnel avant qu'elles ne quittent l'appareil de l'utilisateur ; les niveaux de risque, par exemple, sont envoyés aux contacts sans que des tiers (y compris le gouvernement) aient accès à ces informations ou puissent les relier à un individu (voir 2.3).

- Le consentement des utilisateurs de l'application sera également obtenu pour la collecte et l'utilisation des informations suivantes, comme condition d'utilisation du service, au moyen d'un langage de consentement et d'une politique de confidentialité :
  - L'âge (indiqué par l'utilisateur)
  - Le sexe (indiqué par l'utilisateur)
  - Les conditions médicales (signalées par l'utilisateur)
  - Les symptômes actifs (signalés par l'utilisateur)
  - Le comportement pertinent en cours (signalé par l'utilisateur)
  - *La localisation géographique approximative (mesurée par le GPS)*
  - *Les statistiques sur les mouvements (mesurées par le GPS)*
  - Les informations analytiques (utilisation des fonctionnalités de l'application qui ne révèlent aucune information sensible au sujet de l'utilisateur)

Toutes ces données, à l'exception des informations analytiques, seront introduites dans la fonction d'évaluation des risques des applications (avec le niveau de risque actuel). L'évaluation des risques sera effectuée localement sur l'application installée sur le dispositif de l'utilisateur. Aucune de ces informations ne quittera l'appareil à moins que l'utilisateur ne choisisse de les envoyer à COVI Canada afin d'entraîner le modèle d'apprentissage automatique et contribuer (sous forme agrégée) à la recherche épidémiologique du gouvernement ou d'autres tiers. Les informations analytiques, qui permettent à COVI Canada de suivre certaines actions telles que l'achèvement de l'installation et la mise en place de l'application par les utilisateurs, seront envoyées à COVI Canada sous forme pseudonymisée. Nous considérons que cette collecte est nécessaire pour le fonctionnement de l'application COVI, car il est essentiel que l'application soit largement adoptée afin de maximiser son efficacité par rapport à son objectif principal. Si un pourcentage important d'utilisateurs ne termine pas l'installation de l'application, empêchant ainsi l'application de fonctionner pour propager les niveaux de risque, le fait de savoir cela nous permettra de changer de stratégie de messagerie, d'interface utilisateur ou d'expérience utilisateur afin d'encourager l'achèvement du processus d'installation. Bien que les informations ainsi recueillies quittent l'appareil, il ne s'agit pas d'une divulgation, car elles ne sont fournies qu'à COVI Canada, qui est l'organisation responsable de l'application et de son écosystème d'information. Les informations sont également de nature manifestement non sensible et ne sont associées à aucune information qui permettrait à un tiers de remonter jusqu'à l'appareil d'origine.

2.1.2 Consentement à l'utilisation du résultat officiel d'un test positif

À la réception d'un diagnostic officiel, les utilisateurs de COVI peuvent, de manière facultative, consentir à ce que COVI Canada utilise un résultat de test positif officiel pour actualiser le niveau de risque de l'utilisateur, qui peut alors être utilisé pour améliorer l'estimation du niveau de risque des autres utilisateurs sur leurs appareils. La saisie du diagnostic officiel nécessite une étape d'authentification impliquant une interaction avec les bases de données des autorités de santé publique qui contiennent les informations de diagnostic officiel ; étant donné les exigences juridiques particulières de chaque juridiction, une approche uniforme peut ne pas être possible. En tout état de cause, à l'issue de cette étape d'authentification, il sera expressément demandé à l'utilisateur de consentir à la collecte, à l'utilisation et à la divulgation du résultat officiel du test positif. Si le consentement est accordé, l'application comptabilisera le résultat officiel et l'utilisera pour calculer un nouveau niveau de risque. Le niveau de risque actualisé sera ensuite communiqué aux contacts récents, pour être pris en compte dans l'évaluation des risques de chaque contact. Bien que le résultat officiel lui-même ne soit pas partagé, car l'effet du résultat positif sur le calcul du niveau de risque sera de le conduire vers sa valeur maximale possible, nous traitons le niveau de risque actualisé comme un indicateur de facto du statut d'infection positive et donc la communication de cette information comme une divulgation du résultat. Les personnes contactées ne seront pas informées du nom de leurs contacts ayant reçu un résultat positif à un test de dépistage. L'application COVI de chaque contact calculera le nouveau niveau de risque de ce contact, et les utilisateurs ciblés ne recevront qu'un ensemble de recommandations personnalisées basées sur cette nouvelle évaluation interne du niveau de risque.

2.1.3 Consentement à l'utilisation des données à des fins de recherche

Lors de l'installation ou à tout moment par la suite, les utilisateurs peuvent consentir, sur la base d'option d'adhésion (et peuvent ensuite se désister, à tout moment), à envoyer certaines informations à intervalles réguliers au serveur d'apprentissage automatique de COVI afin d'entraîner les algorithmes d'apprentissage automatique et les modèles épidémiologiques ainsi que pour partager des données agrégées avec le gouvernement et d'autres tiers. Il sera demandé aux utilisateurs de consentir à la collecte et à l'utilisation des informations suivantes, qui seront envoyées aux chercheurs de l'application COVI sous forme pseudonymisées :

- L'âge (indiqué par l'utilisateur)
- Le sexe (indiqué par l'utilisateur)
- Les symptômes actifs (signalés par l'utilisateur)
- Le comportement pertinent en cours (signalé par l'utilisateur)
- *La localisation géographique approximative (mesurée par le GPS)*
- *Les statistiques sur les mouvements (mesurées par le GPS)*
- Statut d'infection positive certifiée (en cas de saisie, conformément au consentement fourni)
- Les informations analytiques (utilisation des fonctionnalités de l'application qui ne révèlent aucune information sensible au sujet de l'utilisateur)

Ces données, outre les informations analytiques, seront utilisées pour améliorer la prévision des risques et les modèles épidémiologiques. Elles serviront également de base à la production de données agrégées au niveau de la population qui seront partagées avec les acteurs gouvernementaux et d'autres tiers, uniquement à des fins liées aux efforts visant à comprendre ou à combattre la COVID-19. Les informations analytiques permettront à

COVI Canada d'évaluer des questions telles que l'efficacité des recommandations concernant la réduction des niveaux de risque.

Toutes les données pseudonymisées resteront sur les serveurs de COVI et tout traitement de ces données sous forme agrégées aura lieu avant que les données cumulées ne soient fournies au gouvernement ou à d'autres tiers.

Il est important de noter, en ce qui concerne le consentement demandé à cette étape, que nous considérons que l'entraînement du modèle d'apprentissage automatique est d'une importance fondamentale à l'objectif principal de l'application, même si nous donnons aux utilisateurs le pouvoir d'y adhérer ou non. Bien que le modèle d'apprentissage automatique soit initialement formé avec des données synthétiques à un degré qui fournit un prédicteur de risque modérément efficace, le modèle nécessite une formation sur des données réelles afin de produire un prédicteur de risque avec un niveau de précision nécessaire pour réduire la propagation de la COVID-19. En règle générale, les lois sur la protection de la vie privée autorisent la collecte des informations nécessaires à la réalisation d'une fonction essentielle d'un service comme une condition de service plutôt que comme une option. En rendant la collecte et l'utilisation de ces informations nécessaires facultatives, le modèle de consentement de COVI va bien au-delà de ce qu'exigent les lois sur la protection de la vie privée. Bien que l'on puisse dire qu'il est contre-intuitif, notre raisonnement s'explique facilement. Si les informations collectées dans le but de former le modèle d'apprentissage automatique sont nécessaires, il est seulement nécessaire qu'une certaine quantité de ces informations soit collectée. Pour un utilisateur donné, il n'est pas nécessaire que des informations soient recueillies auprès de cet utilisateur. Ainsi, en reconnaissance de la nature axée sur la protection de la vie privée de l'application, l'application permet aux personnes de choisir de permettre à COVI Canada d'accéder à ces informations et de les utiliser malgré leur nécessité. Les données analytiques potentiellement sensibles obtenues sont ainsi soumises à ce consentement pour des raisons similaires. Aussi nécessaire soit-il de fournir les meilleures recommandations possibles, nous n'avons pas besoin de suivre les corrélations entre les recommandations faites et les niveaux de risque dans le temps pour chaque utilisateur afin d'évaluer l'efficacité des recommandations faites.

Naturellement, nous espérons que de nombreux Canadiens choisiront de nous fournir volontairement leurs données afin que nous puissions intégrer à l'application une meilleure prévision des risques et de meilleures recommandations, et ainsi créer de meilleurs modèles épidémiologiques pour orienter les politiques publiques. Qu'une personne donnée décide de le faire ou non, les fonctions essentielles de l'application fonctionneront néanmoins pour tout le monde. Nous acceptons ceux qui ne participent pas dans l'espoir que nous aurons une masse critique suffisante de volontaires fournissant des données pour former le modèle d'apprentissage automatique.

Les données collectées seront conservées dans l'entité indépendante mentionnée plus haut dans cette section, et seront détruites dès qu'elles ne seront plus nécessaires. L'utilisateur peut révoquer son consentement à tout moment, après quoi ses données seront supprimées du serveur. S'il ne révoque pas son consentement, ses données seront toujours automatiquement expirées après une période de 90 jours au maximum. Après une période fixe au cours de laquelle aucun nouveau cas n'est survenu, toutes les données restant dans l'ensemble de données non agrégées seront supprimées.

2.2 Limites inhérentes à la protection de la vie privée dans le cadre de la décentralisation de la fonction automatique de traçage des contacts

Bien qu'il existe de nombreux moyens technologiques et cryptographiques pour protéger les informations en transit et en attente, un système automatisé de traçage des contacts est par nature un système de suivi, même si sa portée est limitée. Comme le système doit informer les utilisateurs exposés qu'ils ont été exposés à une personne chez qui l'on a diagnostiqué la COVID-19, le système divulgue des informations sur l'identité des

utilisateurs diagnostiqués. Il existe des risques endémiques pour la vie privée qui ne peuvent être éliminés par des moyens technologiques. Nous pensons qu'il est primordial de reconnaître et d'analyser ces risques inhérents, afin de permettre aux utilisateurs et au gouvernement de prendre des décisions éclairées sur l'ampleur des atteintes à la vie privée qu'ils sont prêts à accepter dans le cadre de la lutte contre la pandémie.

Avant d'entrer dans les détails de notre proposition, nous allons d'abord analyser certains des risques systémiques liés au traçage automatique des contacts décentralisé en envisageant un système abstrait présentant les propriétés souhaitables suivantes :

1. Le traçage des contacts se fait par l'intermédiaire d'une application pour téléphones intelligents, de sorte que lorsque deux téléphones se trouvent à moins de 2 mètres l'un de l'autre, un contact est enregistré.

2. Lorsqu'un utilisateur (Bob) reçoit un diagnostic de COVID-19 (ou voit son niveau de risque augmenter), tous ses contacts (que nous appellerons des Alices) des 14 derniers jours sont informés du fait suivant : le jour X, Alice était à proximité immédiate d'un individu infecté.

Même si Alice elle-même n'est pas directement informée du jour par l'application - par exemple, l'application lui dit seulement qu'elle doit se mettre en quarantaine - cela équivaut, du point de vue de la sécurité, à ce que le téléphone soit informé du jour, puisqu'une application malveillante pourrait extraire l'information. Nous les traitons donc de la même manière dans l'analyse de la protection de la vie privée.

Il existe deux différences principales entre ce modèle de traçage automatique des contacts décentralisé et un modèle plus traditionnel de traçage manuel ou centralisé des contacts :

1. Avec le traçage automatique des contacts décentralisé, si plusieurs utilisateurs diagnostiqués sont en contact avec Alice, celle-ci recevra une notification d'exposition pour chaque individu. Dans le cadre du traçage des contacts standard, Alice ne peut recevoir qu'une seule notification, même si elle a été exposée à plusieurs personnes.

2. Puisqu'il est décentralisé, il est difficile d'empêcher un adversaire d'acquérir des identités multiples lors d'une attaque de type Sybil, alors que dans un modèle plus traditionnel, il peut y avoir des possibilités d'atténuation. Comme le traçage automatique des contacts se fait par l'intermédiaire des téléphones intelligents, un adversaire possédant plusieurs téléphones intelligents peut être en mesure d'acquérir plusieurs identités.

Ces différences permettent une série d'attaques sur les renseignements privés de Bob, qui est un utilisateur envoyant des notifications d'exposition/messages de risque à ses contacts. Pour simplifier la discussion, nous allons décrire les attaques ci-dessous dans le cadre de la notification d'exposition binaire la plus simple pour la COVID-19, mais la plupart des attaques s'appliquent à toute entité qui envoie des messages à toute personne qui s'est trouvée à proximité d'elle. Cela est vrai que le message contient une valeur de risque de transmission pour la COVID-19 ou une liste de symptômes autodéclarés. Nous pensons qu'il est important de reconnaître ces risques comme une base de référence avant de se lancer dans la discussion sur la structure de COVI.

2.2.1 Atteintes à la confidentialité des données médicales

L'une des fuites inhérentes à la protection de la vie privée dans le cadre du traçage des contacts est qu'elle est dérivée de l'historique des lieux où se trouve Bob. Un attaquant qui dispose de suffisamment d'informations sur l'historique de localisation de Bob peut effectuer une attaque de type « linkage » pour connaître l'état de santé de Bob. Heureusement, l'historique de localisation de Bob n'a pas besoin d'être diffusé en soi, mais même dans ce cas, l'envoi d'informations aux contacts de Bob révèle implicitement où se trouvait Bob. Ainsi, les entreprises

qui ont accès ne serait-ce qu'à une partie de l'historique de localisation de Bob peuvent avoir accès à son statut et son diagnostic. Les gérants d'un hôtel en sont un exemple : nous affirmons que l'hôtel peut déterminer le statut de tous ses clients dont il sait qu'ils utilisent l'application.

Commençons par la version la plus simple de l'attaque. L'hôtel a placé un téléphone dans chaque chambre d'hôtel utilisant l'application de traçage des contacts. Si un client, Bob, séjourne dans la chambre 100 le 1er juin, et qu'il a ensuite un diagnostic de COVID-19, alors le téléphone de sa chambre indiquera qu'il y avait un individu infecté dans cette chambre le 1er juin. Comme l'hôtel connaît le registre des clients, il est trivialement en mesure de déterminer que Bob a été diagnostiqué avec la COVID-19, violant ainsi la confidentialité de ses renseignements médicaux.

Bien sûr, cette version très simple de l'attaque peut être partiellement contrecarrée en interdisant à l'hôtel d'avoir 1 000 téléphones. Si vous validez chaque copie de l'application pour que seules de vraies personnes puissent les posséder, alors vous empêchez la version la plus simple de l'attaque, car l'hôtel ne peut pas acquérir 1 000 identités. La validation de l'installation de l'application soulève bien sûr d'autres problèmes de protection de la vie privée, mais ceux-ci peuvent être résolus par d'autres moyens.

Toutefois, si la simple attaque consistant à avoir une copie de l'application dans chaque chambre peut être bloquée par l'enregistrement de l'utilisateur, cela ne bloque pas une version légèrement plus sophistiquée de l'attaque. Supposons que l'hôtel dispose de 1 000 chambres et que seuls 10 téléphones exécutent l'application, ce qui est trivial - par exemple, 10 employés utilisent l'application. Ensuite, la nuit, lorsque tous les clients sont au lit, chaque employé passe devant la moitié de leur porte et n'allume le téléphone qu'à la bonne porte.

Il s'agit en fait d'un codage 10b pour chaque pièce, l'identifiant par l'ensemble d'employés qui sont passés devant leur chambre. Si les employés 1, 3 et 5 passaient devant la chambre de Bob, son code serait alors 1010100000. Comme un codage 10b offre 2_{10} = 1024 possibilités, chaque pièce peut recevoir son propre code unique avec le bon échantillon d'employés passant devant. Plus tard, si les employés 1, 3 et 5 reçoivent des messages indiquant que le 1er juin, ils ont été en contact avec un utilisateur infecté, et qu'aucun des autres employés ne reçoit ce message, alors l'hôtel sait immédiatement que c'est Bob qui a été diagnostiqué.

Bien que cela puisse sembler difficile à coordonner d'un point de vue logistique, il est futile de faire une simulation en plaçant un appareil dans chaque chambre pour simuler le passage des employés selon des schémas spécifiques. Tout ce dont un hôtel a besoin, c'est d'avoir accès à 10 comptes réels, et avec cela, il est facile d'activer et de désactiver les appareils et les identités pour identifier chaque chambre. Ces appareils n'exécutent plus l'application normalement, mais ils simulent le comportement d'une personne réelle passant devant les chambres selon un schéma étrange, et cette attaque ne peut donc pas être facilement détectée/arrêtée par le système de traçage des contacts.

Une autre attaque mathématiquement équivalente est celle appelée attaque « vigilante », dans laquelle un attaquant cherche à révéler qu'un individu qu'il a rencontré est infecté. L'une des motivations peut être que l'attaquant (par exemple, Mallory) veut se venger de Bob pour l'avoir exposée à la COVID-19. Cela est, à bien des égards, mathématiquement équivalent à l'attaque de l'hôtel, mais la différence est que Mallory ne connaît pas l'emplacement de Bob pendant une période déterminée, comme le fait l'hôtel.

En revanche, Mallory connaît son propre historique de localisation et sait quand elle a rencontré d'autres personnes à l'extérieur. Si Mallory peut déterminer la durée d'exposition à la COVID-19 dans une période de 5 minutes, elle peut raisonnablement deviner quand et où elle a rencontré une personne infectée. Si, pendant cette période, la seule personne avec laquelle elle a été proximité était Bob, elle apprend (1) que Bob a été diagnostiqué et (2) que Bob était la source de son exposition, deux fuites d'informations essentielles.

Encore une fois, Mallory pourrait allumer un téléphone séparé pour chaque période de 5 minutes de la journée, mais il existe aussi une version logarithmique de l'attaque. Comme il y a 1 440 minutes dans une journée, il n'y a que 288 périodes de 5 minutes. En utilisant 9 téléphones, Mallory peut de la même manière attribuer un code binaire de 9 bits à chaque période de 5 minutes, et selon les téléphones qui reçoivent la notification d'exposition, Mallory saura quand elle a été exposée. Comme pour l'attaque de l'hôtel, avec un peu d'expertise technique et un accès à 9 identités, Mallory peut composer une application sur un seul téléphone qui prétend être 9 téléphones avec l'application installée.

Dans ce qui précède, nous avons montré qu'il suffit d'un nombre logarithmique d'identités pour qu'un attaquant puisse révéler des informations sur ses contacts ou ses clients. En pratique, de nombreux protocoles proposés de traçage des contacts décentralisé n'en requièrent même pas autant, car ils ne nécessitent pas de validation forte de la part de l'utilisateur lorsque celui-ci tente de déterminer son propre statut d'exposition. Par exemple, dans plusieurs des modèles décentralisés que nous examinerons plus loin, tous les dépistages des contacts se font localement sur le téléphone. Cela est extrêmement performant pour protéger la vie privée des utilisateurs qui n'envoient pas de messages - puisqu'ils ne transmettent aucune information à partir de leur téléphone - mais cela signifie également qu'il n'existe aucun moyen simple d'empêcher un attaquant de manipuler plusieurs fois une interaction avec des contacts.

2.2.2 Attaques ciblant l'historique de localisation de l'utilisateur

La sous-section précédente traitait de la fuite de données médicales, à savoir le statut de diagnostic d'un utilisateur. Cependant, il y a également les fuites des schémas de déplacement des utilisateurs. En particulier, tout utilisateur qui transmet des informations sur son statut de contagion (par exemple, Bob) transmet aussi implicitement des informations sur ses déplacements antérieurs. Cela est bien sûr nécessaire pour qu'Alice puisse établir un contact. Par exemple, dans les systèmes basés sur la technologie Bluetooth, Alice enregistre que Bob diffuse une annonce Bluetooth ; ces annonces sont souvent aléatoires ou pseudo-aléatoires, de sorte qu'elles ne peuvent être mises en relation sans la coopération de Bob. Cependant, une fois que Bob envoie des notifications à tous ses contacts précédents, des informations sur ses déplacements sont au moins transmises à Alice.

Pour simplifier la discussion, nous allons examiner le cas où des incidents d'exposition à la COVID-19 sont peu fréquents. Cela peut sembler une hypothèse étrange en pleine pandémie, mais c'est une hypothèse raisonnable pour le traçage des contacts, car si la plupart des utilisateurs subissent des incidents d'exposition, alors il y a peu de signaux pour informer les utilisateurs qu'ils ont été exposés. Malheureusement, dans ce cas, le temps d'envoi des notifications d'exposition entraîne également la fuite d'un grand nombre d'informations.

Prenons encore une fois le cas de Mallory, qui veut révéler des informations sur Bob. Nous avons vu plus haut qu'avec un nombre logarithmique d'identités, Mallory peut révéler l'heure et le lieu où elle a rencontré Bob. Si cette heure et ce lieu se trouvent dans une aire publique (par exemple dans les transports en commun), Mallory ne pourra peut-être pas identifier exactement qui est Bob. Mais supposons maintenant que Mallory rencontre Bob à plusieurs reprises. Normalement, Mallory ne sait pas nécessairement que ses nombreuses rencontres avec Bob se font avec la même personne. Cependant, une fois que Bob a été diagnostiqué avec la COVID-19, Mallory reçoit une notification pour chacune de ses rencontres avec Bob, dont elle connaît l'heure et le lieu. Comme les expositions sont rares, Mallory est en mesure de déduire que toutes ses notifications d'exposition concernaient probablement la même personne. Ainsi, Mallory est en mesure d'établir un registre partiel des mouvements de Bob.

Notez que cette attaque n'est pas liée à des diagnostics médicaux, mais qu'elle est activée simplement parce que Bob envoie une notification pour chaque heure et pour chaque lieu où il s'est rendu, ce qui constitue la base du traçage des contacts. D'une certaine manière, cette attaque a une portée limitée, car Mallory aurait pu simplement se souvenir de ses rencontres avec Bob d'autres manières (par exemple en filmant puis ensuite, à

l'aide de la technologie, en effectuant une reconnaissance faciale, ou avec des téléphones intelligents plus anciens, les adresses MAC WiFi étaient également traçables). Toutefois, il s'agit toujours d'une fuite du protocole.

Le danger de l'attaque temporelle préméditée peut être amplifié si une grande institution est l'adversaire ; appelons-la Grace. Supposons que Grace place des appareils exécutant le protocole de traçage des contacts et rien d'autre, à de nombreux endroits dans une ville. En utilisant l'attaque préméditée, Grace est alors en mesure d'établir une corrélation entre l'historique de localisation où se trouvent de nombreuses personnes diagnostiquées.

Le signal que Grace reçoit contient un bruit supplémentaire, car avec suffisamment d'appareils, l'hypothèse d'un nombre relativement faible d'incidents d'exposition ne suffit plus pour regrouper les historiques de localisation - de multiples utilisateurs signaleront des actions de diagnostic à peu près au même moment. Cependant, les historiques de localisation sont contigus dans l'espace, et si Grace a suffisamment d'appareils placés autour d'elle, la reconstruction des trajectoires est tout à fait possible. Grace pourrait potentiellement viser un acteur gouvernemental ou une grande entreprise, car ils auraient les moyens de déployer des appareils ou dispositifs sur une vaste zone géographique. Bien entendu, cette fuite d'informations doit être replacée dans son contexte, car un acteur gouvernemental important a déjà accès à de nombreuses autres sources d'informations de suivi, telles que les pings des tours de téléphonie mobile ou les flux de systèmes de télévision en circuit fermé (TVCF). Par conséquent, ce qui empêche réellement ces attaques, c'est le cadre juridique et les normes sociales du pays, ainsi que la force politique de l'opinion publique et des médias. La transparence dans la gestion de COVI est donc d'une importance primordiale pour minimiser ces attaques.

2.2.3 Limitation des risques

Malheureusement, les deux attaques mentionnées ci-dessus fonctionnent pour tous les protocoles avec les propriétés données de traçage des contacts automatique. Bien qu'il existe des solutions technologiques qui peuvent rendre l'exécution plus difficile ou plus irritante pour un attaquant, aucune d'entre elles ne peut réellement arrêter les attaques. Ceci est bien sûr dû au fait que l'identification des contacts avec la COVID-19 est la raison principale pour l'utilisation du traçage des contacts. Dans le monde réel, tout système décentralisé de traçage des contacts comporte en outre des risques supplémentaires dus à la conception du système.

Un défenseur de la protection de la vie privée considérerait raisonnablement ces attaques comme une raison de ne pas utiliser de système décentralisé de traçage automatique des contacts. Cependant, même les partisans de la protection de la vie privée peuvent s'inquiéter de ces compromis, et nous pensons donc qu'il est important de le reconnaître directement, afin que les utilisateurs et le gouvernement puissent trouver un équilibre entre les bénéfices du traçage des contacts pour la santé publique et la quantité de données qui sont exposés. De plus, bien que les solutions technologiques puissent être limitées, nous pensons que des protections juridiques et économiques peuvent être mises en place.

2.3 Choix du protocole pour les messages de risque privés

Bien que, comme nous l'avons vu dans la section précédente, certains risques soient inhérents au traçage des contacts et ne peuvent être éliminés, cela ne nous dispense évidemment pas de la responsabilité d'assurer une protection maximale de la vie privée, tout en réalisant les objectifs du système. Afin de garantir la protection de la vie privée pour les fonctionnalités de base, nous devons mettre en place un système de messagerie privée [67, 68, 69] qui garantit qu'aucune information sur le niveau de risque ou l'historique des contacts d'une personne n'est révélée aux autorités ou aux autres utilisateurs, sauf ce qui est absolument nécessaire pour échanger des messages de risque. Avec l'aide d'un certain nombre d'auditeurs et d'examinateurs externes, nous évaluons

actuellement trois systèmes différents pour le système de messagerie privée, chacun ayant ses propres avantages et ses limites :

1. La structure de notification d'exposition de Google et Apple (GAEN) [13]

2. Le protocole de la coalition TCN [45]

3. Le protocole du NHS utilisant la technologie Bluetooth [51] + un réseau mixte pour l'échange de messages [67]

Les trois systèmes ont leurs avantages et leurs inconvénients. Dans ce livre blanc, nous n'entrons pas dans les détails de la conception au niveau des systèmes, mais nous évaluons à un niveau supérieur les façons dont chacun de ces systèmes de messagerie pourrait être appliqué dans notre conception. Notre développement initial utilise l'option (3) le protocole de la NHS avec la technologie Bluetooth + un réseau mixte, car il existe des limites techniques et pratiques à la structure de notification d'exposition de Google et Apple et au protocole TCN.

Dans la suite de cette section, nous ferons référence aux différents acteurs par leur nom. Nous avons déjà rencontré plusieurs de ces personnages dans la section précédente, mais voici un bref rappel.

Personnages

- Alice, une utilisatrice de l'application. Elle rencontre Bob le jour j.

- Bob, un utilisateur de l'application. Il rencontre Alice le jour j. Plus tard, avec de nouvelles informations, le niveau de contagion estimée de Bob pour le jour j change (communiqué par le biais du niveau de risque), et il veut communiquer ce changement à Alice en privé.

- Grace, le gouvernement (ou une autre autorité centrale). Elle gère le serveur de messagerie central contenant tous les rapports.

- Eve, une espionne passive, qui tente d'obtenir des informations en écoutant la communication, mais ne fait rien d'actif.

- Mallory, un agent malveillant, qui tente de briser le système, et qui essaiera d'envoyer de fausses informations aux serveurs et à d'autres entités. Tout utilisateur ou entité malveillante peut être ou devenir Mallory.

2.3.1 Les API de notification d'exposition de Google et Apple

L'API de notification d'exposition de Google et Apple [13] présente l'avantage considérable d'être directement prise en charge par les fabricants de téléphones intelligents, qui ont un accès de niveau inférieur à Bluetooth, comme aucune autre solution ne peut le faire. Sans trop entrer dans les détails, les téléphones d'Alice et de Bob diffusent en permanence des codes RPI (Rolling Proximity Identifiers – RPIs) via Bluetooth. Ces codes sont dérivés de clés de diagnostic quotidiennes, permettant une régénération ultérieure.

Lorsqu'Alice rencontre Bob, le téléphone d'Alice stocke les codes RPI qu'elle reçoit de Bob. Plus tard, si Bob souhaite envoyer un message de risque à Alice, il envoie les clés de diagnostic à Grace avec un message de risque de transmission quantifié en pièce jointe. Alice télécharge périodiquement toutes les clés de diagnostic de Grace dans une zone géographique donnée, puis régénère localement les codes RPI ; chaque fois qu'elle régénère un code RPI qu'elle a reçie, elle sait que le message lui est destiné, l'informant qu'elle a été exposée à Bob.

Toutefois, pour éviter toute utilisation abusive de son système, l'API de notification d'exposition de Google et Apple impose des limites strictes aux applications participantes, ce qui rend difficile certains types de collecte de données et de cas d'utilisation. Elles n'autorisent pas actuellement le consentement préalable pour le partage des clés de diagnostic, dont nous avons besoin pour une propagation rapide des risques sur le réseau. De plus, elles ne permettent pas aux téléphones d'accéder aux services de localisation ; bien que les applications puissent demander aux utilisateurs leur localisation manuellement, cela limite les types de données épidémiologiques qui peuvent être envoyées à une autorité de santé publique. Heureusement, l'API de notification d'exposition de Google et Apple est en cours d'élaboration, et nous sommes en discussion avec eux pour trouver des solutions à ces limitations.

2.3.2 Coalition TCN

Le protocole de la coalition TCN est une alternative à l'utilisation de la structure officiellement appuyée par Google et Apple [45]. L'absence de soutien de la part de Google et d'Apple impose certaines limites techniques aux capacités de l'infrastructure. Notamment, la communication Bluetooth sur le système iOS est sévèrement limitée. Cependant, l'approche TCN n'a pas les limitations de consentement et de localisation de l'API de notification d'exposition de Google et Apple.

Le protocole TCN est ainsi nommé, car les utilisateurs diffusent entre eux des codes de contacts temporaires, à peu près équivalents aux codes RPI de Google et Apple, via la technologie Bluetooth si les propriétaires des téléphones ont été à proximité les uns des autres [45]. Par exemple, Alice a quatre des codes (TCN) de Bob, et chaque période est de 5 minutes, puis ils ont passé environ 20 minutes dans la limite de distance établie par le protocole. Les téléphones peuvent publier un « rapport » sur un serveur central, en associant un ensemble codes temporaires à une charge de niveau de risque (en utilisant le champ de mémo du rapport TCN). Les rapports TCN permettent au récepteur de régénérer un ensemble de codes temporaires que Bob diffuse sur une période de temps spécifique à l'application (par exemple 6 heures).

Comme pour l'API de notification d'exposition de Google et Apple, les utilisateurs téléchargeront tous les nouveaux rapports d'une zone géographique. Les rapports TCN permettent à l'utilisateur de régénérer les codes temporaires qui ont été diffusés à l'origine, et les utilisateurs peuvent vérifier localement ces codes par rapport à leur journal interne de codes reçus.

À intervalles réguliers, les utilisateurs vérifient les rapports correspondants et mettent à jour leur propre niveau de risque en fonction des informations qu'ils ont reçues. Lorsque le niveau de risque d'un utilisateur change considérablement, l'application envoie ces informations dans toutes les messageries électroniques pertinentes associées à ses contacts des deux dernières semaines. Les utilisateurs peuvent également mettre à jour leur niveau de risque en fonction d'autres informations, telles que les symptômes autodéclarés ou des diagnostics de la COVID-19. Leurs contacts peuvent alors vérifier régulièrement les niveaux de risque associés aux codes que le téléphone a enregistrés.

2.3.3 Le protocole du NHS avec la technologie Bluetooth + réseaux mixtes

L'un des principaux inconvénients de l'API de notification d'exposition de Google et Apple et du protocole TCN est la nécessité de télécharger tous les rapports/clés de diagnostic dans une vaste zone géographique pour qu'Alice puisse déterminer une interaction localement sur son téléphone. Cette approche a l'avantage de révéler très peu d'informations sur Alice lorsqu'elle récupère des messages, mais si les messages de risque sont nombreux, elle nécessite également une bande passante importante. En outre, la plateforme Bluetooth TCN n'a pas été testée de manière aussi approfondie que certains autres systèmes, c'est pourquoi notre déploiement initial utilise donc une troisième option.

Le système de santé publique du Royaume-Uni (National Health Service) a conçu et mis en place une application de traçage des contacts en utilisant la technologie Bluetooth [51] et le code a été validé par des essais sur le terrain. Au lieu de diffuser simplement une annonce BLE, ils créent un canal de communication entre des paires de téléphones, ce qui permet d'envoyer des messages plus longs qu'une annonce BLE. Les messages Bluetooth entre les téléphones contiennent des identifiants chiffrés pour permettre un système de traçage des contacts plus centralisé que celui que nous prévoyons pour COVI. Cependant, leurs bibliothèques et leurs codes permettent d'envoyer des messages Bluetooth pour également soutenir un système plus décentralisé (en modifiant simplement les messages envoyés).

Lorsqu'Alice et Bob se retrouvent à proximité, nous utilisons l'échange de clés Diffie-Hellman [70] via Bluetooth pour générer deux clés chiffrées, l'une pour les messages d'Alice à Bob, l'autre pour les messages de Bob à Alice. L'utilisation de l'échange de clés de Diffie-Hellman empêche Eve ou Mallory de falsifier les messages d'un utilisateur, bien que Mallory puisse bien sûr agir en tant qu'utilisatrice elle-même. Un jeton d'authentification peut ensuite être utilisé pour obtenir une « adresse » et une clé de cryptage grâce à une fonction de hachage unidirectionnelle.

Lorsque Bob envoie son statut de risque à Grace, il envoie des messages de risque chiffrés à la bonne adresse. Alice ne peut vérifier que les adresses où elle s'attend à recevoir ses messages. Ces messages contiennent exactement les mêmes informations sur le risque de transmission que celles qui seraient jointes aux clés de diagnostic de Google et Apple ou aux rapports du protocole TCN.

Le fait qu'Alice se connecte implique un risque d'attaque : Grace peut voir que Bob a envoyé un message qu'Alice a vérifié, et en déduire qu'Alice et Bob se trouvaient à proximité l'un de l'autre. Ainsi, la récupération directe des messages (au lieu de télécharger un lot important) nécessite des couches supplémentaires de confidentialité des communications, que nous obtenons en utilisant des réseaux mixtes (section 2.3.4.)

2.3.4 Dissimuler les schémas d'envoi et de récupération

Malheureusement, le schéma de récupération de la messagerie peut révéler des métadonnées sensibles même si les adresses et les messages eux-mêmes sont cryptés [71]. Par exemple, si l'adresse IP d'Alice est vue en train de vérifier un message envoyé par l'adresse IP de Bob, alors Grace sait qu'il y a eu un contact entre les deux, ce qui lui permet de déduire le graphe social.

Ainsi, nous devons soit cacher les schémas de récupération, soit les schémas d'envoi. Il y a plusieurs façons de cacher les schémas de récupération. Le mécanisme par défaut dans l'API de notification d'exposition de Google et Apple et le protocole TCN consiste à fragmenter la base de données géographiquement, puis Alice télécharge l'ensemble des nouveaux messages dans sa région géographique. L'API Google-Apple et le protocole TCN ne protègent pas l'envoi du message de Bob par les autorités, qui révèle une empreinte Bluetooth, mais comme Alice télécharge la base de données entière, il n'y a aucun moyen de relier les individus entre eux dans une attaque visant le schéma social.

D'autres propositions [72, 15] utilisent plutôt les protocoles Private Set Intersection Cardinality (PSI-CA), qui permettent à Alice d'interroger le serveur, sans révéler ses informations et sans connaître la base de données de Grace. Malheureusement, ces protocoles ont tendance à être très coûteux en matière de calcul et/ou de bande passante, et ne sont donc souvent pas réalisables en pratique pour une mise en place à l'échelle d'un pays entier. En outre, ces protocoles ne fonctionnent pas facilement avec une charge utile de message en plus de la détection des points d'intersection.

Bien que les protocoles utilisés par Google-Apple et TCN dissimulent les schémas de récupération en téléchargeant une base de données fragmentée, cela entraîne un coût important en matière de bande passante. Il faut rappeler que pour se défendre contre les attaques, il faut soit cacher les schémas d'envoi, soit les schémas

de récupération. Dans l'approche NHS + réseaux mixtes, nous utilisons un réseau mixte [73] pour dissimuler les schémas d'envoi à la place.

Nous utilisons le routage en oignon (onion routing) pour coder chacun des messages de risque de Bob [74], qui sont ensuite décryptés par étape par les différents serveurs de réseaux mixtes, qui mélangent également les messages de Bob avec ceux d'autres personnes avant de les transmettre au prochain serveur du réseau, et enfin de remettre les messages à Grace.

Conception des réseaux mixtes Chacun des serveurs mixtes 1...N publie une clé publique $p_1 ... p_N$. Pour les besoins de cette discussion, nous considérons que Grace contrôle le dernier serveur mixte N. Pour chaque message (x,m), qui comprend à la fois une adresse x et un message chiffré m, Bob envoie $p_1(p_2(... p_{N-1}(p_N((x,m)))...))$ au premier serveur mixte, où pi est le chiffrement par la clé publique pi. Le premier serveur mixte supprime le premier niveau de chiffrement, obtenant $p_2(...p_{N-1}(p_N((x,m)))...))$. Le premier serveur mixte attend de recevoir les rapports chiffrés de plusieurs Bob, les regroupe tous et les mélange, en intégrant les messages de différents Bob, puis les transmet par lots au second serveur mixte. Le second serveur mixte fait la même chose. À la fin du protocole, le dernier serveur mixte (contrôlé par Grace) reçoit une série de messages dans le format (x,m), qui ont été dissociés de Bob.

Alice peut alors vérifier directement tous les messages à une adresse x ; bien que Grace apprenne l'ensemble des adresses auxquelles Alice se connecte, ces adresses ne sont pas manifestement liées à Bob, ce qui empêche Grace de faire le lien entre Alice et Bob. Tant que l'un des serveurs mixtes est intègre et qu'il n'y a pas d'attaque active avec des données malformées, les messages de Bob ont été mélangés avec ceux d'autres personnes, et sont donc dissociés de son identité.

Attaque (message de suivi) Si le premier serveur mixte est de connivence avec Grace dans une attaque active, il peut rejeter tous les messages sauf ceux de Bob, en les remplaçant par des messages bidon. De cette manière, ils peuvent déterminer avec quelles messageries Bob est en communication. Afin de répondre correctement à cette attaque, chaque serveur du réseau mixte doit introduire des données bruitées appropriées pour cacher l'identité de Bob [67]. Pour notre mise en œuvre initiale, nous n'incluons pas ce bruit.

Cependant, nous constatons qu'il est possible de détecter cette attaque en nous envoyant des messages par le biais du réseau mixte. S'ils sont rejetés, nous saurons alors que cette attaque est en cours et nous pourrons alors prendre les mesures appropriées. Par ailleurs, ce type de canari peut être mis en œuvre par n'importe quel utilisateur ; des tiers indépendants peuvent vérifier que le premier réseau mixte n'effectue pas une attaque de type « tracking message ».

2.4 Données à option d'adhésion (*opt-in*) pour affiner le modèle d'apprentissage automatique et agrégation de données pour les gouvernements

Les utilisateurs pourront choisir d'envoyer des données pseudonymisées aux serveurs d'apprentissage automatique de COVI afin d'entraîner le modèle d'apprentissage automatique qui détermine les « niveaux de risque » en fonction des contacts passés, des symptômes et des informations démographiques. Si un utilisateur y consent, les informations suivantes sont envoyées à intervalles réguliers (environ tous les jours) au serveur de COVI : un paquet de données pseudonymisées (pour l'apprentissage automatique) et des informations de carte thermique (agrégation spatiale).

2.5 Pseudonymisation de paquet de données

- Âge (en tranches approximatives), sexe, conditions préexistantes

- Symptômes signalés par les utilisateurs
- Statut de diagnostic certifié
- Le nombre et la durée des contacts, ainsi que les niveaux de risque de ces contacts - cela n'inclut pas les messages de risque eux-mêmes, pour prévenir d'éventuelles attaques via les graphiques sociaux, mais seulement les niveaux de risque ainsi que les métadonnées de la date de contact.
- Types de lieux visités et activité. Cela n'inclura PAS les informations relatives à la localisation réelle. Au lieu de cela, le téléphone regroupera localement les lieux par type (par exemple, résidence, épicerie, rue, etc.), et n'enverra que les types de lieux visités.
- Identifiant pseudonymisé - nécessaire pour permettre à l'utilisateur de révoquer ultérieurement son consentement et de supprimer ses données du serveur d'apprentissage automatique COVI.

En utilisant la classification établie par El Emam & Malin [75], nous constatons que ce paquet de données ne comprend pas d'identifiants directs, mais plusieurs quasi-identifiants, dont l'âge, le sexe, les conditions préexistantes et le statut de diagnostic certifié. L'âge est approximativement divisé en tranches de 10 ans afin de réduire le risque de réidentification. Le nombre et la durée des contacts ne sont pas considérés comme un identifiant, car ils ne sont pas une valeur stable dans le temps. De même, nous ne considérons pas les types de lieux visités comme un identifiant, car nous n'incluons pas de lieux spécifiques dans le paquet de données. L'identifiant pseudonymisé est une chaîne générée de manière aléatoire, nécessaire uniquement pour que nous puissions répondre aux demandes de suppression de l'utilisateur. Les informations contenues dans le paquet 1 correspondent donc à des informations non identificatoires comparables à celles que l'on trouve dans une base de données d'essais cliniques.

Le paquet de données pseudonymisées sera regroupé par le téléphone dans un fichier compressé, puis chiffré avec la clé publique de l'organisation COVI Canada avant d'être envoyé au serveur d'apprentissage automatique de COVI. Ces données sont bien sûr sensibles (comme les dossiers médicaux non identificatoires ou les informations sur les essais cliniques), ce qui fait courir un risque de réidentification si une entité malveillante y accède. C'est pourquoi le partage nécessite un consentement actif de l'utilisateur (section 2.4). Les principaux moyens de protection des données pseudonymisées sont juridiques plutôt que techniques. Les techniques de pseudonymisation examinées ici visent uniquement à réduire le risque de divulgation, mais ne peuvent et ne doivent pas être utilisées pour le supprimer. Des protocoles de sécurité standard pour la protection des données, tels que la certification SOC 2, seront utilisés pour protéger ces fichiers.

2.6 Informations géographiques séparées

Les cartes thermiques des niveaux de risque locaux sont essentielles pour permettre aux responsables de la santé publique de repérer et de réagir aux éclosions locales. Pour éviter d'identifier un utilisateur par sa géolocalisation, nous n'envoyons pas la géolocalisation d'un utilisateur directement au serveur, mais seulement des paquets séparés contenant des informations localisées sur les incidences de contact et les niveaux de risque par le biais d'un réseau mixte. Ces paquets ne seront à nouveau envoyés qu'avec le consentement supplémentaire du souhait de fournir sur une base volontaire des données à des fins de modélisation et de statistiques (le même consentement que la volonté de fournir volontairement des données pour l'entraînement de l'estimateur d'apprentissage automatique).

Chacun de ces paquets de données d'occurrences des contacts comprendra des informations de localisation approximatives au niveau de résolution spatiale d'une aire de diffusion de Statistique Canada (une zone ; voir la

section 2.6.1), qui visent entre 400 et 700 personnes. Ils comprendront également quelques autres variables latentes telles que la mobilité et l'aversion au risque.

Pour la modélisation et le suivi des éclosions, plusieurs types de paquets de données sont nécessaires. Ils seront tous envoyés séparément au serveur d'apprentissage automatique via un réseau mixte, de sorte qu'ils ne puissent pas être facilement connectés à un seul utilisateur. De plus, dès que le serveur d'apprentissage automatique reçoit le paquet, il le regroupe immédiatement avec d'autres paquets correspondant aux mêmes zones/jours, en rejetant le paquet original.

Paquet de cartes thermiques Ces informations permettront aux responsables de la santé publique de cartographier les lieux que fréquentent les utilisateurs à haut risque, sans révéler qui sont ces utilisateurs.

- Zone traversée pendant la journée
- Le jour où cette zone a été traversée
- Niveau de risque personnel ce jour-là
- Variable latente de mobilité/aversion au risque (4 bits)
- Ancien niveau de risque si un paquet précédent a été envoyé, mais le téléphone dispose maintenant de meilleures informations sur le niveau de risque pour ce jour-là (par exemple après le diagnostic).

Notez que les données de la carte thermique n'incluent aucun identifiant direct ou quasi identifiant, car aucune de ces informations n'est stable dans le temps. Nous notons cependant que parfois la zone traversée correspondra à la résidence de quelqu'un, c'est pourquoi il est important de ne pas envoyer les coordonnées GPS exactes, mais seulement une aire de diffusion d'au moins 100 personnes. La variable latente de mobilité/aversion au risque est une statistique sommaire qui correspond à peu près à la fréquence des sorties et des rencontres, ainsi qu'aux habitudes d'hygiène, comme le port d'un masque et le lavage des mains. Alors que les habitudes d'hygiène sont relativement stables, le téléphone calculera un seul journal de la mesure dans laquelle les mouvements d'une personne sont risqués, et l'utilisera pour quantifier le niveau de risque pour chaque jour, ce qui contribue à informer sur le risque de transmission. Nous pensons que le calcul de ce niveau de risque n'est pas un identificateur, car il changera au fur et à mesure que les habitudes de mouvement de la personne évolueront dans le temps.

Paquet de cartes des flux Ces informations permettront aux responsables de la santé publique et aux épidémiologistes de cartographier le flux du risque de transmission entre les résidences.

- Zone résidentielle.
- Jour de l'occurrence des contacts.
- Zone de l'occurrence des contacts.
- Niveau de risque reçu d'un autre contact.
- Ancien niveau de risque reçu d'un autre contact (si un paquet précédent a été envoyé, mais a été mis à jour).

Ce paquet comprend le quasi identifiant de la zone de résidence, qui peut identifier un utilisateur parmi 100 personnes possibles. Les trois autres éléments d'information ne sont pas des identificateurs. Ce paquet de

données est plus révélateur que le paquet de cartes thermiques, mais il sera bien sûr immédiatement regroupé dès sa réception par le serveur d'apprentissage automatique.

Ces données seront envoyées via un réseau mixte d'envoi 2.3.4, afin de mélanger les paquets de données avec ceux des autres utilisateurs et de masquer l'adresse IP de l'expéditeur. En plus du regroupement immédiat sur le serveur d'apprentissage automatique dès sa réception, cette procédure permettra de limiter les attaques de géolocalisation, même sur ces données de type option d'adhésion (opt in).

2.6.1 Discrétisation de la géolocalisation

Le brouillage de la géolocalisation sur l'appareil (ou discrétisation de la géolocalisation) pour le partage de données externes est basé sur un processus qui compte plusieurs étapes. Ce processus minimise la latence, la consommation de la batterie et les échanges avec le réseau de diffusion de contenu (RDC) de l'application. La première étape consiste en une recherche de haut niveau de la localisation de l'utilisateur basée sur des coordonnées GPS qui identifient les données géographiques nécessaires à la discrétisation. En utilisant un R-arbre dont le codage se rapporte seulement aux 293 limites géographiques des divisions de recensement définies par Statistique Canada pour le recensement de 2016 [76], l'utilisateur est approximativement localisé dans une province/territoire et dans un groupe de municipalités locales. Ensuite, toutes les limites géographiques plus détaillées (également appelées aires de diffusion) qui coïncident avec le groupe de municipalités identifié sont téléchargées à partir du RDC si elles ne sont pas déjà sur l'appareil. Cet échange avec le RDC est limité aux régions regroupées et ne doit pas entraîner de fuite d'informations de localisation importantes vers des tiers. Ensuite, un deuxième R-arbre est développé sur l'appareil avec les aires de diffusion téléchargées. La position GPS de l'utilisateur peut enfin être discrétisée à l'aide de ce nouvel arbre vers une seule aire de diffusion définie par Statistique Canada (77). L'identifiant unique de l'aire de diffusion est l'index renvoyé comme résultat de la discrétisation de la géolocalisation.

Les aires de diffusion (« zones ») sont des régions géographiques relativement petites et stables dont la population se situe entre 400 et 700 personnes. Il s'agit de la plus petite région géographique pour laquelle Statistique Canada fournit des données de recensement diffusées. Dans les villes plus peuplées, les cas extrêmes peuvent correspondre à un pâté de maisons et contenir plus de 700 personnes. Dans les régions éloignées, les cas extrêmes peuvent s'étendre sur des dizaines de kilomètres et contenir moins de 100 personnes. Nous n'enverrons des données que sur les aires de diffusion qui contiennent au moins 100 personnes.

2.6.2 Procédure de regroupement à l'usage du gouvernement

Comme décrit ci-dessus, les données agrégées seront partagées avec les autorités sanitaires à des fins de santé publique. Le serveur d'apprentissage automatique de COVI joue le rôle d'intermédiaire, tant pour les données de *géolocalisation* immédiatement regroupées que pour les données démographiques/données sur les symptômes du paquet de données pseudonymisées. Ces données comprennent :

- Une carte thermique quotidienne des zones sensibles d'infection et du flux de risque de transmission. Afin de réduire le risque de réidentification, les lieux seront regroupés comme décrit ci-dessus (dans les aires de diffusion).

- Modèles épidémiologiques. Les modèles de risque formés peuvent être ajustés conjointement avec les modèles épidémiologiques et ils constituent bien sûr aussi une forme d'agrégation de données, pour laquelle notre approche est particulièrement bénéfique. Ces modèles informeront les politiques publiques sur les types de contacts et de symptômes les plus à risque pour la propagation de l'infection, ainsi que sur la manière dont les différentes décisions pourraient se concrétiser selon les simulations.

- Informations démographiques agrégées sur les symptômes et le statut de l'infection. Bien qu'elles soient de nature extraordinairement sensible, les données sur la relation entre les symptômes et la démographie sont inestimables pour fournir des informations précises au public et prendre les mesures politiques appropriées pour contrôler au mieux la propagation.

Nous nous efforcerons de respecter la barre des 100 anonymes (au sens de k-anonymat [78]) pour toutes les données agrégées. Nous veillerons à ce que les algorithmes d'apprentissage utilisés soient robustes pour contrer des attaques [79].

De plus, au cours de la pandémie, nous nous attendons à ce que les autorités de santé publique nous demandent d'autres types d'informations agrégées. Si nous pouvons y répondre tout en respectant la barre des 100 anonymes (ou quelque chose de comparable), nous calculerons ces réponses à partir des données pseudonymisées et publierons ces informations.

2.6.3 Politique de stockage des données

Les données pseudonymisées et les paquets de données de zones géographiques à risque nécessaires à l'entraînement de modèles de prédictions statistiques et épidémiologiques seront stockés dans un serveur sécurisé dont l'accès sera limité à certains chercheurs en IA qui entraîneront ces modèles. Ces machines seront gérées par COVI Canada, une organisation à but non lucratif qui se consacre à la gestion de ces données selon les normes les plus élevées de bonne gouvernance et qui a pour seul mandat de protéger la santé, le bien-être, la dignité et la vie privée des Canadiens. Toutes les données sont chiffrées avant de quitter le téléphone de l'utilisateur avec la clé publique de COVI Canada, stockées tout en étant chiffrées et jamais déchiffrées avant d'être utilisées par un chercheur en IA, à l'exception des paquets de données de zones géographiques à risque qui seront immédiatement agrégés puis détruits. Nous espérons obtenir de nouveaux lots de données chaque jour et recycler les modèles de prédiction des risques à ce rythme.

2.6.4 Politique de conservation des données

Toutes les données pseudonymisées non groupées seront automatiquement effacées après une période ne dépassant pas 90 jours. Les utilisateurs ont en outre la possibilité de révoquer leur consentement à tout moment en utilisant l'application, ce qui entraînera la suppression de leurs profils du serveur de COVI. En raison de l'existence de sauvegardes rotatives externes, les révocations peuvent prendre jusqu'à 60 jours pour se répandre complètement (bien que dans la plupart des cas, elles doivent être effectuées dans les 30 jours). Nous notons que pour des raisons techniques, la révocation du consentement ne peut se faire que par le biais de l'application, car le serveur a besoin de connaître l'identifiant pseudonymisé pour déterminer les profils à supprimer. Ainsi, un utilisateur dont le téléphone est réinitialisé ne pourra pas révoquer son consentement ; pour lui, cependant, les données expireront toujours dans les 90 jours.

Les données agrégées et les modèles de risque seront conservés indéfiniment à des fins de recherche et de reproduction. Bien entendu, étant donné que les données agrégées et les modèles de risque sont largement diffusés - dans le premier cas, aux autorités sanitaires gouvernementales, dans le second, aux téléphones des utilisateurs pour la prédiction des risques locaux - il faut supposer que ces données sont accessibles aux utilisateurs malveillants. C'est pour cette raison que les données agrégées et les modèles de risque doivent être non identificatoires, comme décrits ci-dessus. Nous espérons être en mesure de donner de solides garanties de confidentialité k-anonymat [78] pour un tel ensemble de données agrégées, bien que les détails à cet égard restent à déterminer et dépendent de la stratégie de regroupement exacte.

De plus, nous pouvons utiliser les caractéristiques des données pseudonymisées pour générer un ensemble de données synthétiques présentant des caractéristiques similaires à celles des données brutes originales. De

même, cet ensemble de données synthétiques peut être conservé indéfiniment, mais ne doit pas être vulnérable aux attaques de type « linkage », et peut même être rendu public pour être utilisé par d'autres chercheurs.

2.7 Risques résiduels et mesures d'atténuation

Malheureusement, les pirates informatiques, les fraudeurs et autres agents malveillants font partie de la vie, s'attaquant, entre autres, à la confiance qu'ont les utilisateurs dans les institutions. Une fois que COVI sera déployé avec le soutien du gouvernement, il fera partie de l'écosystème et sera vecteur d'attaque. Et, nous devons garder cela à l'esprit lorsque nous concevons nos messages et nos protocoles.

Dans cette section, nous aborderons les deux attaques techniques restantes contre le protocole que nous ne protégeons pas entièrement et des attaques d'ingénierie sociale visant les utilisateurs eux-mêmes. Nous énumérerons ci-dessous certaines des attaques que nous envisageons, ainsi que les moyens potentiels d'atténuation (notez qu'il y a bien sûr un certain chevauchement avec les limitations inhérentes à la protection de la vie privée de la section 2.2, car les risques de COVI font partie d'un ensemble des risques que comporte le traçage des contacts décentralisé en général).

2.7.1 Attaque de type « vigilante »

L'attaque nommée « vigilante », ou encore l'attaque par triangulation est une attaque par laquelle l'attaquant cherche à « démasquer » un individu comme étant infecté. Notre modèle de protection de la vie privée est conçu de telle sorte que (1) il est impossible d'effectuer une telle attaque rétroactivement (c'est-à-dire qu'après que le niveau de risque d'Alice ait augmenté, elle essaie alors de retrouver la source), et (2) cela est techniquement et logistiquement difficile à réaliser.

Une fois qu'une application détermine l'occurrence des contacts, nous ferons en sorte qu'elle oublie les messages originaux qui ont été reçus via Bluetooth ; l'application ne dispose donc plus d'information sur l'heure exacte où le contact a été établi. En oubliant délibérément autant d'informations temporelles que possible, nous espérons réduire le risque d'attaque tout en conservant la possibilité de dépister les contacts et de continuer faire connaître les risques d'infection.

Attaque préméditée Malheureusement, quelqu'un qui a créé à l'avance une version piratée de l'application pourrait lui faire enregistrer tous les contacts et les messages de risque associés, ainsi que l'heure et le lieu exacts du contact. Cette attaque constitue une violation de données particulièrement grave lorsqu'elle est réalisée par une entité (par exemple un hôtel) qui connaît le lieu et l'identité d'un utilisateur particulier, car cette entité peut alors exposer l'état de santé d'un utilisateur.

Comme indiqué précédemment, ce type d'attaque est possible avec n'importe quelle application de traçage des contacts, quelles que soient les garanties (section 2.2). Notre protocole technique ne tente pas d'empêcher cette attaque préméditée, bien que nous devrions noter qu'un attaquant n'a besoin que d'une copie piratée de l'application COVI à l'avance, plutôt que de téléphones supplémentaires. Comme il n'est pas possible de prévoir cette attaque sur le plan technologique, nous devons plutôt explorer des solutions juridiques et économiques.

Attaque rétroactive Le problème de cette attaque multipartite, où plusieurs individus se regroupent pour tenter de déterminer qui les a infectés, est encore plus difficile à prévenir. Cependant, il faut noter que ce problème n'est pas spécifique à une application de traçage des contacts. Si plusieurs personnes tombent malades après une rencontre de groupe, avec ou sans application, elles peuvent se coordonner pour découvrir quel participant à cette rencontre les a infectées. Nous essayons de faire en sorte que notre application n'aggrave pas ce problème en n'exposant pas les détails des événements de contact aux utilisateurs eux-mêmes, mais en fournissant simplement des recommandations qui dépendent d'une mise à jour d'un niveau de risque. Par

exemple, nous ne fournissons pas de clarté sur le niveau de risque (l'application ne donne que des recommandations qui dépendent généralement aussi d'autres facteurs comme le lieu et les conditions médicales) et nous ne disons pas à l'utilisateur quelle occurrence a pu entraîner un changement de ses recommandations, le cas échéant.

2.7.2 Attaques par des autorités malfaisantes (*rogue authority attacks*)

Bien qu'il soit possible pour des individus de réaliser certaines des attaques ci-dessous, nous pensons que le plus grand risque vient ici d'un agent malveillant qui corrompt le serveur de messagerie et le serveur d'apprentissage automatique.

Attaques visant le graphe social (*social graph attacks*) Afin d'empêcher le serveur de messagerie/un agent malveillant de ces attaques, nous cachons soit les schémas d'envoi de Bob avec un réseau mixte, soit les schémas de récupération d'Alice en téléchargeant tous les messages dans une zone géographique. Dans le cas d'un réseau mixte, notez que si nous protégeons le graphe social en cachant les schémas d'envoi de Bob, Alice récupère directement les messages, de sorte que le serveur de messagerie peut voir le nombre d'interactions sociales d'Alice - mais pas avec qui. On peut éviter ce problème en demandant à Alice de récupérer les messages via un réseau mixte ou en anonymisant les serveurs mandataires.

De plus, un certain risque subsiste dans ce type d'attaque sur des sous-groupes où un nombre suffisant d'individus choisissent d'envoyer des données pseudonymisées aux serveurs d'apprentissage automatique COVI. L'agent malveillant peut injecter des niveaux de risque uniques spécifiques en faisant en sorte qu'un dispositif diffuse des niveaux de risque peu communs aux contacts. Ces contacts téléchargent ensuite ces niveaux de risque uniques sur le serveur d'apprentissage automatique pour la carte des flux, révélant ainsi leurs aires de diffusion d'origine, car ces messages peuvent être marqués par le serveur d'apprentissage automatique. Ces niveaux de risque uniques peuvent ensuite être utilisés pour marquer le paquet de données pseudonymisées de la même manière, ce qui peut permettre des attaques par inférence.

Nous atténuons cette vulnérabilité de deux manières : (1) les niveaux de risque sont quantifiés à 4 bits (16 niveaux), et (2) le prédicteur de risque utilise toute la gamme des 16 niveaux. Étant donné que le prédicteur de risque d'apprentissage automatique fournit régulièrement des résultats dans toute la gamme, il n'y aura pas de niveau de risque unique pouvant être utilisé pour le marquage. Le fait que le prédicteur de risque d'apprentissage automatique utilise toute la gamme peut être vérifié de manière indépendante par des tiers, car le prédicteur de risque est une information publique.

Attaque visant l'historique de localisation Une autorité malfaisante (Grace) qui contrôle le serveur de messagerie et qui a déployé des dispositifs de traçage des contacts à travers une ville peut obtenir une quantité importante d'informations sur l'historique de localisation des utilisateurs. Il s'agit là d'une faiblesse inhérente au traçage des contacts décentralisé (section 2.2). Bien que les attaques sur le graphe social soient combattues en cachant les schémas d'envoi de Bob ou de récupération d'Alice, celui des deux qui n'est pas dissimulé peut subir une fuite de sa localisation vers le serveur de messagerie.

Dans le cas de Bob, si les schémas d'envoi ne sont pas cachés (par défaut dans la structure de l'API notification d'exposition de Google et Apple [13] et dans le protocole TCN [45]), alors l'ensemble des communications Bluetooth qu'il a envoyées sont en théorie connues du serveur de messagerie. En déployant un ensemble de dispositifs d'écoute utilisant la technologie Bluetooth à de nombreux endroits, le serveur apprend ainsi une partie de localisation de Bob.

De même, dans le cas d'Alice, si les schémas de récupération ne sont pas cachés (ce qui est le cas par défaut dans l'approche du NHS + réseaux mixtes), alors l'ensemble des messageries auxquelles Alice se connecte sont envoyées au serveur de messagerie. Si le serveur de messagerie déploie des dispositifs Bluetooth actifs à de

nombreux endroits qui se relient au Bluetooth d'Alice, le serveur apprend alors une partie de l'emplacement d'Alice. Dans un système binaire de notification d'exposition post-diagnostic, cette attaque est pire, car seuls certains utilisateurs envoient des notifications (comme Bob), mais chaque utilisateur se connecte (comme Alice). Dans notre système de transmission de messages de risque, de nombreux utilisateurs ont au départ un risque de base nulle, donc la distinction entre Alice et Bob est principalement une question de rôle, car la plupart des utilisateurs joueront les deux rôles.

Une potentielle atténuation partielle de ces attaques sur l'historique de localisation consiste à cacher les schémas d'envoi et de récupération. Par exemple, si un réseau mixte d'envoi est utilisé pour Bob dans les propositions de structure de Google et Apple et du protocole TCN, cela assurerait un certain degré d'anonymat (bien que des traces partielles de 6 à 24 heures puissent encore être disponibles pour Grace). Une autre possibilité serait d'utiliser un réseau mixte de récupération pour Alice dans le cadre de l'approche du NHS + réseau mixte, bien que les réseaux mixtes de récupération soient plus difficiles à mesurer que les réseaux mixtes d'envoi unique.

Nous étudions activement ces deux mesures d'atténuation, mais nous n'avons pas encore tiré de conclusions et ne pouvons pas faire de promesses quant à leur faisabilité/degré de mesure. De plus, nous constatons que même avec ces protections, un agent malveillant qui déploie des dispositifs Bluetooth dans une ville peut se faire passer pour de véritables contacts et recevoir des messages de risque. Ces messages de risque sont eux-mêmes reliés et peuvent révéler partiellement des identités d'expéditeurs des messages, de sorte que cacher l'adresse IP de l'expéditeur par le biais d'un réseau mixte ne constitue de toute façon qu'une protection partielle.

En outre, nous notons que bien que le traçage des contacts puisse être utilisé comme un mécanisme de localisation, il existe déjà de nombreuses autres options disponibles pour les adversaires disposant de ressources gouvernementales, comme l'utilisation de la vidéosurveillance et de la reconnaissance faciale ou des pings des tours de téléphonie mobile. C'est en raison de ces difficultés que nous avons classé l'attaque ciblant l'historique de localisation dans la catégorie des risques restants. Comme il s'agit d'attaques par l'autorité centrale, nous espérons qu'il existe des moyens de défense juridiques et gouvernementaux.

2.7.3 Fuite de renseignements sur la vie privée commise par les entreprises

Une autre source de risque restant est constituée par les entreprises ou institutions, autres que le serveur de messagerie, qui souhaitent obtenir plus d'informations sur un groupe ciblé de personnes qu'elles ont physiquement dans leurs locaux : par exemple, un employeur qui souhaite espionner ses employés. Il existe déjà de nombreux exemples d'employeurs effectuant un suivi de localisation avec, par exemple, des détecteurs de mouvement ou via le système WiFi, et il est important d'examiner comment COVI pourrait apporter un autre moyen de localisation. Une autre version moins malveillante de ce système pourrait être celle d'une épicerie qui voudrait connaître le niveau de risque de ses clients, de sorte que certaines de ces informations pourraient être « divulguées ».

Fuite de renseignements sur l'état de santé Toute personne, y compris une entreprise, peut installer un dispositif se faisant passer pour COVI, qui recevra alors les messages de risque quelques jours plus tard de toute personne passant devant le dispositif. Là encore, il s'agit d'une fuite inhérente au traçage automatique des contacts décentralisés (section 2.2). Heureusement, le délai d'un jour maximum entre la rencontre et la transmission du message de risque rend impossible l'association d'un niveau de risque en temps réel à une personne particulière passant devant le dispositif.

Pour les cas d'utilisation légitime, l'équipe de COVI devrait envisager de fournir une application de veille locale qui ne révèle pas les heures exactes, mais seulement les statistiques de risque cumulé d'un lieu. Ces informations sont similaires aux cartes thermiques que nous prévoyons de fournir aux responsables de la santé publique, et elles ne devraient révéler les mêmes informations agrégées qu'avec le k-anonymat protégeant les renseignements privés.

Bien que nous ne puissions pas empêcher technologiquement le traçage illégitime, des protections légales peuvent être mises en place. Par exemple, dans les conditions de service, nous demandons aux utilisateurs de ne pas pirater le système ; cela n'empêchera pas un attaquant malveillant, mais pourrait au moins décourager les entreprises d'envoyer/de recevoir de faux messages sur le système, surtout si nous pouvons leur fournir une possibilité de veille locale légitime.

Fuite d'informations de localisation hyperlocale Malheureusement, bien que les renseignements médicaux et les informations sur le niveau de risque ne soient disponibles que par le biais du protocole de l'application, que nous contrôlons, il y a une certaine quantité d'informations qui sont divulguées simplement en diffusant les messages Bluetooth. En installant des récepteurs Bluetooth, une entreprise peut trianguler à 2 mètres près la position à tout moment de chaque personne se trouvant dans ses locaux. Bien que cela soit déjà possible en partie grâce aux détecteurs de mouvement et au système WiFi, le Bluetooth permet probablement des versions à plus haute résolution de la même chose. Notez que comme nos messages Bluetooth changent tous les 5 à 15 minutes, un tel système ne sait pas exactement qui est chaque personne qu'il détecte. Cependant, on peut en déduire beaucoup de choses, par exemple, à partir du temps pendant lequel une personne est présente à son bureau (bien qu'il soit également futile d'avoir d'autres moyens pour un employeur de savoir qu'une personne donnée est présente à son bureau).

Un tel système peut également être utilisé pour suivre les déplacements des personnes dans un lieu public (par exemple, une épicerie). Le système ne saurait pas qui est entré, car les messages Bluetooth varient de manière aléatoire, mais il permettrait à l'épicerie de déterminer combien de temps les personnes s'arrêtent devant tel ou tel étalage. Nous ne pensons pas que cette information soit très différente de celle qui peut déjà être obtenue par des caméras de sécurité, des détecteurs de mouvement et le système WiFi, mais il s'agit d'une fuite de données supplémentaire.

2.7.4 L'hameçonnage

L'hameçonnage est une forte réalité sur Internet. Il est caractérisé par un fraudeur se faisant passer pour une entité de confiance et utilisant cette confiance pour convaincre sa cible de faire quelque chose. COVI, en tant que nouvelle entité bénéficiant d'un soutien gouvernemental, devra faire face à des fraudeurs se faisant passer pour lui.

Accès à un faux lien URL COVI lors de la mise en œuvre de l'application. Une fois la campagne publicitaire COVI lancée, les résidents du Canada seront encouragés à télécharger l'application COVI pour faciliter le traçage des contacts. Comme COVI est une application mobile, une attaque initiale évidente consiste pour Mallory à envoyer des messages texte non ciblés au plus grand nombre possible de personnes, en espérant duper Alice avant qu'elle ne télécharge COVI. Ce message texte prétendrait provenir du gouvernement, l'encourageant à visiter le site http://fake-covi-url:ca pour télécharger l'application. Une fois qu'Alice aura visité ce lien URL, des données malveillantes pourront être téléchargées sur son téléphone, ou peut-être que l'URL lui demandera des informations personnelles (par exemple, le NAS, le numéro de carte d'assurance maladie, etc.). L'attaque peut être personnalisée en fonction des informations réelles que COVI demande (par exemple, des informations démographiques), de sorte que si Alice demande conseil à un ami de confiance, avisé en matière de technologie, celui-ci ne se rendra peut-être pas compte qu'Alice se trouve sur une fausse page COVI.

Pour cette raison, il est important de faire très attention aux informations que nous demandons à l'utilisateur de divulguer. Toute information personnelle que nous leur encourageons à divulguer est une information qu'ils peuvent être moins hésitants à divulguer à un site web malveillant.

De plus, un message peut également demander à l'utilisateur de partager le lien de façon virale, en s'appuyant sur le sens du devoir civique d'Alice. Ce message a plus de sens en tant que publication sur les médias sociaux. Par exemple : « Le gouvernement canadien nous demande à tous de télécharger COVI pour aider à combattre la COVID-19. Rendez-vous sur http://fake-covi-url.ca maintenant, et transmettez ce message à tous vos amis pour que nous puissions combattre la COVID-19 ensemble ! » La publication sur les médias sociaux peut être structurée avec un vrai lien vers une source d'information ou un communiqué de presse COVI, pour lui donner un air de légitimité, avec pour seule fausse donnée malveillante l'URL. Dès qu'un tel schéma est découvert, il est donc important d'en avertir les gens.

Installation d'une fausse application. Il est à noter que le lien URL ci-dessus ne doit pas nécessairement mener au téléchargement d'une application. En fait, le lien URL peut même rediriger vers la véritable application après avoir obtenu des informations sur l'utilisateur, afin que l'attaque passe inaperçue. Cependant, dans le cadre d'une attaque en cours, Mallory peut aussi diriger Alice et Bob vers le téléchargement d'une fausse application, ce qui compromet complètement les téléphones des utilisateurs. Cette attaque sera, espérons-le, empêchée par les plateformes de téléchargement d'applications d'Apple et de Google, mais il est possible que quelque chose de ce genre se glisse entre les mailles du filet. Un attaquant peut modifier une application existante pour qu'elle ressemble à COVI, et les gens pourraient recevoir des messages texte des « autorités sanitaires » avec un lien vers cette fausse application.

Faux message texte de diagnostic COVI. Lorsque le serveur de notification de diagnostic est mis en place, les utilisateurs de COVI pourront s'attendre à recevoir un message texte avec un code/URL pour permettre à leurs applications d'envoyer un avis d'infection. Si Mallory sait que Bob a récemment fait un test de dépistage, elle peut lui envoyer un message texte prétendant avoir les résultats de son test à un lien URL spécifié. L'URL peut être n'importe quelle donnée malveillante, comme décrit ci-dessus. Cette attaque est d'autant plus dangereuse que Bob, ayant récemment été en contact avec les autorités sanitaires, pourrait être plus enclin à divulguer un numéro de carte d'assurance maladie ou un NAS, car ce sont des numéros qu'il donne parfois au système de santé. Le site web malveillant peut demander de manière plausible à avoir besoin du nom, de l'adresse, du NAS et du numéro de carte d'assurance maladie de Bob, qui peut ensuite être réutilisé pour un vol d'identité.

Mallory n'a, dans les faits, pas besoin de savoir que Bob a fait un test de dépistage récemment, car elle peut bien sûr se contenter de diffuser le message texte en général. Cependant, heureusement, cette attaque ne semble pas se prêter à une propagation sur les médias sociaux.

Révéler les informations protégées par une application. Parfois, l'hameçonnage est conçu pour amener les individus à effectuer une action pour révéler des données stockées sur le téléphone. Cependant, comme COVI fournit à l'avance la clé chiffrée et le lien URL, ainsi que le serveur de coordination du réseau mixte, l'utilisateur n'a pas la possibilité de révéler sa localisation ou son empreinte Bluetooth à un tiers. Cela diffère des applications qui donnent à Alice et Bob la possibilité d'envoyer leur localisation géographique à un prestataire de soins (par exemple par courriel).

2.7.5 Propagation de la désinformation par le biais des niveaux de risque

Si Mallory veut simplement « voir le monde brûler », elle peut essayer de semer la panique. COVI permet aux utilisateurs de signaler eux-mêmes leurs symptômes qui entrent dans le calcul des niveaux de risque. Si Mallory peut convaincre un nombre suffisamment important d'utilisateurs de soumettre de fausses informations, elle peut être en mesure de briser le système de traçage des contacts et d'inciter la panique au sein de la population. Cette attaque n'est en général possible que jusqu'à ce que les autorités de santé publique soient reliées à COVI pour envoyer un diagnostic officiel confirmé avec un code unique. Après cela, aucun résultat de test non officiel ne peut passer par le réseau ; si les symptômes autodéclarés entrent toujours en ligne de compte dans le niveau de risque, ils jouent un rôle beaucoup moins important et ne peuvent pas influencer le niveau de risque autant qu'un diagnostic prétendu.

Faux résultats de diagnostic Il s'agit d'une variante du faux message texte de diagnostic ci-dessus, mais il doit être effectué par un appel téléphonique pour un effet maximal. Mallory appelle Bob, prétendant avoir des résultats de tests selon la méthode standard utilisée par les autorités de santé publique. Elle dit à Bob qu'il est infecté et qu'il doit informer ses contacts par le biais de COVI. Bien que COVI ne permette pas d'envoyer un diagnostic confirmé sans un code unique provenant des véritables autorités de santé publique, Mallory peut toujours dire à Bob qu'il doit signaler tous les symptômes dans l'application, ce qui augmente son niveau de risque. Ainsi, même sans code, Mallory peut augmenter le niveau de risque des contacts de Bob. Réalisé à grande échelle, cela peut générer de faux résultats et diminuer l'utilité de COVI pour le traçage des contacts. La principale méthode d'atténuation pour cette attaque est simplement le fait que les symptômes autodéclarés jouent un rôle beaucoup moins important dans l'estimation du risque une fois que des diagnostics officiels confirmés avec les autorités sanitaires provinciales sont intégrés.

Incitation à la fausse déclaration Si COVI est sanctionné par le gouvernement, il est possible que des employeurs utilisent COVI comme preuve de maladie. Même si ce n'est pas le cas, Mallory peut convaincre les gens qu'ils pourront obtenir quelque chose en se déclarant eux-mêmes comme étant infectés. Cette méthode est probablement la plus logique en tant que publication virale sur les médias sociaux, un conseil malveillant pour les individus.

Le contenu du message indiquera aux lecteurs que s'ils se déclarent infectés, ils recevront de meilleurs soins médicaux ou pourront s'absenter du travail. Les déclarations sont amplifiées par les messages de risque des contacts. Dans l'ensemble, cela incitera à de l'anxiété et à la surutilisation des ressources médicales. Ces messages ont également pour effet secondaire d'accroître la méfiance à l'égard de COVI, car si les gens croient que d'autres personnes mentent, ils feront eux-mêmes moins confiance aux recommandations de COVI.

Bien qu'il n'y ait aucun moyen de se défendre pleinement contre les fausses déclarations avant l'intégration avec les autorités sanitaires provinciales, nous pouvons intégrer des incitations appropriées dans les messages de confirmation des autodéclarations. En raison des effets négatifs des autodéclarations des utilisateurs, nous recommandons également que le statut de risque et les recommandations de COVI ne soient pas utilisés comme vérification de la maladie par des tiers, ce qui supprime ce type d'incitations.

2.7.6 Recourir au désir de vengeance

Dans sa version finale, nous avons conçu le protocole de protection de la vie privée pour qu'il soit difficile pour une utilisatrice, Alice, de déterminer qui l'a exposée après coup. Il est impossible d'empêcher Alice de déterminer que Bob l'a exposée si Alice le fait de façon préméditée (voir section 2.7.1), car COVI est conçu pour oublier les informations exactes sur l'heure et le lieu de contact avant l'envoi de tout message de risque. Cela rendra plus difficile pour Alice de déterminer, plus tard, que c'est Bob qui l'a exposée, bien que cela soit bien sûr imparfait.

Cependant, un utilisateur qui souhaite se venger peut ne pas comprendre les mesures de protection de la vie privée de l'application. Mallory peut promouvoir auprès d'Alice un service où Alice paye/télécharge une application et Mallory prétendra pouvoir découvrir l'identité de Bob. Pour les besoins de cette attaque, il importe peu que Mallory soit capable de le faire ou non. Mallory peut toujours soit faire en sorte qu'Alice télécharge une application malveillante, soit obtenir des bitcoins d'Alice pour effectuer ce service.

2.7.7 Campagnes de désinformation

Bien que nous ne nous attendions pas à d'actives campagnes de désinformation dans les premières phases du lancement de l'application, nous ne pouvons pas exclure la possibilité de telles attaques, étant donné leur prévalence en ligne de nos jours, parfois même de la part d'acteurs étatiques étrangers [80]. De nombreuses campagnes de désinformation prendront la forme d'attaques d'ingénierie sociale décrites ci-dessus (et peuvent

être évitées en tant que telles), mais il en existe quelques-unes qui sont spécifiques aux motivations et à l'ampleur de ces attaques.

Faux niveaux de risque élevés Un moyen facile de détruire l'utilité de l'application serait de créer un grand nombre de faux rapports de niveaux de risque élevés - par exemple, imaginez 6 millions d'applications signalant un risque d'infection élevé à Montréal, Toronto et Vancouver. Heureusement, Bluetooth fournit une preuve de présence ; en outre, les trois approches envisagées sont conçues pour empêcher les attaques par rediffusion, ce qui rend impossible pour un attaquant de réécouter des messages existants avec de fausses mises à jour. Ainsi, cette attaque exige que l'attaquant ait un dispositif physique présent diffusant des signaux Bluetooth.

Bien que nous nous attendions à ce que l'exigence d'un dispositif physique soit le principal moyen de dissuasion, nous pouvons encore limiter quelque peu l'impact de ce type d'attaque en limitant le nombre de messages par jour par adresse IP et par réseau. Cela peut contribuer à empêcher qu'un seul téléphone prétende être plusieurs téléphones simultanément (bien qu'il s'agisse bien sûr d'une protection imparfaite). De plus, si les serveurs commencent à détecter un nombre anormal de messages provenant d'un bloc de PI, ou d'adresses IP d'origine non canadienne, cela peut être filtré comme une éventuelle campagne de désinformation. De plus, ces mêmes mécanismes de protection contre les abus peuvent aider à se prémunir contre les attaques par déni de service sur l'infrastructure.

Bien entendu, Mallory peut élaborer une version malveillante de COVI qui se contente d'envoyer un maximum de niveaux de risque non diagnostiqués à toutes les personnes qu'elle rencontre en se promenant, cela possiblement plusieurs fois, et de trouver ensuite suffisamment d'adresses IP pour diffuser les faux messages appropriés. Cela est impossible à empêcher, et ressemble au comportement d'une véritable application COVI avec un utilisateur infecté. Elle ne peut pas faire beaucoup de dégâts, car elle ne peut être qu'à un seul endroit à la fois. Cependant, si elle peut convaincre un grand nombre de personnes, réparties dans tout le pays de le faire, elle fragilisera l'utilité des prédicteurs de risque.

Une méthode d'atténuation partielle de cette attaque consiste à ce qu'un modèle d'apprentissage automatique compte plusieurs contacts à haut risque simultanés comme l'équivalent d'un seul, car une application malveillante peut simuler 1 000 patients infectés à un endroit particulier. Dans nos simulations épidémiologiques, nous prévoyons d'étudier si cet ajustement diminuerait la précision des prévisions et si ce n'est pas le cas, cette atténuation pourrait être mise en œuvre.

Fuites de données Un adversaire bien financé peut également cibler la politique de confidentialité du système afin d'empêcher son adoption. La plupart des données de traçage des contacts sont conservées derrière les systèmes de messagerie privés que nous envisageons (section 2.3). Cependant, une cible est le serveur de collecte de données d'apprentissage automatique COVI, qui contiendra des données pseudonymisées sur des centaines de milliers d'utilisateurs. Bien qu'il n'y ait pas d'association entre les données de localisation et les utilisateurs individuels, une violation de données sur le serveur d'apprentissage automatique COVI serait toujours de la même ampleur qu'une violation de données sur les dossiers d'un grand système hospitalier (mais sans aucun identificateur complet), et le serveur devrait donc être traité avec la même prudence. Une telle violation non seulement exposerait des informations personnelles, mais elle porterait également atteinte à la confiance du public dans la confidentialité du système dans son ensemble.

Pour cette raison, les données brutes pseudonymisées ne devraient pas être conservées sous cette forme plus longtemps qu'il n'est absolument nécessaire pour former les modèles. Comme décrit, une certaine durée de conservation est nécessaire pour disposer de suffisamment de données pour former des modèles de risque précis, mais les données brutes devraient être (et seront) automatiquement expirées régulièrement, ne laissant que les données agrégées. En particulier, toute donnée associée à un lieu sera immédiatement cumulée et supprimée, car nous n'avons pas besoin des données individuelles associées à un lieu et elles sont particulièrement sensibles. Voir la politique de conservation des données 2.6.2 pour plus de détails.

3 Détails du modèle épidémiologique

Les données pseudonymisées fournies seront utilisées pour ajuster des modèles épidémiologiques au point de vue individuel qui captent le fil stochastique des événements dans le temps. Ces événements comprendront notamment les déplacements des personnes virtuelles de la simulation, les rencontres entre elles, les événements médicaux et les comportements (comme porter un masque). Ces modèles pourront ensuite être intégrés dans un simulateur que les responsables de la santé publique pourront utiliser pour cartographier géographiquement le développement de la maladie, comprendre les choix des citoyens et mieux définir les facteurs qui ont une incidence sur la contagion.

3.1 Structure du simulateur épidémiologique

Le simulateur est un modèle stochastique basé sur des agents mis en œuvre avec SimPy. Une population humaine est créée dans une ville, et chaque humain se déplace dans la ville selon des tendances concernant la mobilité que génère un modèle de DSE (dossier de santé électronique). Une partie des humains sont atteints de la maladie et, tandis qu'ils se déplacent dans la ville (passant du temps dans des endroits comme leur domicile, leur travail, les transports, les magasins, les hôpitaux, les centres de soins de longue durée, etc.), ils peuvent s'infecter les uns les autres, avoir des symptômes, être hospitalisés, etc. Nous suivons la propagation de la maladie à travers plusieurs paramètres (*R*, taux d'atteinte, etc.) et nous réglons les paramètres du simulateur pour que ceux-ci correspondent aux données réelles et à la sortie d'un modèle de comparaison (SEIR [81]) ajusté selon les données provenant de Wuhan et adapté à la démographie canadienne [36]. Le simulateur produit des séquences de rencontres accompagnées de renseignements sur la transmission de la maladie, et nous les utilisons pour créer un jeu de données pour les modèles d'apprentissage automatique afin de prédire le risque individuel à partir de variables observées, comme les symptômes, les problèmes de santé préexistants et les lieux visités par la personne. Ce prédicteur de risque peut à son tour servir à régler les paramètres du simulateur selon les données qui seraient collectées par une application mobile.

3.1.1 Détails relatifs à la mise en œuvre

La ville est dotée d'une structure de graphe des lieux — notamment les domiciles, les magasins, les parcs, les hôpitaux (comprenant les unités de soins intensifs) et les centres d'hébergement — et de différents modes de transport — y compris la marche, le vélo, le métro/l'autobus, le covoiturage/le taxi et la voiture. À chaque lieu et à chaque moyen de transport sont associées des propriétés relatives à la capacité et à la transmission de la maladie.

Chaque humain possède des caractéristiques individuelles (âge, sexe, problèmes de santé préexistants, prudence, s'il a ou non l'application, fréquence de port du masque, lieu de travail, etc.) échantillonnées selon les renseignements démographiques du Canada. Les humains sont dotés de propriétés épidémiologiques, notamment la charge virale, l'infectiosité et la progression des symptômes durant la maladie, qui dépendent de ces caractéristiques individuelles (voir ci-dessous pour les détails).

De nombreux événements se produisent tandis que les humains se déplacent et que le temps passe :

1. Une rencontre se produit lorsque deux humains sont suffisamment rapprochés dans l'espace et le temps. L'espace est présentement doté d'une résolution d'environ 2 mètres, mais la variance est importante à cause des différences de mise en œuvre de Bluetooth sur différents appareils. Pour mitiger cela, nous utilisons la puissance du signal plutôt que la distance comme donnée d'entrée et nous permettons au prédicteur de tenir compte le mieux possible de l'incertitude inhérente en matière de

distance. Le temps est actuellement divisé en périodes de 15 minutes. Si un des humains est contagieux, il infectera l'autre selon une probabilité proportionnelle à son infectiosité.

2. Si les humains consignent leurs symptômes ou obtiennent un résultat positif ou négatif à un test, un nouveau niveau de risque est calculé. Nous ne modélisons actuellement qu'un seul type de test : les tests de laboratoire à 0 % de résultats faux positifs et 10 % de faux négatifs ; l'ajout d'autres types est prévu.
3. Les humains peuvent contracter un rhume ou une grippe (actuellement, un sous-ensemble aléatoire représentant 1 % de la population, pondéré selon l'âge et d'autres caractéristiques). Cela génère une distribution plus réaliste des symptômes (la COVID-19 n'est pas la seule cause de symptômes). Les allergies saisonnières sont aussi planifiées.

Nous avons appliqué différents degrés de distanciation sociale et d'autres interventions. Nous explorons actuellement leurs effets sur la propagation de la maladie.

L'horodatage de l'infection est le moment précis où un humain a été infecté. Les humains peuvent être infectés par un lieu, p. ex. si un humain très contagieux y était peu de temps auparavant, ou, ce qui est plus probable, par une rencontre avec un autre humain. Nous pistons la source de l'exposition pour chaque humain infecté.

La charge virale est modélisée comme fonction linéaire par morceaux à trois éléments : augmentation, plateau et diminution. L'augmentation commence après un nombre de jours d'incubation (courbe gaussienne ayant pour centre 2,5 jours) et atteint la valeur du plateau après un nombre de jours échantillonné d'une courbe gaussienne ayant pour centre 2,5 jours. Le plateau est présentement échantillonné uniquement selon l'âge, mais nous planifions de le faire dépendre d'autres caractéristiques individuelles (p. ex., problèmes de santé préexistants, changements de comportement) ainsi que de la charge virale initiale lors de la rencontre. Le plateau dure un nombre de jours échantillonné selon une courbe gaussienne ayant pour centre 5 jours. La diminution dure un nombre de jours échantillonnée d'une courbe gaussienne ayant pour centre 5 jours [36, 82].

L'infectiosité est proportionnelle à la charge virale, mais dépend de caractéristiques comme être asymptomatique, être immunocompromis, porter un masque, tousser, etc.

La progression des symptômes dépend de la charge virale [35, 83]. Pour chacun des trois stades, les symptômes sont échantillonnés selon leur prévalence moyenne chez les patients atteints de la COVID-19. Les symptômes commencent un nombre de jours suivant l'infection extrait d'une courbe gaussienne ayant pour centre 2,5 jours.

Une personne a approximativement 40 % de chance d'être asymptomatique (ce qui réduit son infectiosité à 10 % de ce qu'elle aurait autrement été), 15 % d'être très malade (devant être hospitalisée) et 30 % des personnes très malades deviendront extrêmement malades (nécessitant des soins intensifs). Le risque qu'une personne ne récupère jamais est de 0,2 %. Ces valeurs sont échantillonnées individuellement et dépendent de facteurs comme l'âge, les problèmes de santé préexistants, etc. Le simulateur pourrait modéliser la réinfection, mais ne le fait pas pour le moment à cause de l'incertitude qui plane toujours sur la fréquence.

Si une personne porte un masque ou non est présentement échantillonné selon le niveau de prudence de cette personne. Le port du masque est efficace à 98 % pour les travailleurs hospitaliers, et à 32 % pour les autres [84]. Nous validons plusieurs mesures à l'échelle de la population à propos des données simulées, y compris :

1. R ventilé par domicile, hôpitaux et autres lieux.
2. Taux de transmission lors des rencontres.
3. Taux d'atteinte secondaire (#testés positifs/#symptomatiques).
4. Fraction de cas symptomatiques selon l'âge.
5. Correspondance qualitative à une courbe SEIR.

4 Détails de l'apprentissage automatique

Les données pseudonymisées fournies volontairement au serveur d'apprentissage automatique COVI seront utilisées pour l'entraînement de modèles d'apprentissage automatique qui prédisent les risques de contagiosité et permettent d'ajuster un modèle épidémiologique. Les modèles seront entraînés hors ligne afin de ne pas surcharger les téléphones : l'entraînement sur les téléphones exigerait une grande puissance de calcul, car le processus d'entraînement est itératif et long. L'entraînement a lieu hors ligne également parce que nous aurons besoin d'essayer différents modèles pour déterminer la configuration optimale des algorithmes d'apprentissage. Ce n'est qu'ensuite que les algorithmes et les paramètres du modèle prédictif choisi seront envoyés aux téléphones. En ce qui concerne l'algorithme d'apprentissage présentement sur les téléphones, les paramètres devront être réestimés régulièrement (jusqu'à une fréquence quotidienne si nécessaire) et leurs valeurs mises à jour envoyées aux appareils.

Pour aider ce processus et nous assurer que les prédictions sont bien calibrées même dès les premiers jours de l'adoption de l'application, nous entraînons d'avance les modèles d'apprentissage automatique à l'aide de données simulées générées par une version a priori du modèle épidémiologique, laquelle est décrite ci-dessous. Ce modèle épidémiologique est aussi un simulateur qui peut créer des historiques de contacts individuels, de comportements et de transmission virale. Lorsque nous commencerons à collecter les vraies données issues de l'application, nous réglerons les paramètres du simulateur pour qu'ils correspondent aux données à mesure qu'elles sont recueillies. Incorporer l'utilisation de l'application et ses prédictions quant au risque dans la simulation nous permet de modéliser avec précision l'impact de différentes interventions, puisque l'objectif même de COVI est de responsabiliser les citoyens en leur fournissant des renseignements pour mener à un confinement ciblé plutôt qu'uniforme. Consultez la section 3.1 pour les détails relatifs au simulateur épidémiologique.

4.1 Rencontres entre utilisateurs

Lorsque deux téléphones dotés de l'application se rencontrent, ils échangent (avec un décalage allant jusqu'à une journée) de l'information au sujet du risque de l'un et l'autre (plus précisément, à quel point l'application estime que l'utilisateur est contagieux au moment de la rencontre). Plus tard, lorsque d'autres renseignements s'accumuleront sur chacun des téléphones, ces estimations de risque concernant le jour de la rencontre pourront être révisées. Si la révision est suffisamment importante (parce que le niveau de risque a changé d'un niveau distinct à un autre), un message de mise à jour est envoyé à l'autre téléphone. Par exemple, si un utilisateur commence à ressentir des symptômes liés à la COVID-19, le téléphone de cet utilisateur augmentera la contagiosité probable dans les jours précédents et enverra un message de mise à jour à tous les téléphones de gens que cet utilisateur a rencontrés au cours des 14 jours précédents. Cela permet à chaque utilisateur d'obtenir un niveau de risque personnalisé mis à jour et de propager ce risque aux contacts préalables. Si le risque change de manière importante (ce qui est probable si de nouveaux symptômes apparaissent chez un contact ou si l'utilisateur vient d'obtenir un résultat de test positif), un risque mis à jour sera alors propagé par des messages envoyés aux contacts passés de l'utilisateur. Le but de l'estimateur de risque est de prédire la contagiosité présente et passée d'un utilisateur. La première influence les recommandations personnalisées pour que les personnes puissent mieux protéger ceux qui les entourent, tandis que la seconde est envoyée aux contacts passés afin qu'ils puissent être mis au courant du risque qu'ils courent d'être contagieux et puissent agir en conséquence.

4.2 Considérations en matière de protection de la vie privée

Pour une description de la protection de la vie privée du point de vue technique, veuillez consulter la section 2. Nous donnerons ici un résumé des principaux éléments protégeant la vie privée des personnes sur le serveur d'apprentissage automatique utilisé pour entraîner les modèles et recueillir des données épidémiologiques.

4.2.1 Transmission des messages

Lorsqu'un téléphone envoie un message à un autre téléphone par l'entremise du protocole de messagerie protégée, le destinataire ne saura pas de quel appareil (ni son numéro ni son adresse IP) le message provient. Pour fournir une protection accrue contre la stigmatisation, ces messages sont envoyés avec un décalage aléatoire pouvant aller jusqu'à une journée. De cette manière, il n'est pas possible pour Alice, une utilisatrice honnête, de savoir que l'augmentation de son niveau de risque est due à une rencontre avec une personne précise (Bob), à moins qu'elle n'ait eu qu'une seule rencontre ce jour-là (voir la section 2.7 pour les attaques par des entités malveillantes). Pour rehausser encore la protection de la vie privée des utilisateurs, les niveaux de risque sont quantifiés à une précision de 4 bits avant leur échange. Nous faisons remarquer que c'est là une quantité d'information comparable à la transmission à 3 bits du niveau de risque du protocole de Google-Apple [13] et de la note sur les symptômes signalés par l'utilisateur que le projet coEpi joint aux rapports TCN [45].

4.2.2 Données envoyées au serveur d'apprentissage automatique

Comme le décrit la section 2.4, les utilisateurs peuvent choisir de fournir plus de données aux fins de la recherche. Pour les utilisateurs participants, deux types de données sont envoyées au serveur d'apprentissage automatique par l'entremise d'un réseau de mélange ou d'un mandataire afin que le serveur d'apprentissage automatique ne puisse pas savoir de quelle personne les données proviennent. Le premier type est un paquet de données pseudonymisées contenant des renseignements sur la santé ainsi que des détails sur les contacts récents. Le second type est un paquet de géolocalisation distinct qui est dissocié des renseignements au sujet de l'utilisateur que contient le paquet de données pseudonymisées.

Les paquets de données de géolocalisation consistent en un identificateur de lieu correspondant à une zone de dissémination contenant plusieurs centaines de personnes (section 2.6.1), auquel sont jointes des métadonnées (comprenant le niveau de risque et la zone de dissémination du domicile de l'utilisateur). Ces données servent à élaborer des cartes épidémiologiques (section 2.6.1) qui peuvent aider les autorités de santé publique locales à localiser les zones dans lesquelles la maladie se concentre ou se dissémine plus rapidement, ou dans lesquelles la plupart des contacts dangereux se produisent. La géolocalisation brute n'est utilisée localement sur le téléphone que pour déterminer l'identificateur de l'emplacement de la zone de dissémination et pour calculer les facteurs de risque associés à cet emplacement (certains quartiers peuvent par exemple connaître plus de cas). Les zones de dissémination ne sont pas du tout envoyées dans les paquets de données pseudonymisées envoyés au serveur d'apprentissage automatique ; c'est plutôt le facteur de risque dérivé localement sur le téléphone à partir de la zone de diffusion qui est inclus. Il sert ensuite d'entrée dans le prédicteur d'apprentissage automatique.

Les dossiers individuels anonymisés fournis par les utilisateurs participants sont conservés sur un serveur sécurisé doté de protections adéquates pour une base de données de dossiers médicaux pseudonymisés. Comme indiqué à la section 2.6.2, les données agrégées sont communiquées par COVI aux responsables de la santé publique. Ces renseignements agrégés comprennent à la fois ces cartes épidémiologiques et les paramètres du modèle décrit à la section 3.

4.3 Le simulateur comme modèle génératif

Nous allons tenter de résumer certaines des variables aléatoires les plus importantes qui interviennent dans le simulateur épidémiologique. Tout d'abord, le simulateur échantillonne ces variables dans l'ordre selon lequel elles se produiraient dans le temps. Chaque événement peut se produire à un moment précis ou selon un intervalle précis (comme une journée).

En plus des variables aléatoires associées à un moment précis, il y a des variables indépendantes du temps, comme les réponses aux questions que l'utilisateur peut fournir lors de l'installation de l'application. Celles-ci peuvent comprendre, par exemple, l'âge, le sexe biologique et les problèmes de santé préexistants qui pourraient avoir une incidence sur l'évolution de la maladie, mais aussi des questions sur leur domicile (combien de gens y vivent), leur travail (p. ex. s'ils travaillent avec des patients atteints de la COVID-19) et leur comportement (p. ex. s'ils portent un masque). Ces variables statiques peuvent être révisées après l'installation, mais on les considère comme étant des propriétés statiques de la personne. Dans le simulateur, ces variables sont échantillonnées à partir d'une distribution préalable P (statique) initialement basée sur les statistiques connues puis, lorsque les données tirées de l'application seront disponibles, des moyennes de ces réponses à l'échelle de la population. Les déplacements de tous les membres de la population forment un autre type de variable de base. Ces déplacements ne sont pas directement disponibles pour une modélisation, mais ils peuvent être extraits d'autres sources de données sur la mobilité afin de construire un modèle de la manière dont les gens se déplacent typiquement et à quelle fréquence, de leur domicile à leur travail, à l'hôpital, à des magasins, etc. Réunissons ces déplacements dans la variable de la mobilité et appelons leur distribution P(mobilité). L'échantillonnage issu de la mobilité donne des trajectoires hypothétiques pour des personnes hypothétiques qui passent du temps en différents endroits dans la simulation. Ci-dessous, nous constatons que la mobilité d'une personne peut dépendre de la conscience du risque d'être contagieux (parce que la personne ressent des symptômes ou que son téléphone l'avertit de prendre plus de précautions). La politique gouvernementale en matière de santé publique peut aussi influencer la mobilité (p. ex. en permettant ou non à certains types de lieux de fonctionner normalement). Le modèle de mobilité a donc la forme P(mobilité | conscience du risque, politique de santé publique). Lorsqu'ils passent quelques minutes l'un près de l'autre, un *contact* est déclenché. Le simulateur a donc un processus pour capter la distribution conditionnelle P(contacts | mobilité), ce qui revient essentiellement à déterminer quand deux personnes passent cinq minutes ou plus à proximité. Les attributs particuliers des contacts peuvent comprendre leur durée et la distance à laquelle les deux personnes se trouvaient. Certains de ces contacts seront maintenant enregistrés par les gens qui ont un téléphone doté de l'application COVI et constitueront certaines des observations disponibles pour l'entraînement (sans la position précise ni le moment exact de la rencontre, afin de protéger la vie privée des personnes).

Les variables latentes les plus importantes que modélise le simulateur sont le statut d'infection et la contagiosité de chaque personne, chaque jour de la simulation. Une personne peut avoir quatre états : susceptible (non encore infectée), exposée (infectée, mais non encore infectieuse), infectieuse (c.-à-d. contagieuse) ou rétablie (y compris les cas malheureux de décès). Tandis que la personne est infectieuse, une variable continue que nous appelons « contagiosité » transmet de l'information au sujet de la capacité à transférer le virus à une autre personne (ce qui peut être dû à la charge virale et au comportement, comme la toux, etc.). Le simulateur modélise l'évolution temporelle de ces variables latentes comme *P(*infection | contacts, statique), c.-à-d. dépendant des événements de contact et des problèmes médicaux de la personne.

Ces variables chronologiques observées sur le téléphone appartiennent à deux grandes catégories : d'un côté, les symptômes et résultats de tests, qui peuvent être entrés n'importe quel jour, et de l'autre, les observations associées aux *contacts*, p. ex. la caractéristique bruitée et corrompue de la durée et de la distance du contact (avec une estimation de la précision spatiale, car celle-ci peut être différente selon le type de capteur utilisé et les conditions de mesure) et le niveau de risque envoyé par l'autre téléphone quelque temps après le contact lui-même, faisant de tout cela une forme asynchrone de réseau bayésien dynamique fondé. Appelons les

symptômes et résultats de tests entrés des *observations médicales.* Le simulateur a donc un modèle *P(*observations médicales | infection) qui peut être échantillonné séparément pour chaque personne.

Le modèle probabiliste graphique que nous esquissons comprend également plusieurs boucles de rétroaction qui passent par le prédicteur de risque lui-même. Tout d'abord, les sorties du prédicteur de risque sont utilisées pour envoyer des messages à d'autres téléphones qui font partie des variables discutées plus haut et sont observés sur le téléphone destinataire. Deuxièmement, la simulation peut tenir compte de l'influence de l'application sur le comportement en produisant la *conscience* d'être potentiellement contagieux de la personne, ce qui crée une autre boucle de rétroaction avec *P(*mobilité | conscience du risque, politique de santé publique).

Pour résumer, les variables exogènes du système sont la politique de santé publique et les variables statiques relatives à chaque personne. À partir d'elles et des états initiaux de la simulation (une certaine proportion de personnes susceptibles, exposées, infectieuses et rétablies, ainsi que la contagiosité de celles qui sont infectées), la simulation peut se dérouler dans le temps pour plusieurs semaines ou mois relativement à une population hypothétique ayant certaines caractéristiques de mobilité et de densité (qui peuvent être liées à l'environnement, qu'il soit urbain ou rural, par exemple). Durant la simulation, les variables abstraites suivantes sont mises à jour quotidiennement dans cet ordre et suivant la logique ci-dessus :

*P(*mobilité (*t*) | conscience du risque (*t* – 1), politique de santé publique)
*P(*contacts (*t*) | mobilité (*t*), messages (t – 1))
*P(*infection (*t*) | contacts (*t*), statique)
*P(*observations médicales (*t*) | infection (*t*))
*P(*niveau de risques (*t*) | données téléphoniques (*t*))
*P(*messages (*t*) | niveau de risques (*t*))
*P(*conscience du risque (*t*) | niveau de risques (*t*))

où les données téléphoniques (*t*) représentent la séquence des observations disponible dans le téléphone pour calculer les niveaux de risque. Cela comprend les renseignements statiques, les messages reçus, la partie observée des contacts et les observations médicales observées (lorsqu'elles le sont).

Bien sûr, même au cœur d'une pandémie, la propagation d'une maladie est hyperlocalisée. Une seule personne peut causer toute un ensemble de cas d'infection [85]. Même si les modèles épidémiologiques simulés peuvent eux-mêmes prédire de tels événements sporadiques, ils ne peuvent dire aux responsables de la santé publique où précisément ces événements vont survenir. Pour cette raison, COVI fournira aux autorités de la santé publique une carte thermique agrégée des niveaux de risques et événements d'infection (voir la section 2.6.2 pour plus de détails).

4.4 Variables observées et latentes, prédicteur et simulateur génératif

Les données pseudonymisées collectées sur les téléphones serviront à entraîner le prédicteur de risque, mais serviront aussi à entraîner le simulateur épidémiologique ci-dessus, que l'on peut voir comme un modèle génératif duquel on peut échantillonner de nouveaux historiques synthétiques de contacts et de contagion, y compris dans l'avenir. Ce simulateur serait utile pour que les responsables de la santé publique et les épidémiologistes puissent mieux comprendre la propagation de la maladie. Les modules du simulateur sont des distributions conditionnelles (comme ci-dessus, ou ventilés en niveaux de détails plus raffinés) qui caractérisent les événements comme la transmission du virus d'une personne à l'autre, l'apparition de symptômes, le résultat d'un test, ou la façon dont les personnes modifient leur comportement à cause des messages qu'affichent l'application. La structure et les paramètres de ces distributions conditionnelles, après leur ajustement aux données observées, contiendront également des renseignements précis pour la compréhension épidémiologique de la maladie (notamment sur l'importance et les interactions de facteurs particuliers, comme

la distance, la durée ou le port du masque durant une rencontre avec quelqu'un) et au sujet de la réaction des gens aux recommandations.

Appelons le prédicteur de risque Q et le simulateur épidémiologique P par analogie avec les méthodes variationnelles et particulièrement avec les auto-encodeurs variationnels [86], une correspondance étayée ci-dessous. Pour comprendre comment le prédicteur de risque et le simulateur épidémiologique interagissent et s'aident l'un l'autre, il est important de comprendre que le simulateur peut échantillonner des valeurs pour deux types de variables aléatoires : celles (appelons-les X = données téléphoniques) qu'on peut observer sur les téléphones individuels (comme l'occurrence de contacts, les symptômes signalés ou les résultats de tests) et celles qu'on ne peut pas directement observer, nommées variables latentes (comme la charge virale, la contagiosité ou l'état d'infection réelle d'une personne). Appelons les variables latentes non observées, d'une manière ou d'une autre, Z. Le simulateur P est en réalité un modèle génératif $P(X, Z)$ pour la distribution combinée de X et Z dans le temps et parmi les utilisateurs.

Pour une bonne partie de la discussion ci-dessous, nous mettrons l'accent sur les données observées par chaque téléphone, puisque ce sont les seules données dont dispose le prédicteur de risque fonctionnant sur cet appareil. Les variables latentes Z pour le téléphone d'Alice renferment des quantités importantes sur l'état sous-jacent d'Alice, comme si elle est infectée, depuis quel moment et quelle est sa contagiosité depuis lors. Ces variables ne sont pas observées directement, mais elles caractérisent le risque qu'Alice représente pour les autres, et une forme quantifiée de sa contagiosité (le niveau de risque) sera envoyée par Alice aux téléphones des personnes (comme Bob) qu'elle a préalablement rencontrées, comme décrit ci-dessus. Le prédicteur de risque Q est en réalité un estimateur de probabilité $Q(Z \mid X)$ et, en général, on devrait le voir comme étant une approximation de $P(Z \mid X)$.

Pour les modèles génératifs non triviaux comme $P(X, Z)$, typiquement paramétrés comme $P(X, Z) = P(Z \mid X) P(Z)$, il n'y a pas de tractabilité computationnelle au calcul précis de $P(Z \mid X)$. En d'autres mots, alors qu'il est facile de simuler des échantillons et des trajectoires à partir de P, étant donné les paramètres et la structure de P, inverser ce processus pour récupérer le Z latent selon le X observé est difficile. Il existe plusieurs approches d'apprentissage automatique pour réaliser cette approximation, aussi appelée inférence. Peu importe la solution choisie, nous appelons le prédicteur résultant $Q(Z \mid X)$. Nous tiendrons pour acquis qu'on peut tirer un échantillon de $Q(Z \mid X)$ et que si des paires (X, Z) sont observées, on peut les utiliser pour entraîner $Q(Z \mid X)$ par une procédure d'optimisation quelconque, c.-à-d. une méthode d'entraînement supervisé.

4.4.1 Entraîner le prédicteur de risque et le simulateur ensemble

Nous avons donc besoin d'apprendre à la fois un modèle génératif P (faisant ici partie du simulateur épidémiologique) et une méthode approximative Q (le prédicteur de risque) qui détermine les variables Z non observées pour un utilisateur précis, étant donné les données X pour cet utilisateur. Si nous connaissions le P réel, nous pourrions simplement en échantillonner des paires (X, Z) et entraîner Q par apprentissage supervisé (typiquement par maximum de vraisemblance régularisée). Comme première étape d'élaboration d'un prédicteur, nous avons, en fait, bâti un simulateur épidémiologique P fondé sur les statistiques médicales et les statistiques sur la mobilité, et généré un vaste ensemble de trajectoires (p. ex. 30 jours pour une population de 30 000 dans une ville), ce qui a produit de nombreuses paires (X, Z). Nous pouvons donc utiliser ces paires pour entraîner un premier prédicteur de risque qui est fidèle au simulateur, et c'est le prédicteur inclus dans la première version de l'application, avant la collecte de données sur le téléphone.

Lorsque les données sont collectées, nous obtenons une collection de registres X (un par téléphone doté de l'application, pour chaque jour de la collecte de données). Nous pouvons alors utiliser notre prédicteur $Q(Z \mid X)$ pré-entraîné pour obtenir des échantillons des Z correspondants. Une stratégie possible pour entraîner P est donc d'utiliser ces Z inférés et échantillonnés avec les X correspondants observés comme ensemble

d'entraînement de paires (*X, Z*) pour mettre *P* à jour. Cela mènerait à un *P* différent, plus fidèle à la distribution des données sur *X*, et nous devrions alors ré-entraîner *Q* pour qu'il se conforme au nouveau *P* (si *Q* est entraîné plutôt qu'une procédure fixe). Itérer cette procédure est essentiellement l'algorithme de veille-sommeil [87]. Malheureusement, rien ne garantit qu'il optimiserait une fonction objective bien définie. Une variante moderne de cette idée a été introduite dans les méthodes variationnelles amorties comme l'auto-encodeur variationnel [86]. On peut écrire une borne supérieure à la log-vraisemblance log *P(X*) qui implique à la fois *P* et *Q* et peut être optimisée par des algorithmes du gradient. Pour le moment, c'est la méthode que nous préférons pour entraîner ensemble *P* et *Q.* Nous proposons aussi d'utiliser des échantillons tirés de *P* (qui peuvent être en dehors des observations en *X*) pour enrichir l'entraînement de *Q* (d'une manière similaire à l'algorithme veille-sommeil). Cela devrait contribuer à relever le défi représenté par le fait que la distribution des données n'est pas fixe, puisque la société évolue rapidement pour faire face à la pandémie. En gardant *Q* conforme aux connaissances épidémiologiques, non seulement autour des points de données *X*, mais plus généralement, le système devrait être plus robuste.

4.4.2 Inférence variationnelle

On peut doter la log-vraisemblance des données *X* observées de la borne inférieure de cet objectif variationnel, *L*, que l'on appelle aussi limite inférieure prévue, ou ELBO. Ainsi, la maximisation conjointe de *L* à la fois sur *P* et *Q* s'est avérée très efficace pour modéliser les données observées *X* lorsque nous pensons qu'elles sont mieux expliquées en invoquant les variables latentes *Z*, ce qui est exactement notre situation. Les méthodes variationnelles classiques optimisaient *Q* séparément pour chaque *X* donné, alors que les méthodes d'inférence amorties modernes comme l'auto-encodeur variationnel [86] paramétrisent *Q(Z|X*), p. ex., comme un réseau neuronal, et permettent une inférence plus rapide au moment où les données sont fournies. L'optimisation itérative se fait plutôt hors ligne lors de l'entraînement de *Q*, alors que le simple calcul de la sortie du réseau neuronal *Q(Z|X*) est très rapide et peut être effectué sur un téléphone, ce qui rend l'inférence variationnelle amortie attrayante pour l'entraînement du prédicteur de risque de l'application.

L'ELBO peut être optimisée par les algorithmes du gradient habituels. Le flux de calcul typique se déroule comme suit. Compte tenu des données disponibles sur un téléphone *X* jusqu'à une certaine date, on échantillonne les variables latentes *Z* du prédicteur de risque *Q(Z|X*). Ensuite, on calcule le logarithme de la probabilité jointe log *P(X*, *Z*) ainsi que log *Q(Z|X*). Enfin, on peut rétropropager l'ELBO dans *P* afin d'en mettre à jour les paramètres et dans *Q* pour en mettre à jour les paramètres. La seule complication vient du fait que *Z* inclut à la fois des variables continues (les valeurs de contagiosité pour chaque jour) et des variables discontinues (les événements de contagion). La rétropropagation à travers les variables continues est facile et efficace grâce à l'astuce du reparamétrage [88, 86]. La propagation des informations de gradient par l'échantillonnage de variables discontinues est légèrement plus complexe, la technique la plus simple impliquant une approximation de Monte-Carlo du gradient obtenu en échantillonnant (une ou plusieurs fois) les variables discontinues et en utilisant les logarithmes de probabilité de *P* et de *Q* comme signaux de renforcement (de manière similaire à ce qui est fait dans l'algorithme REINFORCE [89] et ses variantes). Des méthodes plus sophistiquées, comme la méthode softmax de Gumbell [90], permettent d'obtenir des estimateurs plus efficaces au prix d'un léger biais dans le gradient.

Notons que *Q* n'est qu'une approximation de l'information stockée dans *P*, de sorte que ses paramètres ne sont pas vraiment libres, car ils doivent être cohérents à la fois avec *P* et avec les données observées. Dans notre cas, *P* sera assez compact et aura peu de liberté, de sorte que la complexité de *Q* découle de la nécessité de compiler, dans le réseau neuronal, non seulement les connaissances sur *P,* mais aussi la manière d'effectuer le type d'inversion ou d'inférence nécessaire pour passer des données aux variables latentes (alors que *P* passe des variables latentes aux données).

De nombreuses questions sont toujours ouvertes en ce qui concerne l'amélioration de cette conception initiale. La tension entre la précision de la prédiction du risque et la protection de la vie privée nous force à développer

de nouvelles méthodes d'apprentissage automatique. Le fait de ne pas connaître le graphe de contact complet, par exemple (pour des raisons de confidentialité, mais également parce que les simulations s'échelonneraient difficilement si nous devions représenter le graphe de contact de tous les Canadiens), nous oblige à imaginer de nouvelles façons d'effectuer des inférences, potentiellement à différentes échelles spatiales. Notre mise en œuvre actuelle élude cette question en ne faisant qu'une inférence au niveau des individus, mais il faut travailler davantage pour tirer des conclusions à plus grande échelle sur la propagation du virus par un réseau d'individus.

Bien sûr, d'autres techniques de relaxation dans le graphe (pour communiquer les informations sur les risques par l'entremise du réseau téléphonique et converger vers un consensus) ou d'inférence pourraient être plus efficaces que celles qui sont proposées ici. Grâce au fait que la probabilité de transmission virale est minime pour un contact précis, lorsque le risque d'un utilisateur augmente de manière importante (p. ex. en raison d'un résultat de test positif), l'information se diffuse à travers le réseau et s'éteint exponentiellement rapidement (atteignant rapidement l'effet zéro en raison de la quantification des niveaux de risque). Les futurs travaux sur l'aspect de l'apprentissage automatique devront manifestement inclure des comparaisons entre différentes approches. Les mesures utilisées doivent également dépendre de l'application. À court terme, nous utilisons le logarithme de la probabilité du modèle, la précision et le rappel des personnes infectées par l'entremise de leur niveau de risque.

Toutefois, en fin de compte, la meilleure façon d'évaluer les différentes méthodes est de voir dans quelle mesure elles parviennent à freiner la propagation du virus pour une quantité donnée de mobilité moyenne. Comment devrait-on mesurer cela ? In silico, une telle mesure peut être obtenue en effectuant une simulation avec le prédicteur dans la boucle du comportement et en évaluant l'impact de celui-ci sur la mobilité et la propagation du virus, ainsi que la meilleure façon de parvenir à un compromis entre le nombre d'heures hors isolement (p. ex. au travail) et le maintien du taux de reproduction R en dessous de 1. Une autre question en suspens est de déterminer comment adapter les simulations à de grandes populations. Les simulations actuelles ont été effectuées pour jusqu'à 30 000 personnes. Comment les adapter à la taille d'un pays comme le Canada ? Bien que les effets locaux soient les plus importants pour comprendre la propagation et la croissance du virus, le déconfinement soulèvera des questions sur la propagation du virus entre les régions et les pays, et nous avons besoin d'outils de calcul appropriés à cette échelle.

4.4.3 Confidentialité des messages sur le risque

Comme les utilisateurs envoient des messages de risque à tous leurs contacts, il peut y avoir des fuites de données dans les messages de risque eux-mêmes. Cela est vrai, peu importe le protocole/mécanisme utilisé pour le traçage des contacts (pour l'API Google-Apple, il s'agit d'un message de risque sur 3 bits [13], pour la proposition CoEpi, il s'agit d'une liste de symptômes autodéclarés [45]), et est endémique à l'envoi de messages de risque. Dans un cas extrême, à des fins illustratives, si le message de risque incluait le NAS de l'utilisateur, le protocole ne respecterait pas du tout la vie privée, car les utilisateurs diffuseraient effectivement leurs identifiants avec un certain décalage par l'entremise du système de messagerie privé. Même dans des scénarios plus réalistes, les messages de risque eux-mêmes peuvent contenir des renseignements corrélables. Par exemple, si Alice et Bob se croisent tous les jours et que la plupart des gens ne sont pas infectés, Alice recevra de Bob le même niveau de risque pour une série d'événements de contact, ce qui lui permettra de déduire que tous les messages de Bob viennent de lui. Bien que nous cachions ces informations aux utilisateurs de l'application, un utilisateur malhonnête avec une application piratée pourrait, potentiellement, enregistrer les messages réels.

Pour cette raison, l'espace des messages de risque possibles doit être suffisamment petit pour que chaque utilisateur puisse nier de manière plausible avoir été l'expéditeur du message de risque. C'est pour cette raison que nous quantifions les niveaux de risque à 4 bits, en ayant des niveaux de risque entre 1 et 16. Avec seulement 16 niveaux de risque, Bob peut nier de manière plausible être l'expéditeur d'un message de risque.

Nous avons donc utilisé les manières suivantes pour réduire la quantité d'information disponible pour l'inférence et l'entraînement du prédicteur d'apprentissage automatique afin d'accroître la protection de la vie privée :

- Nous avons éliminé l'accès à la localisation et à la trajectoire exactes de chaque personne (la géolocalisation est plutôt convertie en statistiques de la zone de dissémination, qui sert d'entrée pour le prédicteur de risque, et les statistiques de chaque zone sont collectées pour le modèle épidémiologique) ;
- Nous avons supprimé l'accès à l'heure exacte des rencontres (pour protéger la vie privée de l'autre personne), en ne conservant que la date ;
- Nous avons éliminé les renseignements sur la structure globale du graphe des contacts (qui a rencontré qui) : il ne reste que la vue de chaque téléphone sur les contacts non identifiables spécifiques à ce téléphone et leur niveau de risque ;
- Nous avons éliminé la connaissance exacte nécessaire pour permettre de faire correspondre sans ambiguïté différentes rencontres comme appartenant à un même contact (nous avons plutôt des informations bruitées dérivées du moment approximatif de la réception des messages de mise à jour) ;
- Nous ne sommes pas en mesure de modéliser la distribution jointe des zones géographiques et des questionnaires médicaux par utilisateur, en fractionnant les données envoyées au serveur d'apprentissage automatique en deux types de paquets allant dans des fichiers différents dont les entrées ne peuvent plus être jumelées (l'un est indexé par zones, et l'autre par identifiants pseudonymisés).

Passons en revue ce qui est perdu avec chacun de ces éléments et comment nous atténuons ces problèmes. La situation géographique est sans doute importante, car certaines zones ont une probabilité de base d'infection plus élevée dans la population et, par exemple, une densité, des aspects culturels, etc. qui influencent la probabilité de contacts dangereux. Nous atténuons ce problème en utilisant les informations démographiques de Statistique Canada comme approximation de la localisation géographique réelle. Cela peut, en fait, contribuer à obtenir de meilleures généralisations dans les zones pour lesquelles moins de données sont collectées.

L'heure exacte des rencontres pourrait nous renseigner sur les circonstances de la contagion (p. ex., au travail pendant la journée ou pendant la nuit en partageant un lit), mais il s'agit d'une information trop sensible lorsque nous voulons éviter la stigmatisation (identifier les personnes rencontrées qui présentent un risque élevé). De même, le graphe de contact est crucial du point de vue de la protection de la vie privée (la plupart des gens ne veulent pas que quelqu'un piste les personnes qu'ils ont rencontrées). Il serait intéressant de voir comment cela affecte la précision des prédictions, mais nous avons choisi de ne pas envisager de méthodes (comme la *propagation des conviction par boucle*) qui nécessiteraient le graphe de contact complet pour effectuer des inférences et de l'apprentissage.

La capacité à faire correspondre différents événements de contact comme appartenant à la même personne (sans nécessairement savoir qui était cette personne) est importante pour estimer correctement les probabilités d'infection. Pour l'illustrer, envisageons deux scénarios : sur une période de quelques jours, Alice a N = 100 rencontres de 15 minutes chacune avec Bob (ils vivent ensemble), tandis que dans un autre scénario, Alice a N = 100 rencontres de 15 minutes chacune, mais cette fois, chaque rencontre concerne une personne différente. Dans le premier cas, en première approximation, la probabilité qu'Alice soit infectée peut augmenter pour se rapprocher de celle de Bob. Dans le second cas, la probabilité qu'Alice ait été infectée est très élevée, car il suffit que l'une des 100 personnes ait été contagieuse pour qu'elle ait une chance significative d'avoir été infectée. C'est pourquoi il est très important de faire la différence entre les contacts répétés et non répétés. Bien que le protocole de communication ne permette pas de faire une association exacte entre les différents contacts, il est possible d'obtenir une association probabiliste. Nous avons conçu un algorithme de groupement basé sur les niveaux de risque et l'heure d'arrivée des messages de mise à jour pour regrouper les différents contacts en blocs correspondant hypothétiquement à la même personne. Nous utilisons ces étiquettes bruitées comme

entrée supplémentaire pour le prédicteur. La séparation des données de questionnaires et de localisation a fait l'objet d'une discussion à la section 2.6.

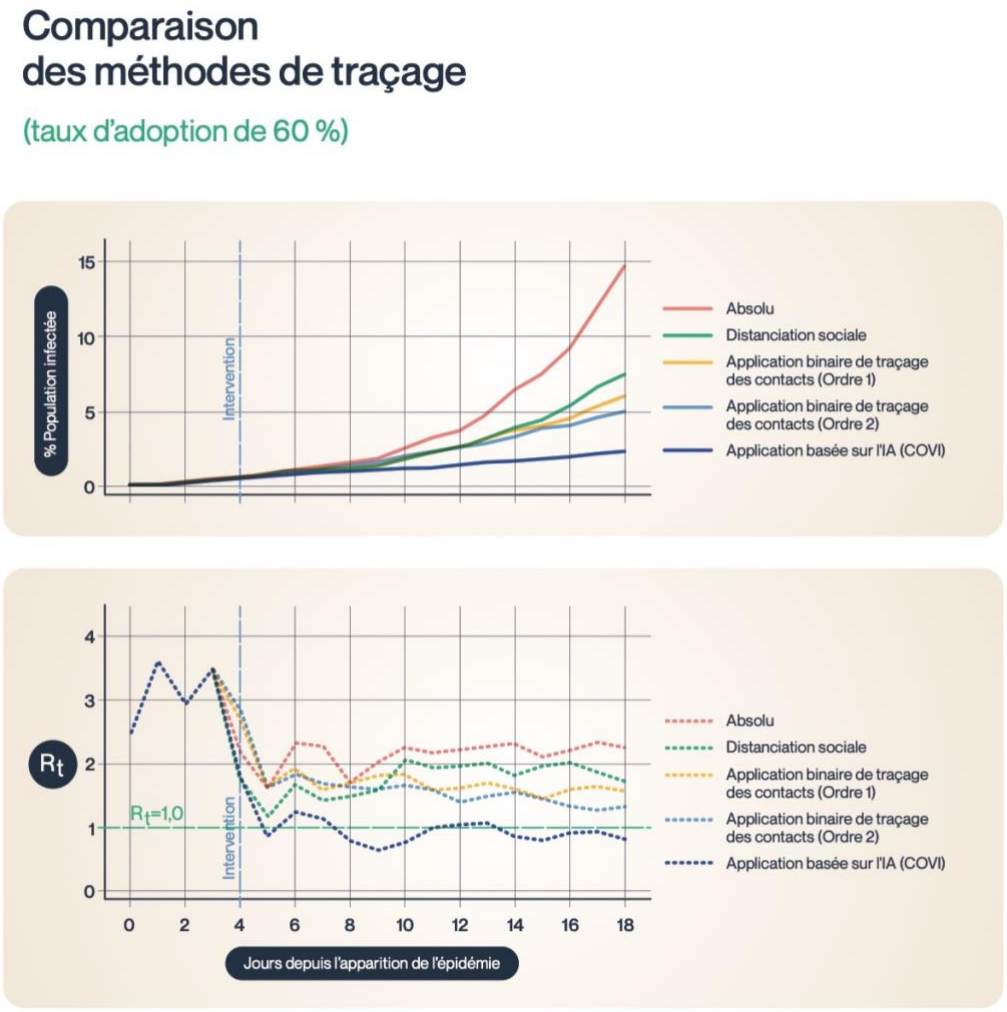

Figure 1 : Comparaison de quatre scénarios différents : non atténué (comportement préconfinement), distanciation sociale (même politique de mobilité pour tous les agents), traçage numérique binaire (méthode standard utilisée dans les applications de traçage numérique sans IA), et application basée sur l'IA mettant en œuvre une version simple du prédicteur d'apprentissage automatique de COVI (basée sur un petit modèle *Transformer*).

En haut : estimation du facteur de reproduction (R_t) au fur et à mesure de l'avancement de la simulation. Nous constatons un gain substantiel en R_t en utilisant le traçage sous une forme ou une autre, mais un gain beaucoup plus important avec la prédiction du risque basé sur l'apprentissage automatique. En bas : évolution du nombre de cas accumulés par rapport au nombre de jours écoulés. Le jour de l'intervention (jour 4) est le jour où les différentes politiques de mobilité sont mises en place.

4.5 Résultats préliminaires sur l'impact de l'apprentissage automatique

Nous avons entraîné un prédicteur basé sur l'apprentissage automatique de manière supervisée à l'aide des données générées par le simulateur, en utilisant une heuristique simple de traçage de contact pour générer les messages. Afin d'obtenir des résultats préliminaires mesurant l'impact de l'utilisation de l'apprentissage automatique pour prédire la contagiosité et obtenir des messages sur le niveau de risque, nous avons ensuite utilisé ce prédicteur à l'intérieur du simulateur pour influencer le comportement des agents selon quatre niveaux de recommandations associés à différents seuils de niveaux de risque. Cela nous a permis de simuler (avec des graines (seeds) aléatoires différentes de celles utilisées pour générer les données d'apprentissage) l'impact du prédicteur d'apprentissage automatique sur le facteur de reproduction R_t de la maladie et la croissance des infections dans une petite population pilote de 1000 personnes (R_t est indexé selon le temps t, car il peut évoluer dans le temps, en fonction des recommandations envoyées aux citoyens par les autorités de santé publique et leur application). La simulation a été réalisée en supposant que 60 % de la population utilisait l'application et que les utilisateurs au niveau de recommandation le plus élevé (niveau de quarantaine) subissaient un test.

Notons que le nombre d'entrées du prédicteur est variable, car il dépend du nombre de contacts. C'est pourquoi nous avons utilisé une architecture d'apprentissage profond de type *Transformer* [91], qui peut également capter les dépendances dans l'ensemble de l'historique de l'utilisateur (14 jours et la liste de tous les contacts) sans souffrir du problème de la disparition du gradient qu'ont les réseaux neuronaux récurrents [92]. La contagiosité prédite par le *Transformer* pour chacun des 14 derniers jours est convertie en un message de 16 niveaux (4 bits) envoyé aux contacts du jour correspondant. La conversion de la sortie réelle au niveau de risque a été faite en choisissant des seuils rendant les 16 cases de fréquence approximativement égale. Le niveau de risque pour aujourd'hui a été converti en un niveau de recommandation. Nous avons utilisé quatre niveaux de recommandation dans cette simulation :

1. Niveau de recommandation 1 (niveaux de risque 0 et 1) : encouragement au lavage des mains, activation du facteur d'hygiène pour réduire l'infection.
2. Niveau de recommandation 2 (niveaux de risque 2 et 3) : comme le niveau 1, avec en plus le port d'un masque et se tenir à 2 mètres des autres. Cela permet de maintenir le facteur d'hygiène. Cela active également le port d'un masque en dehors de la maison. L'efficacité du port du masque est définie différemment selon que la personne est un travailleur de la santé ou non (car les premiers ont tendance à avoir de meilleurs masques).
3. Niveau de recommandation 3 (niveaux de risque 4 et 5) : comme le niveau 2, avec en plus la pratique d'une plus forte distanciation sociale, qui réduit de moitié la durée des contacts. Cela empêche également les personnes de se rendre dans des endroits qui n'ont pas déjà été visités, ce qui les rend plus prudentes.
4. Niveau de recommandation 4 (niveaux de risque 6 à 15) : il s'agit du niveau de quarantaine ; comme le niveau 3, avec en plus la recommandation de subir un test, en fonction de la disponibilité du test. Les personnes mises en quarantaine travaillent à domicile si elles le peuvent et restent chez elles sauf si elles sont hospitalisées (il peut donc y avoir de nouvelles infections à domicile). La probabilité qu'elles sortent dans les magasins ou les parcs est réduite par un facteur de 10, mais chaque fois qu'elles sortent, elles n'explorent pas, c'est-à-dire qu'elles ne se rendent pas à plus d'un endroit.

Une simulation différente a ensuite été réalisée pour comparer quatre scénarios différents :

- Sans atténuation : les agents se comportent conformément aux statistiques de mobilité avant le confinement, ce qui donne une valeur de R_t légèrement supérieure à 2 et une croissance exponentielle rapide de la population infectée.
- Distanciation sociale : tous les agents respectent la même politique de mobilité, et un paramètre global (correspondant à la force des politiques de distanciation sociale) contrôle le nombre de contacts dangereux. Un réglage maximal de ce paramètre conduirait à un confinement total.

- Traçage numérique binaire (second ordre) : les agents se mettent en quarantaine s'ils ont été en contact avec une personne dont le test est positif, ou s'ils ont été en contact avec une personne qui a été en contact avec une personne dont le test est positif. Nos simulations donnaient de meilleurs résultats avec le second ordre comparé au premier ordre.
- Application basée sur l'IA : les agents utilisent le prédicteur de COVI pour moduler leur distanciation sociale et leur auto-isolement.

Le paramètre global contrôlant la force de la distanciation sociale a été modulé séparément pour les trois dernières méthodes afin d'égaliser la mobilité globale (le nombre de contacts). En effet, l'application basée sur l'IA aurait autrement tendance à être favorisée, car plus de personnes auraient généralement tendance à obtenir une forme de recommandation d'être prudentes, alors que le traçage numérique binaire ne concerne que les contacts immédiats (ou de second ordre) des patients dont le test a été positif. Les résultats sont présentés dans la figure 1, et suggèrent que la prédiction du risque basée sur l'apprentissage automatique pourrait substantiellement réduire le R_t, comparativement au traçage numérique standard et à une politique uniforme de distanciation sociale.

Cette visualisation met l'accent sur l'avantage de l'IA en termes de réduction du nombre de cas pour un certain niveau général fixe de mobilité, mais il est possible de montrer comment, pour un choix fixe de R_t (obtenu en diminuant globalement la mobilité avec une distanciation et un isolement accru), on peut obtenir plus de mobilité avec un prédicteur basé sur l'apprentissage automatique. Plus de détails sur cette simulation ainsi que sur le code utilisé seront fournis dans un prochain rapport technique qui se concentrera sur les aspects de simulation et d'apprentissage automatique de ce projet.

5 Responsabiliser les citoyens

La prise de décision est difficile, particulièrement en temps de crise où le rapport signal-bruit et l'impact potentiel des décisions sont plus importants que dans la vie de tous les jours. Des approches telles que l'économie comportementale postulent que cette situation complexe conduit les gens à commettre des erreurs prévisibles [93]. Les décideurs politiques emploient donc des stratégies de « coup de pouce » pour « réparer » ces erreurs et aligner le comportement des gens sur certaines normes. Cette stratégie est certainement pertinente dans certains contextes : par exemple, une consigne invitant la population à rester à la maison ou des instructions quant au lavage de mains pendant 20 secondes [94].

Cependant, si des approches comme l'économie comportementale [95] sont souvent utilisées pour atteindre des objectifs de changement de comportement prescrits par les organisations et les gouvernements, elles comportent également un risque : l'efficacité d'une stratégie basée sur le paternalisme libertaire (une caractérisation habituelle de la stratégie du « coup de pouce ») dépend des motivations sous-jacentes de la population à laquelle elle est appliquée. À court terme, les objectifs imposés de l'extérieur peuvent être promus par des messages clairs, cohérents et prescriptifs ; néanmoins, pour garantir un succès à long terme, il faut établir un lien avec les motivations et les préférences de chaque utilisateur.

Ainsi, lors d'une longue période de crise, s'assurer que les citoyens peuvent accéder à des informations fiables et les comprendre, en internalisant ce qui est le plus pertinent pour leur propre situation, ne suffit pas ; il faut plutôt s'assurer qu'un retour d'information subsiste dans l'échange d'informations. Le public doit donc avoir la possibilité d'exprimer ses objectifs et ses motivations, et ces objectifs et motivations doivent en fait servir à orienter les éléments axés sur l'information qui constituent l'objectif initial des messages. Cette approche offre aux utilisateurs une expérience qui leur permet de se faire entendre et de percevoir les informations qui leur sont fournies comme un soutien à leur épanouissement personnel plutôt que comme une contrainte extérieure.

Jusqu'à présent, dans cette crise, les citoyens ont été invités (ou, dans certains cas, forcés) à réduire fortement leurs activités quotidiennes habituelles, avec des mesures de confinement étendues et strictes adoptées et appliquées par différents ordres de gouvernement. À ce jour, la sévérité de ces mesures a rendu la prise de décision assez simple : les restrictions sont si importantes et généralisées que les citoyens ont une marge de manœuvre minimale pour interpréter ce qui leur est demandé.

Toutefois, à mesure que la crise atteindra de nouvelles phases, cette situation changera considérablement. Les services rouvriront progressivement, d'abord de manière limitée, puis plus rapidement (avec de possibles ajustements pour tenir compte de nouveaux foyers d'éclosion). Les types de décisions auxquelles le public sera confronté durant cette phase de la crise seront probablement plus exigeants sur le plan cognitif, car les recommandations appropriées dépendront davantage de la situation précise de chacun (par opposition à une approche plus large). De même, l'acceptation des mesures prises jusqu'à présent aura affaibli dans une certaine mesure la capacité de décision de la population. Dans l'ensemble, les citoyens seront à la fois moins certains de ce qu'il faut faire et moins motivés à le faire. Si une approche motivée de l'extérieur est toujours en place (par opposition à un modèle de réalisation de soi), des tensions croissantes entre le désir interne de liberté personnelle et la nécessité apparemment externe de santé et de sécurité publiques pourraient se manifester. Les implications de décisions quotidiennes, même minimes peuvent être profondes (par exemple, « Devrais-je rencontrer un ami dans un café ? »).

Nous devons donc élaborer minutieusement des outils qui permettront aux citoyens de prendre ces décisions sur la base de données probantes qui tiennent compte (1) de leur niveau de risque potentiel pour les autres, (2) de leur vulnérabilité en cas d'infection et (3) de leurs préférences en matière de risque. Plutôt que de s'appuyer sur des pressions coercitives, COVI favorise la prise de conscience, la responsabilisation et la réalisation de soi. L'application sert à instaurer un climat de transparence afin que chaque utilisateur puisse prendre la meilleure décision possible à la fois pour lui-même et pour les personnes qui l'entourent. Cette transparence est dictée par les préférences (notamment les préférences en matière de risque) que les utilisateurs expriment par leur utilisation de l'application. Les sections ci-dessous décrivent ce que cela implique concrètement, en présentant les principes fondamentaux qui orientent les décisions de conception de base et les considérations éthiques.

5.1 Préférences des utilisateurs pour une expérience d'un bout à l'autre

Alors qu'une approche strictement épidémiologique de la crise de COVID-19 pourrait privilégier la minimisation du risque encouru par les citoyens, la nécessité de protéger également à tout prix les libertés civiles et le bien-être économique exige que tout outil visant à servir les citoyens à long terme tienne compte de leurs préférences personnelles. Concevoir des outils en partant de l'hypothèse que chaque utilisateur vise uniquement à minimiser son risque d'infection pourrait conduire à une transmission de messages qui seraient ignorés par les utilisateurs.

Étant donné que l'un des principes directeurs de COVI est l'engagement à formuler des recommandations concrètes, il est important que ce message soit transparent, pertinent, engageant, responsabilisant et qu'il réponde aux préférences exprimées par l'utilisateur. Pour atteindre les principaux objectifs de santé publique, les utilisateurs doivent considérer COVI comme un outil qui les aide à se réaliser plutôt que comme un outil permettant à des acteurs extérieurs de leur dire quoi faire. Cette section examine la manière dont les divers mécanismes et cadres de mesure doivent être mobilisés pour obtenir les préférences personnelles, ce qui est essentiel pour atteindre les objectifs de santé publique.

COVI est fondé sur l'hypothèse de la capacité à agir personnellement, c'est-à-dire que les recommandations sont conçues pour éclairer le choix plutôt que pour le modifier (une distinction subtile, mais critique). Nous adoptons une approche fondée sur les faits et nous basons l'expérience de l'utilisateur sur des principes de clarté et de collaboration. Les cadres de meilleures pratiques et les instruments de collecte de données

énumérés ci-dessous donnent un aperçu de la manière dont ce principe s'applique à tous les aspects de la conception et de l'évolution de l'application.

5.1.1 Mesurer les comportements des utilisateurs dans l'application

En mesurant soigneusement les types d'informations avec lesquelles un utilisateur interagit, nous pouvons créer une expérience plus significative pour lui. Les recherches antérieures en science cognitive des applications mobiles liées à la santé nous ont appris que l'attention, la mémoire et le traitement des récompenses sont essentiels pour déterminer des modèles de prestation efficaces pour les recommandations en matière de santé [96]. Les données sur l'engagement recueillies à l'aide d'outils d'analyse dans l'application et affichées sur un tableau de bord interne de l'engagement nous permettent de tirer parti d'une combinaison de cadres de science comportementale établis dans les domaines de la santé publique [97], de la gestion comportementale des produits [98] et du changement de comportement éthique [99] pour ajuster l'architecture informationnelle de l'application et personnaliser l'expérience de l'utilisateur, en donnant la priorité aux informations qui intéressent l'utilisateur.

Il est important de noter que la divulgation de certains types d'informations serait inappropriée, même si celles-ci s'avèrent très attrayantes. Par exemple, une carte montrant en temps réel les utilisateurs infectés attirerait certainement beaucoup d'attention, mais la stigmatisation des utilisateurs infectés et les autres impacts négatifs de cette situation sur la société seraient énormes. L'information doit être personnalisée pour promouvoir l'engagement, sans toutefois que cela se fasse à n'importe quel prix. Les cadres éthiques décrits ci-dessous doivent donc être utilisés pour déterminer comment la poursuite de l'engagement des utilisateurs doit être équilibrée par rapport à d'autres considérations cruciales, telles que le bien-être psychosocial et l'inclusion.

Comme mentionné précédemment, les analyses intégrées (des utilisateurs consentants) contribuent aux algorithmes d'apprentissage automatique qui construisent des modèles épidémiologiques. La combinaison des outils de l'interface utilisateur et l'apprentissage automatique nous permet de fournir des messages adaptés aux différents groupes d'utilisateurs, tout en respectant les messages de santé publique des différentes autorités.

5.1.2 Déterminer les besoins non satisfaits grâce à des sondages auprès de la population

Lorsque l'on travaille avec de grands échantillons de données, il est facile de confondre la population de l'échantillon avec la population globale. Le principe démocratique de la mission directrice de COVI nous aide à garder cette différence en vue. Pour comprendre comment ces préférences des utilisateurs s'inscrivent dans la population générale, y compris les groupes sous-représentés, nous effectuons périodiquement une série d'exercices de collecte de données (tels que des sondages quantitatifs) avec des échantillons représentatifs de Canadiens. Par exemple, alors que les données en application peuvent montrer que les utilisateurs sont généralement satisfaits des options de confidentialité, les données de sondage pourraient révéler que la population en général est plus susceptible d'interpréter l'application comme une réduction de l'autonomie - ce qui a été démontré comme réduisant l'efficacité des messages des applications liées à la santé au Québec et en Alberta [100].

Les instruments de sondage que nous avons conçus se concentrent sur l'identification des principales croyances, attitudes et besoins non satisfaits. Ils se concentrent également sur la façon dont les Canadiens perçoivent le risque, sur les mesures qu'ils envisagent de prendre (sur la base de cette perception du risque) et sur la façon dont ils consomment les informations pour orienter les gestes qu'ils poseront.

Des sondages préliminaires ont été menés à l'aide de MTurk, une plateforme fréquemment utilisée dans la recherche en sciences sociales. Bien que cette plateforme soit connue pour s'orienter vers les membres les plus jeunes, les plus instruits et les moins riches de la société, elle offre néanmoins un accès à une population au-delà des utilisateurs de notre application. En comparant les données démographiques de notre population de

répondants aux données démographiques du recensement disponibles auprès de Statistique Canada, les résultats de ces sondages sont ensuite pondérés pour refléter proportionnellement la population canadienne (et ses sous-groupes). Parmi les utilisateurs qui fournissent leurs données pseudonymisées aux serveurs d'apprentissage automatique de COVI, des ajustements similaires sont effectués (sur la base des informations démographiques qu'ils saisissent) afin de fournir une image représentative des Canadiens en général.

Pour nous assurer que les instruments de collecte de données saisissent suffisamment les besoins non satisfaits et les obstacles comportementaux potentiels, nous nous référons à la théorie des réseaux sociaux [101, 102]. Ces cadres conceptuels guident la conception des instruments de collecte de données tels que les sondages ou les questionnaires d'entrevues. Les résultats de cette recherche permettent de définir les priorités en matière de développement de fonctionnalités et de transférer les changements dans la feuille de route du produit. Les résumés de ces informations nous permettent également de nous connecter aux équipes de sensibilisation du public afin d'aligner leurs messages sur les préférences des Canadiens (à la fois pour augmenter l'adoption de l'application et pour avoir un impact positif sur les Canadiens qui cherchent à contribuer à la lutte contre la crise sans utiliser l'application).

5.1.3 Appliquer des listes de vérification de la facilité d'utilisation pour améliorer l'expérience

Un certain nombre de listes de vérification de la facilité d'utilisation et de l'expérience de l'utilisateur sont appliquées lors de la création et du test de l'application. Ces listes portent sur les composantes cognitives et affectives de la convivialité de l'application (103). Dans ce contexte, l'utilisabilité se concentre sur cinq éléments décrits par J. Nielson (104) : 1) l'apprentissage, c'est-à-dire la facilité avec laquelle les utilisateurs peuvent accomplir les tâches de base dès la première fois qu'ils découvrent la conception ; 2) l'efficacité, c'est-à-dire la rapidité avec laquelle ils peuvent accomplir les tâches une fois qu'ils ont appris la conception, 3) la mémorisation : lorsque les utilisateurs reprennent la conception après une période de non-utilisation, comment peuvent-ils rétablir facilement leurs compétences ? 4) les erreurs : combien d'erreurs les utilisateurs commettent-ils, quelle est la gravité de ces erreurs et comment peuvent-ils s'en remettre facilement ? 5) la satisfaction : dans quelle mesure est-il agréable d'utiliser la conception ? Des principes clés tels que l'évaluation heuristique sont utilisés pour détecter et supprimer tout défaut de conception potentiel qui entrave la convivialité.

En plus de garantir que les informations et les fonctionnalités sont compréhensibles par les utilisateurs, l'objectif de ces audits d'expérience utilisateur est de s'assurer que l'engagement des utilisateurs est maximisé. L'une de nos principales priorités pour améliorer l'expérience utilisateur est d'examiner comment nous pouvons responsabiliser les utilisateurs et exploiter leur motivation intrinsèque afin qu'ils restent engagés à long terme. Pour cela, nous nous appuyons sur la recherche en matière de motivation humaine, comme la théorie de l'autodétermination, l'auto-efficacité et l'autorégulation (105, 106, 107), pour nous assurer que l'expérience utilisateur est conçue pour un changement de comportement à long terme et motivé par l'utilisateur.

Au-delà de l'application des meilleures pratiques en matière d'expérience utilisateur, nous travaillons également avec un certain nombre de partenaires industriels qui sont des leaders dans la création d'interfaces, et nous fournirons des conseils et des validations sur les nouvelles versions de l'application ainsi que sur les nouvelles fonctionnalités qui sont en cours de déploiement. Ces engagements reflètent le fait que l'état de l'art dans la création d'interfaces utilisateurs (et c'est vraiment autant de l'art que de la science) n'est pas documenté de manière approfondie dans les articles de recherche ou autres sources écrites. Au contraire, la pointe d'expérience utilisateur n'est parfois visible que dans le travail et les idées de ses principaux praticiens, c'est un domaine de pratique professionnelle d'abord et de recherche ensuite.

5.2 La compréhension de l'utilisateur est priorisée et vérifiée plutôt que supposée

Comme indiqué tout au long de ce document, la protection de la vie privée est une considération essentielle pour le projet COVI - tant parce qu'elle est une priorité sociale et démocratique que parce qu'elle joue un rôle déterminant dans l'adoption de l'application. Les recherches ont montré que les autorisations de partage de données sont fortement influencées par l'expérience en matière de respect de la vie privée, l'angoisse relative à l'informatique et la perception des facteurs de contrôle qui, s'ils ne sont pas pris en compte, peuvent avoir un effet profondément négatif sur les préoccupations d'un segment important d'utilisateurs en matière de respect de la vie privée [108]. Bien qu'il existe des exemples de la manière dont les préjugés cognitifs tels que le surchoix et l'actualisation hyperbolique peuvent être exploités dans la conception de la plateforme pour contraindre le consentement (109), nous opérons avec la ferme conviction qu'il s'agit d'une stratégie contraire à l'éthique et non durable.

Pour cette raison, il est impératif que toutes les autorisations de partage soient soigneusement élaborées et validées tout au long de l'expérience de l'application. Pour démontrer de manière transparente les implications de l'utilisation de COVI sur la vie privée, et pour garantir que les citoyens qui décident d'utiliser l'application le fassent en toute confiance, l'application ne peut être conçue ou mise en œuvre d'une manière qui suppose qu'un utilisateur dispose d'un temps, d'une attention ou d'une compréhension infinie pour explorer les implications sur la vie privée.

Pour valider que la conception permet effectivement le consentement éclairé, des techniques et des cadres fondés sur des données probantes doivent être utilisés pour comprendre dans quelle mesure les mécanismes de divulgation réussissent à informer les utilisateurs sur les principales caractéristiques - en particulier autour du concept de vie privée et de partage des données. En particulier, les fonctionnalités liées aux tests d'utilisabilité mentionnées ci-dessus sont incluses pour mesurer les choix faits par les utilisateurs.

5.2.1 Test au niveau de la population sur le partage des préférences

De nombreuses recherches montrent que les préférences des Canadiens en matière de partage des données sont fondées sur des convictions discutables quant aux politiques de collecte et d'utilisation des données (110). Afin de révéler les véritables préférences au niveau de la population, la recherche doit être menée de manière à exposer clairement les politiques d'utilisation des données, à en assurer la compréhension et à susciter des réponses franches. Ainsi, en plus de se référer à la littérature existante sur le sujet (111), nos sondages cherchent à comprendre les préférences des Canadiens en matière de vie privée et de partage des données. Il est important de noter que les tests nous permettent d'identifier les principales lacunes dans les connaissances que les Canadiens peuvent avoir sur ce que le partage des données peut impliquer dans le contexte d'une application de suivi numérique de contacts. Les caractéristiques et la rédaction des textes dans l'application ont été et continuent d'être influencées par les thèmes généraux identifiés dans cette recherche concernant les préférences des Canadiens en matière de vie privée et de partage des données.

5.2.2 Les conditions clés sont soulignées et progressivement divulguées

De nombreuses recherches sur l'ergonomie cognitive soutiennent l'idée que les gens ont une bande passante cognitive limitée et se rabattent sur l'heuristique pour simplifier leur prise de décision lorsqu'ils interagissent avec des systèmes complexes [112]. En fait, il existe une série de preuves suggérant que la rareté cognitive est en corrélation avec un comportement de divulgation d'informations plus important (54). Cela signifie que le consentement pourrait probablement être contraint en surchargeant les utilisateurs (qui sont probablement déjà sous pression émotionnelle) d'informations, mais une telle approche est contraire aux principes éthiques guidant COVI et probablement non viable.

Nous devons plutôt nous assurer que les éléments clés des conditions générales sont bien compris par les utilisateurs, et non pas simplement acceptés au hasard. Pour ce faire, nous utilisons une approche de divulgation progressive à plusieurs niveaux, qui s'est avérée équilibrer l'expérience des utilisateurs et la transparence du système (113). Par exemple, une couche supérieure illustrant les implications en matière de protection de la vie privée peut être reliée à une deuxième couche un peu plus textuelle ; celle-ci peut ensuite être reliée à la section plus longue de la FAQ sur le site web, qui à son tour renvoie les utilisateurs à la politique de protection de la vie privée dans son intégralité.

Les utilisateurs reçoivent ainsi des informations sur la divulgation en fonction de leur niveau d'intérêt et de connaissance du sujet. Ceux qui sont satisfaits d'une vue de haut niveau en reçoivent une, tandis que les utilisateurs plus intéressés peuvent continuer à approfondir de plus en plus les détails jusqu'à ce qu'une réponse soit apportée à leurs questions.

5.2.3 La compréhension de l'utilisateur est vérifiée plutôt que supposée

Le consentement est souvent obtenu en présentant aux utilisateurs un bloc de texte de « conditions générales » et en supposant qu'ils le liront et le comprendront. Or, cette hypothèse a été invalidée par des preuves empiriques ; en fait, un consentement donné dans de telles conditions ne peut être considéré comme pleinement éclairé. Afin de rendre le consentement plus significatif, il est essentiel de savoir quelles informations affectent le plus les décisions des utilisateurs, et de s'assurer que ces informations sont transmises de manière à ce qu'ils soient le plus susceptibles de les lire et de les comprendre.

Nous prenons un certain nombre de mesures pour y parvenir. Tout d'abord, nous appliquons l'analyse interne à l'application pour estimer la compréhension des utilisateurs, par exemple en examinant le taux moyen d'abandon des utilisateurs à différentes couches d'informations divulguées. Deuxièmement, nous soumettons des questionnaires de compréhension dynamique à un échantillon aléatoire d'utilisateurs, ce qui nous permet de comprendre quelles informations ont été ou non internalisées. Enfin, les outils de divulgation sont révisés de manière itérative sur la base du retour d'information de ces mesures, afin de s'assurer qu'ils répondent au mieux au comportement réel des utilisateurs.

5.3 La responsabilisation des utilisateurs pour se protéger et protéger les autres est maximisée

Comme indiqué au début, l'un des principaux défis de l'assouplissement des mesures de confinement sera la perte d'instructions claires et uniformes concernant le comportement des citoyens. Afin d'assurer une clarté continue et de promouvoir l'autonomisation des utilisateurs, l'application COVI s'appuie sur des méthodes éprouvées. En particulier, la recherche sur les messages de santé publique suggère que deux facteurs sont particulièrement prédictifs du sentiment de maîtrise de soi : la conscience de la santé (déclenchée par la volonté de se protéger et de se dépasser) et les connaissances en matière de santé (déclenchées par le capital social qui lie et relie les individus).

[55]. En ciblant explicitement ces résultats dans ses messages, COVI peut accroître la collaboration en vue d'obtenir des résultats favorables en matière de santé publique parmi ses utilisateurs.

En plus de tirer parti des cadres de messages de santé publique qui optimisent l'autonomisation, un aspect unique de COVI est sa capacité à fournir des informations personnalisées aux utilisateurs sur les mesures qu'ils peuvent prendre en toute sécurité en fonction de leur contexte individuel. Comme indiqué ci-dessus, ces informations sont personnalisées en fonction des préférences de l'individu (afin de promouvoir l'épanouissement personnel, ce qui est différent de la conformité) et alignées sur la politique définie par les autorités de santé publique (qui restent les décideurs légitimes, même si COVI offre des outils puissants pour informer ces politiques et les transmettre aux citoyens).

Grâce à un algorithme prédictif basé sur l'apprentissage automatique, COVI fonctionne avec des niveaux de risque scalaires, plutôt que de simples binaires de contact/pas de contact avec une personne infectée. La nature prédictive et scalaire de l'algorithme facilite à son tour une approche proactive et progressive permettant aux utilisateurs de gérer leur risque en limitant leurs mouvements par degrés à mesure que leur niveau individuel de risque d'infection augmente, avant même qu'ils aient la certitude d'être infectés ou non (ou d'avoir eu un contact direct avec une personne infectée).

5.3.1 Les croyances et les attitudes des Canadiens face à la crise sont suivies de près

Pour s'assurer que les messages correspondent aux préférences des utilisateurs (pour qu'ils soient responsabilisants plutôt que perçus comme imposés), l'évolution des croyances et des attitudes des Canadiens face à la crise (y compris les dimensions sanitaires, sociales et économiques) doit être suivie de près. Pour ce faire, diverses approches sont utilisées de concert.

Des sondages à grande échelle auprès de la population canadienne ont été et continuent d'alimenter notre compréhension de la façon dont les Canadiens perçoivent cette crise, de leurs préférences, etc. Ils sont complétés par nos données internes, qui (comme indiqué ci-dessus) sont pondérées par le recensement afin de refléter la population canadienne.

Ces instruments s'appuient sur des recherches menées dans divers domaines, notamment les messages de santé, les communications de crise et la recherche sur les préférences révélées/énoncées dans le domaine des sciences du comportement. Des recherches antérieures ont démontré que les messages de santé ont un fort effet de cadrage, notamment en montrant que les messages axés sur les bénéfices sont plus efficaces que les messages axés sur les pertes pour promouvoir des comportements de prévention (114). Il existe actuellement peu de recherches sur les effets de cadrage des messages liés à la COVID-19, une lacune que les recherches liées à COVI cherchent à combler. Les recherches de pointe en matière de communication de crise ont mis en évidence l'importance de donner la priorité à la transparence, à la confiance et à la responsabilisation des utilisateurs (55). Les théories et les cadres relatifs aux préférences sont examinés dans la section ci-dessus et alimentent également les outils décrits ici, en particulier à mesure que la base de connaissances s'accroît et que la crise évolue (ainsi que les perceptions de la crise).

Ces résultats nous aident à formuler des variantes de messages et à concevoir des fonctionnalités, qui font ensuite l'objet d'un test A/B dans l'application pour déterminer les effets des différentes approches envisagées. Nous y parvenons en utilisant notre moteur d'analyse et nos questions personnalisés dans l'application (qui sont intégrés dans l'algorithme d'apprentissage automatique sur les serveurs d'apprentissage automatique de COVI Canada, comme indiqué ci-dessus), ainsi que par des sondages externes.

5.3.2 La désensibilisation lors de l'utilisation de l'application est prise en compte

Un effet bien connu des communications de crise est la désensibilisation aux messages dans le temps, autrement dit la « fatigue de l'alerte ». En fait, Baseman et ses collègues [115] ont constaté que chaque message de santé publique supplémentaire envoyé au cours d'une semaine entraînait une diminution statistiquement significative de 41,2 % des chances de se souvenir du message. La désensibilisation est également connue pour diminuer la perception du risque, ce qui peut affecter le respect des recommandations par les utilisateurs. Étant donné que certains messages disponibles via l'application sont critiques (par exemple, une prescription de rester à la maison basée sur votre risque d'infection, ou la vulnérabilité à la COVID-19 en raison d'une condition préexistante), il est important que ces messages soient pris au sérieux par les utilisateurs. En bref, nos messages internes doivent être soigneusement structurés selon des pratiques éprouvées [116], en veillant à ce que les alertes de faible priorité ne créent pas de bruit qui empêche les alertes de haute priorité d'être prises au sérieux et de faire l'objet d'une action. Il y a un équilibre délicat à trouver ici, car le désir de stimuler l'implication à l'égard

de l'application (surtout au début) pourrait créer une pression pour accroître l'urgence des messages de l'application, mais ultimement, il faut laisser de la place au-delà des messages de base pour que les messages urgents se distinguent.

Pour cette raison, l'effet des messages de l'application sur la désensibilisation des utilisateurs doit être évalué en permanence. C'est pourquoi, il faut réaliser des sondages auprès des utilisateurs (y compris des microsondages à question unique diffusés dans l'application à des moments soigneusement choisis), des sondages auprès de la population et des analyses dans l'application (en examinant les taux, le caractère opportun et la conformité aux alertes de priorité faible ou élevée). La conception des sondages s'appuie sur des cadres de communication de crise, des messages publics et des messages sur la santé.

5.3.3 Communiquer de façon crédible

Les recherches sur la communication de crise et des risques [117] ont montré que les gens recherchent des informations simples, cohérentes et crédibles. Compte tenu de ces préférences connues, les utilisateurs sont susceptibles d'être extrêmement sensibles à toute indication selon laquelle les informations présentées dans la demande sont biaisées, périmées ou fausses ; le caractère opportun et la crédibilité des informations doivent être identifiés de manière proactive et démontrés aux utilisateurs. Les experts et les chercheurs s'accordent également à dire que la communication des autorités au public devrait inclure des informations explicites sur les incertitudes liées aux événements ; le degré de certitude concernant les informations devrait donc être clairement communiqué. Une collaboration étroite avec les autorités pour garantir la validité et la fiabilité des informations est essentielle en tant que stratégie d'atténuation de ces incertitudes, la source des données devant être mise en évidence pour l'utilisateur.

5.3.4 L'information est mise à jour régulièrement et de façon visible

Nos recherches sur les applications numériques existantes relatives à la COVID-19 indiquent que les utilisateurs accordent beaucoup d'importance à la réception d'informations actualisées. C'est pourquoi il faut veiller tout particulièrement à ce que les informations disponibles dans l'application soient aussi à jour que possible et que le dernier horodatage de mise à jour soit toujours disponible et pertinent pour l'utilisateur. L'ensemble des fonctionnalités initiales de l'application intègre des API qui sont mises à jour quotidiennement, dans le but de rendre les données disponibles encore plus rapidement. Alors que nous nous dirigeons vers des sources de données propriétaires pour les visualisations dans l'application, il est important de se coordonner avec les équipes de développement pour s'assurer que les horodatages sont (1) facilement disponibles et pertinents pour les utilisateurs, et (2) qu'ils sont significatifs pour les utilisateurs. Par exemple, nous communiquerons clairement si les horodatages font référence à la date d'infection, à la date d'exécution du test ou à la date de disponibilité du résultat du test.

5.4 Promotion du bien-être psychosocial des utilisateurs

En raison de la nature sensible du contenu communiqué par l'application - ainsi que du niveau général de stress causé par la crise - il est essentiel d'aborder à la fois la conception de l'application et la rédaction avec une grande empathie. Cela signifie notamment qu'il faut à tout moment réduire au minimum le stress excessif pour l'utilisateur et accorder une attention particulière aux groupes à risque élevé qui sont susceptibles de subir un stress encore plus important.

Un stress accru a un impact négatif sur la prise de décision. Ainsi, étant donné que l'objectif de COVI est de redonner le pouvoir aux utilisateurs par une meilleure prise de décision, ajouter du stress va directement à l'encontre de notre objectif. En outre, le bien-être psychosocial joue un rôle important dans la définition des récits individuels et collectifs sur nos efforts pour faire face à la crise, ce qui a des implications démocratiques

importantes (par exemple, les solutions du gouvernement et des organisations sont principalement considérées comme les principaux moyens de résoudre les problèmes d'action collective ; ce projet représente l'expérience d'une nouvelle forme décentralisée de coordination qui est axée sur la protection de la vie privée et l'autonomisation des citoyens, et qui peut avoir des répercussions sur les décisions futures concernant la manière de faire face aux défis sociétaux à grande échelle). Il existe également des raisons plus tangibles de promouvoir le bien-être psychosocial : un stress et une anxiété accrus sont liés à un fonctionnement immunitaire réduit [118, 119, 120] et un stress excessif diminuerait donc notre résilience biologique à la COVID-19.

Voici quelques-unes des stratégies que nous employons pour promouvoir le bien-être psychosocial.

5.4.1 Créer des fonctionnalités permettant aux utilisateurs d'évaluer les risques pour leur bien-être psychosocial

Pendant une pandémie mondiale, certaines personnes peuvent être moins attentives à leur bien-être mental, et plusieurs sont dans une position où elles doivent faire des compromis difficiles entre les préoccupations de santé et d'autres questions (financières, professionnelles, familiales/sociales, culturelles/religieuses). Même la santé mentale de personnes qui ne contractent pas le virus peut être compromise en raison d'une anxiété et d'un stress supplémentaires. En effet, les personnes ayant des problèmes de santé mentale préexistants font partie des populations les plus vulnérables en période de crise et d'isolement [121]. C'est pourquoi nous incluons des fonctionnalités dans l'application pour aider les individus à évaluer leur bien-être et fournissons des ressources en matière de santé mentale adaptées au profil de l'utilisateur (âge, lieu, etc.).

5.4.2 Fournir des ressources en matière de santé mentale

Comme indiqué, nous fournirons aux utilisateurs des ressources (par exemple, des exercices de thérapie psychologique et des services de soutien en santé mentale) pour les aider à progresser de manière significative vers un meilleur bien-être psychosocial. Dans les cas où les utilisateurs signalent des problèmes graves liés à la santé mentale, un système de triage simple devrait être créé qui les incite à contacter un praticien de santé mentale ou à demander un soutien immédiat. Comme la santé mentale reste quelque peu taboue dans la société, nous avons fait de la communication de la normalité des problèmes de santé mentale une priorité, en particulier dans une période comme celle-ci (par exemple en utilisant des normes sociales pour indiquer qu'un nombre X de personnes dans leur quartier a utilisé cette ressource).

5.4.3 Créer une distribution positive de la valeur émotionnelle (valence) des messages

Étant donné que la plupart des messages relatifs à la crise de COVID-19 sont extrêmement négatifs, il faut veiller à réduire le stress imposé par les messages ainsi formulés, d'autant plus que des recherches antérieures ont montré que la formulation des messages peut entraîner des changements importants dans la réaction du public [122]. Pour ce faire, il faut utiliser des tests utilisateurs pour mesurer la valeur émotionnelle des messages et s'assurer que la valeur des messages est formulée de manière positive et communiquée efficacement (tout en veillant à ce qu'ils ne soient pas interprétés à tort comme étant sans gravité). En fait, tous les messages utilisés dans l'application ont été adaptés pour éviter de susciter des sentiments négatifs et de faire peser une charge inutile sur l'état mental des utilisateurs. Chaque itération de l'application continuera à refléter les tests continus que nous effectuons à cet égard, afin de valider les approches et d'améliorer l'effet de l'application sur le bien-être des utilisateurs.

5.4.4 S'attaquer de manière proactive aux risques de stigmatisation et autres dynamiques sociales liées à la vie privée

L'application COVI a été conçue pour protéger la vie privée et responsabiliser ses utilisateurs. Si une attaque technique sur l'infrastructure est un vecteur important à prendre en compte pour les atteintes à la vie privée, le

téléphone de l'utilisateur doit également être considéré comme un point de vulnérabilité au sein du système. Bien que nous souhaitions fournir un contenu transparent et informatif aux utilisateurs, ce même contenu vu par une autre personne regardant son téléphone pourrait constituer un compromis insoutenable pour la vie privée. Par exemple, un niveau de risque affiché en évidence pourrait intéresser les utilisateurs, mais un commerçant ou un employeur pourrait également exiger de voir l'écran du niveau de risque des utilisateurs comme condition pour être admis sur le site. Comme le suggèrent Bruns et ses collègues [123], ce genre de dynamique peut créer de profonds risques de stigmatisation, qui doivent être traités par des stratégies d'atténuation des risques soigneusement planifiées et des normes sociales appropriées protégeant les droits des personnes. Il convient de noter que le consentement peut ne pas être un obstacle suffisant pour un employé qui n'a pas d'autre choix que d'accepter les directives de son employeur ou de perdre son emploi. Une protection juridique supplémentaire de la vie privée devrait être envisagée et mise en place en fonction des préférences collectives des citoyens et de l'intérêt de protéger les plus vulnérables de la société. Des mesures véritablement volontaires (par exemple, rester à la maison en cas de risque élevé) sont préférables, mais ne devraient pas avoir de coût personnel (par exemple, perdre son emploi).

Pour répondre à ces préoccupations, nous nous appuyons sur des scénarios de modèles de menace pour nous assurer que nous tenons compte des façons dont la vie privée d'un utilisateur pourrait être compromise par une personne qui a accès à ses informations par le biais de son propre téléphone. Ces modèles de menace nous permettent d'effectuer des évaluations de la vie privée structurellement analogues à celles créées par l'équipe chargée de l'infrastructure de protection de la vie privée.

5.5 L'inclusion des utilisateurs pour reconnaître la diversité de leurs besoins

La diversité et l'inclusion sont importantes dans le contexte de COVI pour plusieurs raisons. Tout d'abord, la question de la justice sociale et de l'équité. Toute application qui se présente comme un catalyseur d'un mouvement national doit soutenir tous les membres de la population, en donnant à chacun la possibilité d'agir et de se joindre à ce mouvement. Une application qui est moins accueillante ou moins efficace pour certains segments de la population compromet sa prétention à un effort véritablement national. En outre, comme l'utilité de l'application COVI augmente rapidement à mesure que la base d'utilisateurs se développe, la promotion de l'inclusion est importante pour des raisons instrumentales dans la mesure où elle augmente le pouvoir de servir l'ensemble de la base d'utilisateurs pour faire face à la crise.

Étant donné cette aspiration, il est important de considérer comment divers groupes, en particulier ceux qui sont déjà marginalisés, sont susceptibles d'interagir avec l'application. Pour ce faire, nous déployons des stratégies telles que les suivantes.

5.5.1 Audits d'inclusion au niveau de la population

L'équipe de COVI doit utiliser des instruments de collecte de données pour identifier les principales lacunes démographiques dans l'utilisation de l'application. Les dimensions mises en évidence comme étant pertinentes comprennent : le sexe, la race, l'âge, la langue, le revenu, l'éducation, le secteur d'emploi, la composition de la famille, la région, le milieu rural/urbain, le statut autochtone, la santé mentale/physique, la capacité mentale/physique, la situation du logement. Cela peut être réalisé en comparant les données démographiques des utilisateurs aux données de recensement de Statistique Canada. Ces informations seront partagées avec l'équipe chargée de la sensibilisation du public afin d'éclairer ses stratégies visant à atteindre une base d'utilisateurs diversifiée. Ils seraient également partagés avec les autorités de santé publique afin de les aider à comprendre quels sous-groupes de population sont ou ne sont pas bien représentés dans les données qu'elles reçoivent de COVI Canada (y compris les données agrégées et la modélisation épidémiologique).

La qualité des données est un élément essentiel pour le succès de cette approche. Les utilisateurs ne sont pas tenus de fournir ce niveau d'informations démographiques. Ils ont plutôt la possibilité de fournir autant ou aussi peu de ces dimensions qu'ils le souhaitent, sachant que plus ils en donnent, plus leur expérience est personnalisée. En outre, pour les utilisateurs qui choisissent de fournir leurs données au serveur d'apprentissage machine de COVI, leur contribution de données démographiques est essentielle pour l'entraînement des algorithmes d'apprentissage automatique et donc, pour fournir à ces utilisateurs individuels et à d'autres comme eux des recommandations précises.

5.5.2 Intégrer les dimensions de la diversité dans nos autres analyses

Au-delà de la simple compréhension de la façon dont la base d'utilisateurs se rapporte à la diversité des sous-populations au Canada, il est important d'identifier et d'évaluer toute différence significative dans la façon dont les sous-populations interagissent avec l'application COVI et en tirent profit. Par exemple, si les résidents urbains plus jeunes, plus instruits et plus aisés sont surreprésentés parmi les utilisateurs de l'application (un scénario plausible étant donné que ce groupe démographique est également surreprésenté parmi les propriétaires de téléphones intelligents), cela pourrait conduire à la construction d'un modèle épidémiologique mieux adapté à la réalité de certains utilisateurs que d'autres. Si une telle situation n'était pas identifiée et traitée, elle pourrait conduire à des recommandations moins précises à des utilisateurs différents de ce groupe - fournissant des protections de santé moins efficaces à certains utilisateurs qu'à d'autres.

Pour identifier et relever ces défis potentiels, des cadres d'évaluation des biais dans la recherche clinique sont mis à profit [124] et les analyses décrites dans ce document sont ventilées selon ces dimensions de la diversité. En outre, les dimensions de la diversité sont introduites dans l'algorithme d'apprentissage automatique pour identifier les différences épidémiologiques et comportementales entre ces sous-groupes. La procédure d'apprentissage pour la prédiction des risques et la modélisation épidémiologique peut ensuite être modifiée pour augmenter le poids des groupes sous-représentés, en utilisant une méthode d'échantillonnage par importance ou de pondération par importance. Par exemple, si les Autochtones ont des préférences différentes de celles des autres Canadiens dans la gestion de cette crise, et réagissent donc différemment aux messages, cela pourrait être identifié par l'approche analytique utilisée ici, afin de garantir que les Autochtones reçoivent des messages qui favorisent leur épanouissement personnel plutôt que de leur imposer des messages qui intègrent les préférences des autres.

5.5.3 Engagement auprès des populations à risque

Les analyses ci-dessus expliquent le processus d'évaluation de la sous-représentation démographique et d'adaptation de l'expérience des utilisateurs (y compris les recommandations en matière de santé) en fonction de la démographie. Toutefois, les équipes du projet partent du principe que certains groupes démographiques importants sont si fortement sous-représentés parmi les utilisateurs que cela crée un risque que l'ensemble du sous-groupe ne soit pas bien servi par COVI. De plus, si COVI représente un effort pour soutenir et responsabiliser les Canadiens en cette période de crise, alors les groupes à risque (par exemple, les Canadiens en situation d'itinérance ou vivant dans des situations de logement précaires) sont les plus susceptibles d'être systématiquement exclus de l'utilisation de COVI.

Pour ces raisons, il est important pour l'équipe de l'application d'ouvrir le dialogue avec les entités qui représentent les groupes de la population les plus menacés par la marginalisation. La marginalisation structurelle est prévisible parmi la population des personnes âgées, les personnes n'ayant pas accès aux téléphones intelligents ou aux données mobiles et les personnes handicapées, bien qu'une vérification plus approfondie basée sur les cadres établis pour l'innovation inclusive soit nécessaire [125]. Depuis le début, nous avons été proactifs dans l'idéation de solutions de rechange pour accéder aux groupes de personnes qui risquent d'être marginalisés. En outre, notre feuille de route comprend un ensemble de caractéristiques très spécifiques qui permettraient à l'application d'atteindre, bien que de manière plus limitée, les membres des groupes à risque -

par exemple, des portefeuilles de garde (*custodial wallets*) pour ceux qui ne possèdent pas de téléphone portable, mais qui y ont accès. Enfin, des recherches antérieures ont montré que l'implication de parties prenantes des communautés à risque en tant qu'experts du domaine dans la co-création de solutions peut améliorer considérablement la manière dont « les scientifiques des données abordent le développement de corpus et d'algorithmes qui affectent les personnes des communautés marginalisées et savoir qui impliquer dans ce processus » [126]. Ainsi, un engagement direct avec les communautés marginalisées est nécessaire tout au long du projet.

6 Discussion

Examinons maintenant brièvement certaines des critiques souvent formulées à l'égard du traçage de contacts et examinons comment COVI se comporte à cet égard. Les plus grandes critiques concernant le traçage de contacts numérique portent généralement sur la confidentialité, la confiance et l'adoption.

Une question importante qui se pose est de savoir s'il vaut la peine ou non de prendre quelque risque que ce soit avec la vie privée si nous ne sommes pas certains qu'une telle application serait utile. Des études suggèrent qu'environ la moitié de la population doit accepter de l'utiliser pour vaincre le virus, un seuil élevé pour l'adoption. Toutefois, s'il existe une chance qu'une application comme COVI puisse réussir à être utilisée à ce niveau, nous devons la prendre, car les conséquences de ne pas le faire sont trop importantes : la différence entre un facteur de reproduction de 1,5 et un facteur de reproduction de 0,9 est énorme en terme de pertes humaines, sans parler du fait que l'auto-isolement ciblé ouvre la possibilité qu'une grande partie de la population puisse jouir d'une plus grande liberté et puisse travailler tout en maintenant le virus à distance.

Bien sûr, dans un pays démocratique où nous valorisons la liberté et la responsabilité, nous ne pouvons pas rendre une application comme celle-ci obligatoire : la seule option est la confiance. Pour qu'un système comme COVI fonctionne, il faut que les gens fassent confiance à l'organisation qui la gère ; d'où l'importance de la protection de la vie privée et d'une organisation à but non lucratif axée uniquement sur la pandémie de la COVID-19 pour gérer les efforts. De même, COVI Canada et les gouvernements doivent avoir confiance que la plupart des citoyens agiront de manière responsable lorsqu'ils comprendront ce qui est en jeu, c'est-à-dire la vie et la santé de leurs concitoyens.

Cependant, même si l'utilisation de COVI n'était pas suffisante pour un traçage automatique efficace des contacts afin d'estimer les risques de contagion, elle renforcerait tout de même les efforts de traçage manuel, en permettant aux citoyens à haut risque (selon les estimations de COVI) d'entrer en contact de manière proactive avec les autorités de santé publique (avant d'être appelés par la santé publique, si tant est qu'ils l'aient été), gagnant ainsi de précieux jours pendant lesquels des contagions seraient susceptibles de se produire. En outre, COVI pourrait jouer un rôle important en matière de modélisation et les prévisions épidémiologiques. En effet, il suffit qu'une petite fraction de la population consente à partager ses données pour les modèles d'apprentissage automatique afin d'améliorer considérablement ce qui est actuellement possible pour la compréhension et la prévision épidémiologiques dans le cadre de différentes politiques de santé publique.

Sur la question de la vie privée, le problème présente de multiples facettes, et COVI en aborde les différents aspects de différentes manières. Une préoccupation majeure est la stigmatisation qui pourrait résulter de l'application. Nous veillons à ce que les tierces parties ou les personnes avec lesquelles un contact COVI est établi ne puissent généralement pas déduire facilement le niveau de risque d'une personne (à moins qu'il ne soit évident pour d'autres raisons). Passer du temps avec une personne diagnostiquée ne la marquerait pas comme un paria, car cette information resterait anonyme. Même la manière dont l'infectiosité est communiquée à l'utilisateur, c'est-à-dire sous la forme de recommandations qui peuvent dépendre d'autres facteurs, rend plus difficile pour quelqu'un comme un conjoint ou un employeur, simplement en regardant son téléphone, d'obtenir une lecture explicite du niveau de risque (par exemple, nous évitons le type de schéma de couleurs évident mis

en œuvre dans d'autres applications). COVI n'est pas destiné à être utilisé comme un passeport d'immunité, en raison des préoccupations que cela soulèverait pour les droits de la personne et la dignité.

Une autre considération importante en matière de protection de la vie privée est la crainte que les organismes gouvernementaux aient accès à la trajectoire détaillée et au réseau de contacts d'une personne. Là encore, cela est évité dans la mesure du possible, grâce à l'approche décentralisée de la gestion des données de risque, aux mécanismes cryptographiques utilisés pour envoyer les messages de risque, aux mécanismes de protection de la vie privée du côté de l'apprentissage automatique, et à la création d'un modèle de gouvernance des données solide (avec COVI Canada, une organisation à but non lucratif) pour conserver les informations pseudonymisées pertinentes sur le plan médical (comme les réponses aux questionnaires) pendant une période de trois mois et à une mission bien définie de protection de la santé, de la vie privée et de la dignité des citoyens en ce qui concerne la gestion des données recueillies. En outre, la nature pseudonymisées des données volontaires, même facultatives, rend difficile le suivi des personnes, car leur numéro de téléphone, leur adresse IP, leur nom ou d'autres informations permettant de les identifier ne seraient pas recueillis et ne seraient donc pas accessibles à quiconque.

Pour promouvoir la confiance, COVI Canada aura des règles ouvertes concernant sa gouvernance, le libre accès au code et aux modèles épidémiologiques agrégés, et sera continuellement surveillé par son conseil d'administration, des comités d'experts internes et des évaluations externes de groupes universitaires indépendants et de représentants gouvernementaux, afin de s'assurer qu'il reste fidèle à sa mission. L'ensemble du modèle de gouvernance de COVI Canada est construit autour des valeurs fondamentales de légitimité, de responsabilité, de transparence et d'efficacité. COVI se conforme aux lois canadiennes sur la protection de la vie privée et aux principes mis en avant dans la déclaration commune des commissaires à la protection de la vie privée fédéral, provinciaux et territoriaux du 7 mai 2020 [48]. Ces principes comprennent le consentement et la confiance, la conformité à la loi, la nécessité et la proportionnalité, la finalité, la dépersonnalisation, la durée limitée des mesures, la transparence, la responsabilité et le déploiement de garanties. Ce livre blanc public se veut un exemple de transparence : nous avons essayé d'être explicites sur les risques que la recherche des contacts comporte pour la vie privée, et nous espérons que les utilisateurs finaux seront d'accord avec notre proposition étant donné les circonstances exceptionnelles de la pandémie.

La mission unique de COVI Canada, qui consiste à soutenir les Canadiens dans leur lutte contre la COVID-19, et son caractère à but non lucratif garantissent que les données recueillies ne seront jamais utilisées à des fins commerciales ni vendues à des entreprises privées. Elles ne peuvent pas être utilisées pour la surveillance ou pour faire appliquer une quarantaine par les gouvernements. Les données sont toutes stockées au Canada et seront supprimées dès que la pandémie sera terminée. Le conseil d'administration de COVI Canada sera présidé par un juge canadien à la retraite, et le modèle de gouvernance comprend un conseil consultatif d'experts composé de leaders d'opinion reconnus dans des domaines pertinents tels que la santé publique, l'éthique, les droits de la personne et la protection de la vie privée. COVI adhère à la Déclaration de Montréal pour le développement responsable de l'IA et a été élaboré avec le soutien de l'UNESCO. Nous reconnaissons les conséquences malheureuses et inacceptables qu'une technologie pourrait avoir sur les groupes marginalisés et c'est pourquoi COVI Canada continuera de travailler avec les organisations des droits de la personne, les groupes de la société civile et les experts en droit et en sciences sociales pour prévenir les biais algorithmiques, renforcer l'accessibilité de la technologie et assurer une représentation inclusive à tous les niveaux de son modèle de gouvernance. COVI Canada sera démantelé à la fin de la pandémie, mais la science et la technologie développées pourront nous aider si des situations similaires se présentent dans le futur.

7 Conclusion

Au fur et à mesure que la pandémie progresse, un objectif important est de tirer parti des stratégies numériques pour minimiser la propagation de la COVID-19 tout en empêchant la violation de la vie privée et les intrusions aux libertés civiles. Alors que les ressources sanitaires et économiques sont fortement sollicitées, il est critique de pouvoir réduire efficacement et rapidement la propagation de la COVID-19 afin de réduire la morbidité et la mortalité associées à la maladie. Toutefois, à notre avis, cela ne doit pas se faire au détriment des libertés civiles qui sont au cœur des sociétés démocratiques. COVI est une solution numérique qui combine le traçage des contacts, la science de l'interface utilisateur et l'apprentissage automatique avec de solides protections de la vie privée tout en donnant le pouvoir à chacun d'agir en connaissance de cause.

Nous ne pouvons pas éliminer tous les risques et les compromis en matière de protection de la vie privée qui sont endémiques dans le domaine du traçage des contacts, mais en combinaison avec une gouvernance indépendante et à but non lucratif COVI vise à gagner la confiance du public et à contribuer à une action collective responsable contre la pandémie.

Nous considérons COVI comme une occasion de renforcer une forme de démocratie où le pouvoir est véritablement entre les mains des citoyens : ils décident d'utiliser ou non cette technologie, en évaluant les risques et les avantages pour eux-mêmes et leur communauté en fonction de leurs valeurs. Bien sûr, cela nécessite un débat public pour aider les citoyens à comprendre les enjeux et à se doter des outils qui correspondent le mieux à nos valeurs, et ce débat démocratique est une composante essentielle du plan de COVI Canada. Malgré les nombreux défis associés au lancement d'une telle stratégie, l'équilibre obtenu par COVI représente une étape importante pour faire progresser l'utilisation de la santé numérique et de l'apprentissage automatique afin de lutter contre une crise mondiale majeure. COVI donne le pouvoir aux individus en leur fournissant des informations factuelles et personnalisées sur leur niveau de risque, leur permettant ainsi d'agir en conséquence et de manière responsable pour protéger leurs proches et leur communauté. COVI fournit également aux services de santé publique des données agrégées qui peuvent être essentielles pour élaborer des politiques appropriées. Ultimement, nous pensons que COVI donnera aux Canadiens le pouvoir de se protéger, de limiter la propagation du virus et de faciliter une levée intelligente et sécuritaire des mesures de distanciation sociale par une action collective et démocratique dans leur vie quotidienne.

Remerciements

Nous aimerions remercier Sumukh Aithal, Behrouz Babaki, Henri Barbeau, Edmond Belliveau, Vincent Berenz, Olexa Bilaniuk, Amélie Bissonnette-Montminy, Pierre Boivin, Emélie Brunet, Joé Bussière, Gaétan Marceau Caron, René Cadieux, Pierre Luc Carrier, Hyunghoon Cho, Anthony Courchesne, Linda Dupuis, Justine Gauthier, Joumana Ghosn, Gauthier Gidel, Marc-Henri Gires, Simon Guist, Deborah Hinton, Bogdan Hlveca, Bernd Holznagel, Samuel Huberman, Shrey Jain, Jameson Jones-Doyle, Dilshan Kathriarachchi, Giancarlo Kerg, Soundarya Krishnan, David Lazar, Frédéric Laurin, Sacha Leprêtre, Stéphane Létourneau, l'équipe de Libeo, Alexandre Limoges, Danielle Langlois, Vincent Martineau, Lucas Mathieu, Philippe Matte, Rim Mohsen, Eilif Muller, Ermanno Napolitano, David Noreau, Ivan Oreshnikov, Satya Ortiz-Gagné, Jean-Claude Passy, Marie Pellat, Dan Popovici, Daniel Powell, Brad Rabin, Catherine Saine, Victor Schmidt, Shanya Sharma, Kareem Shehata, Pierre-Luc St-Charles, Marie-Claude Surprenant, Melisande Teng, Julien Tremblay-Gravel, David Wu et Lenka Zdeborova pour leur aide. Nous aimerions également remercier le Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG), les Instituts de recherche en santé du Canada (IRSC), CIFAR, le Fonds québécois de la recherche sur la nature et les technologies (FRQNT) et Scale AI pour leur financement. A.S. est financé par la bourse de chercheur clinicien (Junior 1) du Fonds de la recherche en santé du Québec, le prix Lucien McGill et le fonds de l'Initiative interdisciplinaire en infection et immunité de l'Université McGill. Y.W.Y. est financé par le Fonds d'action COVID-19 de l'Université de Toronto.

Références

[1] B. Gates, « Responding to covid-19 a once-in-a-century pandemic? » New England Journal of Medicine, 2020.

[2] N. Fernandes, « Economic effects of coronavirus outbreak (covid-19) on the world economy, » Disponible : SSRN 3557504, 2020.

[3] R. M. Anderson, H. Heesterbeek, D. Klinkenberg, et T. D. Hollingsworth, "How will country-based mitigation measures influence the course of the covid-19 epidemic?" The Lancet, vol. 395, no. 10228, pp. 931–934, 2020.

[4] E. T. Barometer, « January 20, 2019, » 2019. [En ligne]. Disponible : https://www.edelman.com/sites/g/files/aatuss191/files/2019-02/2019_Edelman_Trust_Barometer_Global_Report_2.pdf

[5] R. Niehus, P. Martinez de Salazar Munoz, A. Taylor, et M. Lipsitch, « Quantifying bias of covid-19 prevalence and severity estimates in Wuhan, China that depend on reported cases in international travelers, » 2020.

[6] U. Irfan, "The case for ending the COVID-19 pandemic with mass testing," April 2020. [En ligne]. Disponible : https://www.vox.com/2020/4/13/21215133/coronavirus-testing-covid-19-tests-screening

[7] J. Flint, S. Burton, J. Macey, S. Deeks, T. Tam, A. King, M. Bodie-Collins, M. Naus, D. MacDonald, C. McIntyre et al., "Assessment of in-flight transmission of sars–results of contact tracing, Canada." Canada communicable disease report= Relevé des maladies transmissibles au Canada, vol. 29, no. 12, p. 105, 2003.

[8] L. Ferretti, C. Wymant, M. Kendall, L. Zhao, A. Nurtay, L. Abeler-Dorner,¨ M. Parker, D. Bonsall, et C. Fraser, "Quantifying sars-cov-2 transmission suggests epidemic control with digital contact tracing," Science, 2020.

[9] D. Tang, « Contact-tracing strategies for sars-cov-2 eradication**** brouillon, » 2020.

[10] J. Bay, A. Tan, C. S. Hau, L. Yongquan, J. Tan, et T. A. Quy, « BlueTrace: A privacy-preserving protocol for community-driven contact tracing across borders, » 2020. [En ligne]. Disponible : https://bluetrace.io/static/bluetrace_whitepaper-938063656596c104632def383eb33b3c.pdf

[11] J. Chan, S. Gollakota, E. Horvitz, J. Jaeger, S. Kakade, T. Kohno, J. Langford, J. Larson, S. Singana-malla, J. Sunshine et al., « Pact: Privacy sensitive protocols and mechanisms for mobile contact tracing, » arXiv preprint arXiv:2004.03544, 2020.

[12] R. L. Rivest, J. Callas, R. Canetti, K. Esvelt, D. K. Gillmor, Y. T. Kalai, A. Lysyanskaya,

[13] Apple et Google, "Privacy-preserving contact tracing," April 2020. [En ligne]. Disponible : https://www.apple.com/covid19/contacttracing/

[14] "Pan-European Privacy-Preserving Proximity Tracing," April 2020. [En ligne]. Disponible : https://pepp-pt.org/

[15] N. Trieu, K. Shehata, P. Saxena, R. Shokri, et D. Song, “Epione: Lightweight contact tracing with strong privacy,” Avril 2020. [En ligne]. Disponible : https://sunblaze-ucb.github.io/privacy/projects/epione.html

[16] F. Ordonez, “Ex-officials call for $46 billion for tracing, isolating in next coronavirus package,” NPR, Avril 2020. [En ligne]. Disponible : https://www.npr.org/2020/04/27/845165404/ex-officials-call-for-46-billion-for-tracing-isolating-in-next-coronavirus-packa

[17] W. F. Flanagan, « Equality rights for people with aids: Mandatory reporting of hiv infection and contact tracing, » McGill LJ, vol. 34, p. 530, 1988.

[18] M. L. Levine, “Contact tracing for hiv infection: a plea for privacy,” Colum. Hum. Rts. L. Rev., vol. 20, p. 157, 1988.

[19] H. Cho, D. Ippolito, et Y. W. Yu, « Contact tracing mobile apps for covid-19: Privacy considerations and related trade-offs, » arXiv preprint arXiv:2003.11511, 2020.

[20] R. Raskar, I. Schunemann, R. Barbar, K. Vilcans, J. Gray, P. Vepakomma, S. Kapa, A. Nuzzo, R. Gupta, A. Berke et al., « Apps gone rogue: Maintaining personal privacy in an epidemic, » arXiv pré-impression arXiv:2003.08567, 2020.

[21] J. Burgess, “A contact-tracing procedure,” British Journal of Venereal Diseases, vol. 39, no. 2, p. 113, 1963.

[22] J. Millar, “A well-intentioned but unproven app could reinforce biases and create confusion and stress, something developers must take more time to consider,” Avril 2020. [En ligne]. Disponible : https://policyoptions.irpp.org/magazines/april-2020/five-ways-a-covid-19-contact-tracing-app-could-make-things-worse/

[23] D. K. Gillmor, “Principles for technology-assisted contact-tracing,” Avril 2020. [En ligne]. Disponible : https://www.aclu.org/sites/default/files/field_document/aclu_white_paper_-_contact_tracing_principles.pdf

[24] Inria, “Proximity tracing applications: The misleading debate about centralised versus decentralised approaches,” Avril 2020. [En ligne]. Disponible : https://github.com/ROBERT-proximity-tracing/documents/blob/master/Proximity-tracing-discussion-EN.pdf

[25] D. Meyer, “Controversy around privacy splits Europe’s push to build COVID-19 contact-tracing apps,” Fortune, April 2020. [En ligne]. Disponible : https://fortune.com/2020/04/20/coronavirus-contact-tracing-privacy-europe-pepp-pt-dp3t-covid-19-tracking/

[26] A. Greenberg, “Clever cryptography could protect privacy in Covid-19 contact-tracing apps,” Wired, April 2020. [En ligne]. Disponible : https://www:wired:com/story/covid-19-contact-tracing-apps-cryptography/

[27] A. Sharma, R. A. Harrington, M. B. McClellan, M. P. Turakhia, Z. J. Eapen, S. Steinhubl, J. R. Mault, M. D. Majmudar, L. Roessig, K. J. Chandross et al., “Using digital health technology to better generate evidence and deliver evidence-based care,” Journal of the American College of Cardiology, vol. 71, no. 23, pp. 2680–2690, 2018.

[28] C. Troncoso, M. Payer, J.-P. Hubaux, M. Salathe, J. Larus, E. Bugnion, W. Lueks, T. Stadler, A. Pyreglis, D. Antonioli, L. Barman, S. Chatel, K. Paterson, S. Capkun, D. Basin, J. Beutel, D. Jackson, B. Preneel, N. Smart, D. Singelee, A. Abidin, S. Guerses, M. Veale, C. Cremers, R. Binns, et C. Cattuto, “Decentralized privacy-preserving proximity tracing,” April 2020. [En ligne]. Disponible : https://github.com/DP-3T/documents/blob/master/DP3T%20White%20Paper.pdf

[29] J. Chen, J. Lieffers, A. Bauman, R. Hanning, et M. Allman-Farinelli, “The use of smartphone health apps and other mobile health (mhealth) technologies in dietetic practice: a three country study,” Journal of Human Nutrition and Dietetics, vol. 30, no. 4, pp. 439–452, 2017.

[30] R. Dandekar, S. G. Henderson, M. Jansen, S. Moka, Y. Nazarathy, C. Rackauckas, P. G. Taylor, et A. Vuorinen, « Safe blues: A method for estimation and control in the fight against covid-19, » 2020.

[31] B. J. Zikmund-Fisher, A. Fagerlin, et P. A. Ubel, “is 28% good or bad? evaluability and preference reversals in health care decisions,” Medical Decision Making, vol. 24, no. 2, pp. 142–148, 2004.

[32] E. Kangethe, V. Kimani, D. Grace, G. Mitoko, B. McDermott, J. Ambia, C. Nyongesa, G. Mbugua, W. Ogara, et P. Obutu, “Development and delivery of evidence-based messages to reduce the risk of zoonoses in nairobi, kenya,” Tropical animal health and production, vol. 44, no. 1, pp. 41–46, 2012.

[33] G. of Canada, “Coronavirus disease (COVID-19): Guidance documents,” 2020. [En ligne]. Available: https://www.canada.ca/en/public-health/services/diseases/2019-novel-coronavirus-infection/guidance-documents.html

[34] M. Hunger, L. Schwarzkopf, M. Heier, A. Peters, R. Holle, K. S. Group et al., “Official statistics and claims data records indicate non-response and recall bias within survey-based estimates of health care utilization in the older population,” BMC health services research, vol. 13, no. 1, p. 1, 2013.

[35] X. He, E. H. Lau, P. Wu, X. Deng, J. Wang, X. Hao, Y. C. Lau, J. Y. Wong, Y. Guan, X. Tan, X. Mo, Y. Chen, B. Liao, W. Chen, F. Hu, Q. Zhang, M. Zhong, Y. Wu, L. Zhao, F. Zhang, B. J. Cowling, F. Li, et G. M. Leung, “Temporal dynamics in viral shedding and transmissibility of covid-19,” Nature Medicine, 2020.

[36] S. A. Lauer, K. H. Grantz, Q. Bi, F. K. Jones, Q. Zheng, H. R. Meredith, A. S. Azman, N. G. Reich, et J. Lessler, “The incubation period of coronavirus disease 2019 (covid-19) from publicly reported confirmed cases: estimation and application,” Annals of internal medicine, 2020.

[37] K. Leung, J. T. Wu, D. Liu, et G. M. Leung, “First-wave covid-19 transmissibility and severity in China outside Hubei after control measures, and second-wave scenario planning: a modelling impact assessment,” The Lancet, 2020.

[38] C. J. Wang, C. Y. Ng, et R. H. Brook, “Response to COVID-19 in Taiwan: Big Data Analytics, New Technology, and Proactive Testing,” JAMA, 2020.

[39] P. Regan, “Legislating privacy,” 1995.

[40] D. J. Solove, “Understanding privacy,” 2008.

[41] M. Kundera, “The unbearable lightness of being,” 1984.

[42] M. J. Keith, J. S. Babb Jr, C. P. Furner, et A. Abdullat, « Privacy assurance and network effects in the adoption of location-based services: an iphone experiment. » in ICIS, 2010, p. 237.

[43] J. Q. Whitman, “The two western cultures of privacy: Dignity versus liberty,” Yale LJ, vol. 113, p. 1151, 2003.

[44] R. Bayer et A. L. Fairchild, « Surveillance and privacy, » 2000.

[45] T. Coalition, “TCN protocol,” April 2020. [En ligne]. Disponible : https://github.com/TCNCoalition/TCN

[46] « COVID Watch, » https://covid-watch.org/, 2020.

[47] H. K. Patil et R. Seshadri, “Big data security and privacy issues in healthcare,” in 2014 IEEE international congress on big data. IEEE, 2014, pp. 762–765.

[48] O. of the Privacy Commissioner of Canada, “Joint statement by federal, provincial and territorial privacy commissioners: Supporting public health, building public trust: Privacy principles for contact tracing and similar apps.” May 2020. [En ligne]. Disponible : https://www.priv.gc.ca/en/opc-news/speeches/2020/s-d 20200507/

[49] A. Cavoukian, “Privacy by design: the 7 foundational principles,” January 2011. [En ligne]. Disponible : https://www.ipc.on.ca/wp-content/uploads/Resources/7foundationalprinciples.pdf

[50] J. Penney, S. McKune, L. Gill, et R. J. Diebert, “Advancing Human Rights by Design in the Dual Use Technology Industry,” Columbia Journal of International Affairs, December 2018. [En ligne]. Disponible : https://jia.sipa.columbia.edu/advancing-human-rights-design-dual-use-technology-industry

[51] NHS, “Nhs covid-19 app,” April 2020. [En ligne]. Disponible : https://www.nhsx.nhs.uk/covid-19-response/nhs-covid-19-app/

[52] A. P. Gregg, B. Seibt, et M. R. Banaji, « Easier done than undone: asymmetry in the malleability of implicit preferences. » Journal of personality and social psychology, vol. 90, no. 1, p. 1, 2006.

[53] M. Friese, M. Wanke,¨ et H. Plessner, « Implicit consumer preferences and their influence on product choice, » Psychology & Marketing, vol. 23, no. 9, pp. 727–740, 2006.

[54] G. A. Veltri et A. Ivchenko, “The impact of different forms of cognitive scarcity on online privacy disclosure,” Computers in Human Behavior, vol. 73, pp. 238 – 246, 2017. [En ligne]. Disponible : http://www.sciencedirect.com/science/article/pii/S0747563217301693

[55] C. Chang, “Self-control-centered empowerment model: Health consciousness and health knowledge as drivers of empowerment-seeking through health communication,” Health Communication, vol. 0, no. 0, pp. 1–12, 2019, pMID: 31480856. [En ligne]. Disponible : https://doi.org/10.1080/10410236.2019.1652385

[56] M. Gachter,¨ D. A. Savage, et B. Torgler, « The relationship between stress, strain and social capital, » Policing : An International Journal of Police Strategies & Management, 2011.

[57] D. Devakumar, G. Shannon, S. S. Bhopal, et I. Abubakar, « Racism and discrimination in covid-19 responses, » Lancet (London, England), vol. 395, no. 10231, p. 1194, 2020.

[58] L. O. Gostin, E. A. Friedman, et S. A. Wetter, “Responding to covid-19: How to navigate a public health emergency legally and ethically,” Hastings Center Report, 2020.

[59] C. Wenham, J. Smith, et R. Morgan, “Covid-19: the gendered impacts of the outbreak,” The Lancet, vol. 395, no. 10227, pp. 846–848, 2020.

[60] “NOVID,” April 2020. [En ligne]. Disponible : https://novid.org/

[61] T. White, R. Fenwick, I. Becker-Mayer, J. Petrie, Z. Szabo, D. Blank, J. Colligan, M. Hittle, M. Ingle, O. Nash, V. Nguyen, J. Schwaber, A. Veeraghanta, M. Voloshin, S. V. Arx, et H. Xue, « Slowing the spread of infectious diseases using crowdsourced data, » March 2020. [En ligne]. Disponible : https://www.covid-watch.org/article

[62] L. Kelion, “Coronavirus : German contact-tracing app takes different path to NHS,” BBC News, May 2020. [En ligne]. Disponible : https://www.bbc.com/news/technology-52650576

[63] J. Tidy, “Coronavirus : Israel enables emergency spy powers,” BBC News, March 2020. [En ligne]. Disponible : https://www.bbc.com/news/technology-51930681

[64] V. Liu, M. A. Musen, et T. Chou, « Data breaches of protected health information in the United States, » Jama, vol. 313, no. 14, pp. 1471–1473, 2015.

[65] S. Warren et L. Brandeis, “The right to privacy,” Harvard Law Review, vol. 4, pp. 193–220, 1890.

[66] W. Prosser, “Privacy,” California Law Review, vol. 48, pp. 383–423, 1960.

[67] J. Van Den Hooff, D. Lazar, M. Zaharia, et N. Zeldovich, “Vuvuzela: Scalable private messaging resistant to traffic analysis,” in Proceedings of the 25th Symposium on Operating Systems Principles, 2015, pp. 137–152.

[68] N. Tyagi, Y. Gilad, D. Leung, M. Zaharia, et N. Zeldovich, “Stadium: A distributed metadata-private messaging system,” in Proceedings of the 26th Symposium on Operating Systems Principles, 2017, pp. 423–440.

[69] H. Corrigan-Gibbs, D. Boneh, et D. Mazieres,` ‘Riposte: An anonymous messaging system handling millions of users,’ in 2015 IEEE Symposium on Security and Privacy. IEEE, 2015, pp. 321–338.

[70] R. C. Merkle, ‘Secure communications over insecure channels,’ Communications of the ACM, vol. 21, no. 4, pp. 294–299, 1978.

[71] B. Greschbach, G. Kreitz, et S. Buchegger, ‘The devil is in the metadatanew privacy challenges in decentralised online social networks,’ in 2012 IEEE International Conference on Pervasive Computing and Communications Workshops. IEEE, 2012, pp. 333–339.

[72] A. Berke, M. Bakker, P. Vepakomma, R. Raskar, K. Larson, et A. Pentland, “Assessing disease exposure risk with location histories and protecting privacy: A cryptographic approach in response to a global pandemic,” arXiv preprint arXiv:2003.14412, 2020.

[73] D. L. Chaum, ‘Untraceable electronic mail, return addresses, and digital pseudonyms,’ Communications of the ACM, vol. 24, no. 2, pp. 84–90, 1981.

[74] M. G. Reed, P. F. Syverson, et D. M. Goldschlag, ‘Anonymous connections and onion routing,’ IEEE Journal on Selected areas in Communications, vol. 16, no. 4, pp. 482–494, 1998.

[75] K. El Emam et B. Malin, ‘Concepts and methods for de-identifying clinical trial data,’ Paper commissioned by the Committee on Strategies for Responsible Sharing of Clinical Trial Data, 2014.

[76] S. Canada, ‘Census division (CD),’ 2016.

[77] S. C. G. Division, Introducing the dissemination area for the 2001 census: An update. Geography Division, Statistics Canada, 2001.

[78] L. Sweeney, ‘k-anonymity: A model for protecting privacy,’ International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, vol. 10, no. 05, pp. 557–570, 2002.

[79] M. Fredrikson, S. Jha, et T. Ristenpart, ‘Model inversion attacks that exploit confidence information and basic countermeasures,’ in Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, 2015, pp. 1322–1333.

[80] R. K. Garrett, ‘The echo chamber distraction: Disinformation campaigns are the problem, not audience fragmentation.’ 2017.

[81] M. Y. Li et J. S. Muldowney, “Global stability for the seir model in epidemiology,” Mathematical biosciences, vol. 125, no. 2, pp. 155–164, 1995.

[82] S. L. Chang, N. Harding, C. Zachreson, O. M. Cliff, et M. Prokopenko, “Modelling transmission and control of the covid-19 pandemic in australia,” arXiv preprint arXiv : 2003.10218v2, 2020.

[83] L. Ferretti, C. Wymant, M. Kendell, L. Zhao, A. Nurtay, L. Abeler-Drner, M. Parker, D. Bonsall, and C. Fraser, ‘Quantifying sars-cov-2 transmission suggests epidemic control with digital contact tracing,’ Science, 2020.

[84] T. Jefferson, C. B. Del Mar, L. Dooley, E. Ferroni, L. A. Al-Ansary, G. A. Bawazeer, M. L. van Driel, N. S. Nair, M. A. Jones, S. Thoring, et J. M. Conly, ‘Physical interventions to interrupt or reduce the spread of respiratory viruses (review),’ Cochrane Database of Systematic Reviews, 2020.

[85] M. J. Kim et S.Denyer, “A travel log of the times in South Korea: Mapping the movements of coronavirus carriers,” The Washington Post, March 2020. [En ligne]. Disponible : https://www.washingtonpost.com/world/asia_pacific/coronavirus-south-korea-tracking-apps/2020/03/13/2bed568e-5fac-11ea-ac50-18701e14e06d story.html

[86] D. P. Kingma et M. Welling, ‘Auto-encoding variational bayes,’ in Proceedings of the International Conference on Learning Representations (ICLR), 2014.

[87] G. E. Hinton, P. Dayan, B. J. Frey, et R. M. Neal, ‘The’ wake-sleep” algorithm for unsupervised neural networks,” Science, vol. 268, no. 5214, pp. 1158–1161, 1995.

[88] Y. Bengio, N. Leonard,´ et A. Courville, « Estimating or propagating gradients through stochastic neurons for conditional computation, » arXiv preprint arXiv:1308.3432, 2013.

[89] R. J. Williams, “Simple statistical gradient-following algorithms for connectionist reinforcement learning,” Machine learning, vol. 8, no. 3-4, pp. 229–256, 1992.

[90] E. Jang, S. Gu, et B. Poole, « Categorical reparameterization with gumbel-softmax, » arXiv preprint arXiv:1611.01144, 2016.

[91] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, et I. Polosukhin, « Attention is all you need, » NeurIPS, 2017.

[92] N. R. Ke, A. G. A. P. GOYAL, O. Bilaniuk, J. Binas, M. C. Mozer, C. Pal, et Y. Bengio, “Sparse atten-tive backtracking: Temporal credit assignment through reminding,” in Advances in neural information processing systems, 2018, pp. 7640–7651.

[93] R. A. Heiner, “The origin of predictable behavior,” The American economic review, vol. 73, no. 4, pp. 560–595, 1983.

[94] R. Dreibelbis, A. Kroeger, K. Hossain, M. Venkatesh, et P. Ram, “Behavior change without behavior change communication: Nudging handwashing among primary school students in bangladesh,” International Journal of Environmental Research and Public Health, vol. 13, no. 1, p. 129, Jan 2016. [En ligne]. Disponible : http://dx.doi.org/10.3390/ijerph13010129

[95] T. C. Leonard, “Richard h. thaler, cass r. sunstein, nudge: Improving decisions about health, wealth, and happiness,” 2008.

[96] D. Hilty et S. Chan, “Human behavior with mobile health: Smartphone/ devices, apps and cognition,” Psychology and Cognitive Sciences - Open Journal, vol. 4, pp. 36–47, 12 2018.

[97] S. Mitchie, S. Ashford, F. Sniehotta, S. Dombrowski, A. Bishop, et D. French, « A refined taxonomy of behaviour change techniques to help people change their physical activity and healthy eating behaviors: the calo-re taxonomy, » Psychol Health, vol. 26, pp. 1479–1498, 2011.

[98] N. Eyal et R. Hoover, « Hooked: A guide to building habit-forming products, » 2013.

[99] J. M. Jachimowicz, S. Chafik, S. Munrat, J. C. Prabhu, et E. U. Weber, “Community trust reduces myopic decisions of low-income individuals,” Proceedings of the National Academy of Sciences, vol. 114, no. 21, pp. 5401–5406, 2017.

[100] K. Kongats, J. A. McGetrick, K. D. Raine, C. Voyer, et C. I. Nykiforuk, « Assessing general public and policy influencer support for healthy public policies to promote healthy eating at the population level in two canadian provinces, » Public Health Nutrition, vol. 22, no. 8, p. 14921502, 2019.

[101] L. Festinger, A theory of cognitive dissonance. Stanford university press, 1962, vol. 2.

[102] J. Krause, D. P. Croft, et R. James, “Social network theory in the behavioural sciences: potential applications,” Behavioral Ecology and Sociobiology, vol. 62, no. 1, pp. 15–27, 2007.

[103] P. Zaharias et A. Poylymenakou, “Developing a usability evaluation method for e-learning applications: Beyond functional usability,” Intl. Journal of Human–Computer Interaction, vol. 25, no. 1, pp. 75–98, 2009.

[104] J. Nielsen, “Usability metrics: Tracking interface improvements,” Ieee Software, vol. 13, no. 6, pp. 1–2, 1996.

[105] N. Ryan, “Willpower: Rediscovering the greatest human strength, by roy f. baumeister and john tierney,” 2012.

[106] A. Bandura, “Self-efficacy: Toward a unifying theory of behavioral change,” Advances in Behaviour Research and Therapy, vol. 1, no. 4, pp. 139 – 161, 1978, perceived Self-Efficacy: Analyses of Bandura’s Theory of Behavioural Change. [En ligne]. Disponible : http://www.sciencedirect.com/science/article/pii/0146640278900024

[107] A. K. Koch et J. Nafziger, “Self-regulation through goal setting*,” The Scandinavian Journal of Economics, vol. 113, no. 1, pp. 212–227, 2011. [En ligne]. Disponible : https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9442.2010.01641.x

[108] K. Degirmenci, “Mobile users information privacy concerns and the role of app permission requests,” International Journal of Information Management, vol. 50, pp. 261 – 272, 2020. [En ligne]. Disponible : http://www.sciencedirect.com/science/article/pii/S0268401218307965

[109] A. E. Waldman, “Cognitive biases, dark patterns, and the privacy paradox,” Current Opinion in Psychology, vol. 31, pp. 105 – 109, 2020, privacy and Disclosure, En ligne and in Social Interactions. [En ligne]. Disponible : http://www.sciencedirect.com/science/article/pii/S2352250X19301484

[110] M. D. Rice et E. Bogdanov, “Privacy in doubt: An empirical investigation of canadians’ knowledge of corporate data collection and usage practices,” Canadian Journal of Administrative Sciences/Revue Canadienne des Sciences de l’Administration, vol. 36, no. 2, pp. 163–176, 2019. [En ligne]. Disponible : https://onlinelibrary.wiley.com/doi/abs/10.1002/cjas.1494

[111] O. of the Privacy Commissioner of Canada, “2018-19 survey of Canadians on privacy,” 2019.

[112] M. S. Young, K. A. Brookhuis, C. D. Wickens, et P. A. Hancock, “State of science: mental workload in ergonomics,” Ergonomics, vol. 58, no. 1, pp. 1–17, 2015, pMID: 25442818. [En ligne]. Disponible : https://doi.org/10.1080/00140139.2014.956151

[113] A. Springer et S. Whittaker, “Progressive disclosure: Empirically motivated approaches to designing effective transparency,” in Proceedings of the 24th International Conference on Intelligent User Interfaces, ser. IUI 19. New York, NY, USA: Association for Computing Machinery, 2019, p. 107120. [En ligne]. Disponible : https://doi.org/10.1145/3301275.3302322

[114] K. M. Gallagher, J. A. Updegraff, A. J. Rothman, et L. Sims, « Perceived susceptibility to breast cancer moderates the effect of gain-and loss-framed messages on use of screening mammography. » Health Psychology, vol. 30, no. 2, p. 145, 2011.

[115] J. G. Baseman, D. Revere, I. Painter, M. Toyoji, H. Thiede, et J. Duchin, “Public health communications and alert fatigue,” BMC health services research, vol. 13, no. 1, p. 295, 2013.

[116] C. Rossmann, Content Effects: Health Campaign Communication. American Cancer Society, 2017, pp. 1–11. [En ligne]. Disponible : https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118783764.wbieme0127

[117] C. for Disease Control and Prevention, “CERC: Psychology of a crisis,” 2019. [En ligne]. Disponible : https://emergency.cdc.gov/cerc/ppt/CERC_Psychology_of_a_Crisis.pdf

[118] S. C. Segerstrom et G. E. Miller, « Psychological stress and the human immune system: a meta-analytic study of 30 years of inquiry. » Psychological bulletin, vol. 130, no. 4, p. 601, 2004.

[119] J. N. Morey, I. A. Boggero, A. B. Scott, et S. C. Segerstrom, “Current directions in stress and human immune function,” Current opinion in psychology, vol. 5, pp. 13–17, 2015.

[120] R. Glaser, J. Sheridan, W. B. Malarkey, R. C. MacCallum, et J. K. Kiecolt-Glaser, “Chronic stress modulates the immune response to a pneumococcal pneumonia vaccine,” Psychosomatic medicine, vol. 62, no. 6, pp. 804–807, 2000.

[121] H. Yao, J.-H. Chen, et Y.-F. Xu, “Patients with mental health disorders in the covid-19 epidemic,” The Lancet Psychiatry, vol. 7, no. 4, p. e21, 2020.

[122] D. R. Garfin, R. C. Silver, et E. A. Holman, “The novel coronavirus (covid-2019) outbreak: Amplification of public health consequences by media exposure.” Health Psychology, 2020.

[123] D. P. Bruns, N. V. Kraguljac, et T. R. Bruns, “Covid-19: Facts, cultural considerations, and risk of stigmatization,” Journal of Transcultural Nursing, vol. 0, no. 0, p. 1043659620917724, 0, pMID: 32316872. [En ligne]. Disponible : https://doi.org/10.1177/1043659620917724

[124] J. C. Stone, K. Glass, J. Clark, Z. Munn, P. Tugwell, et S. A. Doi, “A unified framework for bias assessment in clinical research,” International Journal of Evidence-Based Healthcare, vol. 17, no. 2, pp. 106–120, 2019.

[125] G. George, A. M. McGahan, et J. Prabhu, “Innovation for inclusive growth: Towards a theoretical framework and a research agenda,” Journal of Management Studies, vol. 49, no. 4, pp. 661–683, 2012. [En ligne]. Disponible : https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-6486.2012.01048.x

[126] W. R. Frey, D. U. Patton, M. B. Gaskell, et K. A. McGregor, “Artificial intelligence and inclusion: Formerly gang-involved youth as domain experts for analyzing unstructured twitter data,” Social Science Computer Review, vol. 38, no. 1, pp. 42–56, 2020. [En ligne]. Disponible : https://doi.org/10.1177/0894439318788314

*Ce livre blanc est une traduction du document : ArXiv:2005.08502. Vous êtes invité à soumettre vos corrections et commentaires à : medias@mila.quebec.