UNIVERSITÉ DU QUÉBEC À MONTRÉAL

BIG DATA, INTELLIGENCE ARTIFICIELLE ET PRÉVENTION DU SUICIDE

THÈSE PRÉSENTÉE COMME EXIGENCE PARTIELLE
DU DOCTORAT EN PSYCHOLOGIE

PAR

CARL MARIA MORCH

JUIN 2019

DÉDICACE

Ce travail est dédié aux personnes qui consacrent leur vie à rendre les technologies plus justes et plus équitables dans l'intérêt du plus grand nombre. J'espère, par ce travail, avoir contribué à diffuser leurs efforts.

REMERCIEMENTS

Merci à ma femme Rebecca Samulon, pour tout et bien plus. Merci d'avoir été une aide morale, matérielle, humaine. Merci pour ton optimisme à toute épreuve. Merci d'avoir fait une pause entre les peupliers et les palmiers. Avec tout mon amour immense. Mes pensées aimantes vont aussi à ma fille Faye, née pendant mes années de doctorat.

Merci à Brian pour son suivi, son soutien matériel et les opportunités. Merci de m'avoir toujours encouragé à tenter des choses.

Merci également à mes parents, mes sœurs vaillantes et ma famille pour m'avoir soutenu. Pensées très spéciales à mes amis de Montréal qui m'auront accompagné et soutenu pendant ce travail de longue haleine (pensées super spéciales à la rue Berri et Kizu). Ce travail est également dédié à mes amis de France, les BFF en particulier.

Merci à François Lauzier-Klopp, Louis-Philippe Bourdieu, Luc Dargis-Couillard, Fanny Mi-Suisse, DL et tous mes collègues. Et enfin, merci à Elsie Titouan et au vénérable Professeur Tommy Saïas pour son mentorat. Enfin, merci à André Mondoux pour sa passion contagieuse et à Dominic Martin pour m'aider à tracer mon dans le monde de l'éthique.

LISTE DES ABRÉVIATIONS

AI : Artificial Intelligence

BD : Big Data

TIC : Technologies de l'Information et de la Communication

MHSP : Mental Health and Suicide Prevention

RÉSUMÉ

Les objectifs de cette thèse sont de présenter :

- les différentes utilisations actuelles du Big Data et de l'Intelligence Artificielle en prévention du suicide
- les enjeux éthiques associés à l'utilisation et au déploiement de ces technologies en prévention du suicide
- un outil éthique, validé, pour aider ceux qui voudraient utiliser ces technologies à prendre des décisions en connaissance des risques potentiels. Cette thèse est composée de trois articles.

Le premier article est une étude de portée par recension systématique des écrits. Elle a consisté à explorer les bases de données informatiques, de sciences humaines, de psychologie et biomédicales. Son but était un aperçu étendu des utilisations actuelles en prévention du suicide. Elle a suivi le protocole PRISMA afin de répondre aux standards les plus rigoureux en matière de recension des écrits. Après la collecte de données, 4891 articles ont été filtrés. Ce sont finalement 58 études qui ont été codifiées par une équipe de recherche, soumises à un interjuge, puis analysées. Les résultats de cette étude de portée montrent que Big Data et Intelligence Artificielle sont aujourd'hui autant utilisés. La plupart des études provenaient de pays anglophones. Les études étaient de nature hétérogène. Les techniques prédominantes en Intelligence Artificielle étaient l'apprentissage machine (Machine Learning) et le Natural Language Processing (NLP). Les techniques prédominantes en Big Data étaient l'analyse de média sociaux (type Twitter) ou de recherches Internet (type Google). La plupart des études recensées portaient sur de l'analyse linguistique et de l'infoveillance. Les sources principales de données étaient les dossiers médicaux électroniques. Plus de 75% des articles ne comportaient pas de mention d'enjeux éthiques.

Le deuxième article a consisté à analyser 10 rapports internationaux en matière d'éthique et d'intelligence artificielle. Une équipe a extrait 329 recommandations. A partir de celles qui pouvaient s'appliquer au domaine de la santé, une liste de contrôle (checklist) a été consolidée (en 42 items). Cette liste a été soumise à un processus de consultation Delphi en deux rondes, auprès d'un panel international de 16 experts et professionnels. La consultation a permis d'aboutir à une liste finale de 38 items, répartis en quatre catégories : "Description du Système Intelligent Autonome", "Vie privée et transparence", "Sécurité", "Risques en Santé", "Biais".

Le troisième article a synthétisé les discussions éthiques trouvées dans les articles de l'étude de portée (article 1). Des recommandations ont été formulées à partir de la littérature en éthique, Big Data, Intelligence Artificielle et santé mentale. Cette étude tend à montrer que nous sommes dans une phase d'adoption enthousiaste de ces technologies en prévention du suicide. Elles sont jugées très prometteuses mais aussi

porteuses de nombreux défis éthiques, sociaux et scientifiques. Les recommandations appellent à davantage de transparence dans les publications sur les défis éthiques rencontrés, davantage de clarté dans le processus de traitement et d'analyse des données.

Mots clés : Big Data, Artificial Intelligence, Suicide Prevention, Machine Learning, Ethics

TABLE DES MATIÈRES

LISTE DES TABLEAUX

Article 1 **Page**

Article 2 **Page**

LISTE DES FIGURES

Article 1 **Page**

Article 2 **Page**

INTRODUCTION DE LA THÈSE

Actualité du domaine de recherche

Les données informatiques sont au cœur de l'actualité internationale. De l'affaire Snowden, à la possible amende de 5 milliards de dollars contre Facebook, en passant par la nouvelle réglementation européenne (RGPD), la privée fait désormais partie du débat public. Parallèlement, l'intelligence artificielle (IA) connaît une popularité croissante. De nombreux rapports publics (en France, au Canada, aux États-Unis, en Europe) et de consortiums (comme IEEE) ont été publié afin de cerner les promesses et risques associés.

Dans ce contexte, la province du Québec souhaite devenir la capitale mondiale de l'Intelligence Artificielle (IA). Elle dispose de structures pour attirer les sociétés étrangères en technologies numériques (ex. Montréal International), elle souhaite créer un organisme international de régulation des enjeux liés à l'IA. Avec le gouvernement fédéral, elle subventionne largement de façon directe et indirecte la recherche et l'investissement dans ces domaines (ex. à travers le CIRFA ou les Fonds de Recherche du Québec). Notre sujet s'inscrit ainsi dans une actualité contrastée : d'un côté les technologies font l'objet d'investissement massifs, et de l'autre elles suscitent de nombreuses inquiétudes. Nous avons tâché de faire le lien entre ce paradoxe et notre domaine : la Psychologie.

Démarche

Le présent projet de recherche est né du travail effectué pendant une année auprès de l'Autorité de Régulation des Jeux en Ligne en France. À cette occasion, j'étais chargé d'assister à la prévention du jeu pathologique et l'établissement de critères de détection du jeu pathologique dans l'Internet. Cette dernière mission se fondait sur l'utilisation

des données générées par les joueurs et collectées à large échelle, en temps réel, auprès de tous les opérateurs de jeux agréés en France.

Nous devions également accompagner les entreprises dans l'application de leurs obligations réglementaires en matière de « jeu responsable ». C'est à cette occasion que j'ai été confronté aux enjeux complexe de deux bouleversements technologiques majeurs : le « Big Data » (BD) et de l'Intelligence Artificielle appliqué à la psychologie et au secteur de la santé. J'ai tout particulièrement été intéressé par les situations où des joueurs d'argent en ligne étaient identifiés comme potentiellement suicidaires ou en détresse psychologique. Mon projet de thèse a été motivée par le souhait de travailler en profondeur sur les enjeux éthiques et sociaux associés à l'utilisation du BD ou de l'IA pour les repérer ou leur proposer de l'aide.

Objectifs du projet

Une des questions qui ont dirigé ce travail de thèse fut : comment se saisir du potentiel des technologies numériques tout en prenant en compte les défis éthiques et sociaux qu'ils soulèvent ?

En effet, le champ de la santé mentale a été transformé durablement par l'adoption des Technologies de l'information et de la communication (TIC) comme le BD et l'IA : de la prévention en passant par la thérapie (Conway & O'Connor, 2016; Mohr et al., 2013). Chercheurs, décisionnaires et praticiens doivent désormais intégrer ces nouveaux outils mais aussi intégrer les risques qu'ils soulèvent pour les usagers (Luxton, 2014a, 2014b).

Un champ permettra d'illustrer de façon précise l'importance de ces enjeux : la prévention du suicide. Ce domaine est un des principaux enjeux de santé publique en matière de santé mentale (Canada, 2014; WHO, 2014). Les TIC y sont très présentes, que ce soit dans la digitalisation des lignes d'écoute (Mishara & Côté, 2013), ou dans l'offre d'aide *via* des applications sur téléphones intelligents (Labelle, Bibaud-De

Serres, & Leblanc, 2013). Mais aujourd'hui, le BD et l'IA soulève des questionnements très larges (Poulin, 2014) qui dépassent les cadres habituels de la prévention du suicide (Mishara & Weisstub, 2013b).

Les objectifs de notre thèse seront ainsi de : 1. Étudier l'utilisation du BD et de l'IA en santé mentale, à travers le domaine de la prévention du suicide, 2. Étudier les enjeux éthiques associés à ce domaine, 3. Proposer un outil pour une utilisation éthique du BD et de l'IA, adapté à la prévention du suicide et à son cadre de référence la santé mentale.

Structure de la thèse

Un premier chapitre présentera les principaux concepts en lien avec le suicide. Le but de cette partie sera de présenter l'importance générale de la prévention du suicide, puis ses caractéristiques lorsqu'elle est faite en ligne.

Nous présenterons dans un chapitre 2 les concepts de BD et d'IA. Nous les placerons dans le contexte de la prévention du suicide. Ce chapitre est une étude de portée sur les utilisations actuelles. Elle a été faite au moyen d'une recension systématique des écrits.

Cette étude de portée montrera que la plupart des études actuelles intègrent peu de considérations éthiques lorsqu'elles utilisent l'IA ou le BD en prévention du suicide. Pour répondre à ce manque, un outil a été développé pour accompagner au niveau éthique, tout chercheur, décisionnaire ou professionnel intéressé par l'utilisation de l'IA pour prévenir le suicide ou l'utiliser en santé mentale : le Protocole Canadien. Cet outil est une liste de contrôle (*checklist*) qui synthétise les principaux défis éthiques relevés dans la littérature internationale en matière d'éthique et d'IA. Ce Protocole Canadien est présenté dans un article (chapitre 3) qui présentera son élaboration et son processus de validation. Ce dernier a été fait au moyen d'une stratégie éprouvée en santé pour consulter des experts : la méthode Delphi.

Suite à ces deux chapitres, des recommandations éthiques ont été formulées dans un troisième article (chapitre 4). La direction de thèse souhaitait que soit opérée une

synthèse des expériences et connaissances développées et leurs implications. L'objectif sera de promouvoir un usage éthique et socialement responsable de ces technologies quand on veut les utiliser pour aider des personnes en détresse. Ce chapitre, avec la conclusion de la thèse, visera à synthétiser les principaux enjeux éthiques actuels en matière de prévention du suicide à l'aide de l'Intelligence Artificielle.

Références bibliographiques

Canada. (2014). Mental Health and Wellness. Retrieved from http://cbpp-pcpe.phac-aspc.gc.ca/public-health-topics/mental-health-and-wellness/

Conway, & O'Connor. (2016). Social media, big data, and mental health: Current advances and ethical implications. *Current Opinion in Psychology, 9*, 77-82. doi:10.1016/j.copsyc.2016.01.004

Labelle, Bibaud-De Serres, & Leblanc. (2013). Innovating to Treat Depression and Prevent Suicide: The IPhone@ PSY ASSISTANCE Application. In *Suicide Prevention and New Technologies* (pp. 166-180): Springer.

Luxton. (2014a). Artificial intelligence in psychological practice: Current and future applications and implications. *Professional Psychology: Research and Practice, 45*(5), 332-332.

Luxton. (2014b). Recommendations for the ethical use and design of artificial intelligent care providers. *Artificial Intelligence in Medicine, 62*(1), 1-10.

Mishara, & Côté. (2013). Prevention, Suicide new technologies: Towards evidence based practice. In (pp. 1-23). New York: Palgrave Macmillan.

Mishara, & Weisstub. (2013b). Challenges in the control and regulation of suicide promotion and assistance over the Internet. *Suicide Prevention and New Technologies. Palgrave Macmillan, Hampshire, UK*, 63-75.

Mohr, Ph, Burns, Ph, Schueller, Ph, . . . S. (2013). Behavioral Intervention Technologies : Evidence review and recommendations for future research in mental health. *General Hospital Psychiatry*. doi:10.1016/j.genhosppsych.2013.03.008

Poulin. (2014). Big Data Custodianship in a Global Society. *SAIS Review of International Affairs, 34*(1), 109-116. doi:10.1353/sais.2014.0002

WHO. (2014). *Preventing suicide: A global imperative* (9241564776). Retrieved from

1. CHAPITRE 1
PRÉSENTATION GENERALE DES CONCEPTS : LE SUICIDE ET SA PRÉVENTION DANS L'INTERNET

1.1. Définitions des concepts

1.1.1. La santé mentale

La santé est définie par l'Organisation mondiale de la santé (OMS) comme un état de complet bien-être physique, mental et social, qui ne peut être réduit à l'absence de symptômes (WHO, 1948). La santé mentale est vue comme une des composantes essentielles au bien-être des individus. L'Agence de Santé Publique du Canada, la définit plus précisément comme un :

« sentiment positif de bien-être émotionnel et spirituel qui respecte l'importance de la culture, de l'équité, de la justice sociale, des interactions et de la dignité personnelle » (Canada, 2014).

1.1.2. Le suicide

Le suicide est un des défis majeurs en matière de santé mentale (WHO, 2014). Les atteintes auto-infligées ou comportements autodestructeurs (« self-harm » et « self injurious-behaviors ») sont des phénomènes hautement complexes qui mettent à l'épreuve les personnes qui y sont confrontés (Nock, 2014). Parmi les atteintes à soi-même, le suicide en est la forme la plus dramatique. Ce dernier est un sujet historiquement et universellement bien documenté (Mishara & Tousignant, 2004). Cependant, le suicide est compris et accepté de façon très variable selon les cultures (DeLeo, Bille-Brahe, Kerkhof, & Schmidtke, 2004). À l'heure actuelle, il n'existe pas encore de définition universellement partagée. Le nombre de définitions fréquemment

utilisées oscille entre six (DeLeo et al., 2004), huit (Maris, Berman, & Silverman, 2000), douze (Windfuhr & Kapur, 2011), voire quinze (Silverman, 2006). Un des défis de la suicidologie sera de surmonter l'absence de nomenclature internationale et standardisée (Silverman, 2011).

Néanmoins, une définition simple et essentielle peut être de définir le suicide, dit "complété", comme "une mort intentionnelle causée par soi-même" (Mishara & Tousignant, 2004, p. 32). Ou encore, le suicide est "un acte intentionnel auto-infligé qui résulte en une mort" (Maris, 2002). Du point de vue médico-légal, en Amérique du Nord, les coroners doivent vérifier (O'Carroll et al., 1996) :

- Que la mort ait lieu suite à des blessures (*injuries*), suffocation, empoisonnement ;
- Qu'elle ait été auto-infligée ;
- Qu'elle soit intentionnelle.

Cette approche est limitée, car elle ne définit pas ce que signifie s'auto-infliger la mort et limite les causes de mort à celles pré-mentionnées (Silverman, 2006).

1.1.3. Tentatives suicidaires

Le CDC[1] définit la tentative de suicide comme un comportement non-fatal, dirigé envers soi-même, qui peut potentiellement blesser, guidé par l'intention de mourir comme résultat du comportement[2] (Crosby, Gfroerer, Han, Ortega, & Parks, 2011). Les tentatives de suicide sont intégrées dans les conduites suicidaires et peuvent se

---

[1] Center for Disease Control and Prevention

[2] « A non-fatal self-directed potentially injurious behavior with any intent to die as a result of the behavior »

décliner en plusieurs types : avorté, interrompu, non complété (Posner, Melvin, Stanley, Oquendo, & Gould, 2007).

Mais l'inclusion des tentatives de suicide dans la catégorie « parasuicidaire » (dont font partie les idéations suicidaires) a été sujet à débat scientifiques (Posner, Brodsky, KseniyaYershova, & Mann, 2014). L'intention de se tuer n'étant pas toujours claire, il a été décidé de davantage prendre en compte une distinction des comportements selon qu'ils étaient « fatals » ou « non-fatals » (Posner et al., 2014).

L'OMS estime que le nombre de tentatives de suicide est beaucoup plus important que le nombre de suicides (WHO, 2014). Mais le travail d'estimation a deux limites : les tentatives de suicide sont souvent sous-évaluées, et il n'existe que deux sources d'information (l'auto-déclaration dans les enquêtes et les dossiers médicaux) (WHO, 2014) . Au Québec, l'Institut National de la Santé Publique du Québec a estimé en 2008 qu'environ 0,5% de la population avait rapporté avoir tenté de se suicider dans les 12 derniers mois (Inspq, 2007). En se fondant sur les données issues des enquêtes populationnelles, il a été également observé que plus de femmes rapportaient avoir tenté de se suicider que des hommes (Inspq, 2007).

1.1.4. Impact du suicide

Le suicide a un taux mondial de 11,4 personnes pour 100 000 habitants (WHO, 2014). Tuant plus de 800 000 personnes par an, il demeure cependant une façon "résiduelle" de mourir, bien inférieure en nombre aux homicides ou aux accidents (DeLeo et al., 2004). Néanmoins, il est fréquemment sous-évalué (Bertolote & Wasserman, 2009).

Le report du nombre de morts par suicide a donné lieu à de nombreux débats et à une grande variabilité selon les pays (Oquendo et al., 2011). Les méthodes de calcul ne sont pas les mêmes selon la définition de suicide employée, et selon les critères retenus pour les enquêtes. Il a été évoqué le terme de "tour de Babel" en suicidologie (O'Carroll et al., 1996).

Le Centres de Contrôle et de lutte contre les maladies aux États-Unis estime que cette variabilité nuit à la qualité des données collectées sur le suicide, aux comparaisons dans le monde (Crosby et al., 2011). Ces arguments vont dans le sens des partisans d'une classification internationale commune, pour mieux comprendre, comparer et prévenir le suicide (Posner et al., 2014).

Nous venons de passer en revue les concepts de suicidologie pertinents avec notre sujet. Comme nous l'avons vu, la prévention du suicide prend plusieurs formes. Elle fait l'objet de nombreux débats théoriques. Et parallèlement, elle est devenu un enjeu international de santé publique (WHO, 2014). Enfin, elle est bouleversée par l'utilisation des TIC, et en particulier par l'utilisation du Big Data (Reidenberg & Gordon, 2012) et de l'Intelligence Artificielle (Poulin, Thompson, & Bryan, 2016).

1.2. Ampleur du phénomène

1.2.1. Le Suicide

Le suicide est un phénomène universel et de grande ampleur. À l'heure actuelle, l'OMS estime que 804 000 personnes se donnent la mort chaque année, soit un taux standardisé selon la classe d'âge de 11,4 pour 100 000 habitants par année (WHO, 2014). Il représente 50% des morts violentes chez les hommes, 71% chez les femmes. Il affecte toutes les catégories d'âge, de genre et socio-économiques (Mishara, Tousignant, 2004). Il touche tout particulièrement les personnes au-delà de 70 ans et les jeunes de 15 à 29 ans (WHO, 2014). L'OMS a décrété une situation d'urgence face à l'ampleur du phénomène et a lancé un plan d'action international en 2014 pour engager les pays membres à réduire les taux de suicide de 10% d'ici 2020 (WHO, 2014).

Au Canada, le taux selon l'âge est de 9,8 pour 100 000 en 2012, soit inférieur de 11.1% par rapport au taux de 11,0 calculé en 2000 (WHO, 2014). Le taux de suicide est cependant encore particulièrement élevé chez les hommes (14,9) et assez faible chez les femmes (4,8) (WHO, 2014). La situation varie cependant selon les provinces.

Le Québec est une des trois provinces canadiennes aux taux de suicide les plus élevés : 13,3 pour 100 000 habitants (Légaré & Gagné, 2015). L'Institut National de Santé Publique du Québec (INSPQ) estime que 1 101 personnes seraient décédées par suicides en 2013 (Légaré & Gagné, 2016). Là encore, le suicide touche davantage les hommes (taux ajusté de 20,7 pour 100 000) que les femmes (6,1 pour 100 000). La classe d'âge la plus touchée par le suicide est celle de 35-49 ans pour les hommes et de 50-64 pour les femmes.

1.2.2. Le suicide et l'Internet

Internet est souvent perçu en prévention du suicide comme un outil à double tranchant (Lewis & Seko, 2016a; Tam, Tang, & Fernando, 2007), à la fois moyen privilégié pour accéder à des moyens d'exprimer sa détresse, de trouver de l'aide, mais aussi accès facile à des renseignements sur les moyens de se suicider (Grella & Gilmore, 2002).

Chercheurs et professionnels ont d'abord cherché à évaluer l'impact possible de l'utilisation d'Internet sur le risque suicidaire. En effet, même s'il semble que les taux de suicide n'aient pas augmenté depuis la création d'Internet (Galtung & Hoff, 2007), il est admis que l'utilisation d'Internet (en lien avec le suicide ou l'automutilation) est commune chez les jeunes adultes, en particulier chez ceux avec des comportements et/ou pensées suicidaires (Mars, Heron, Biddle, Donovan, & Holley, 2015). Afin d'avoir une meilleure connaissance des pratiques numériques des personnes suicidaires, Harris et collègues ont interrogé près de 290 personnes (diagnostiquées à risque) pour étudier leurs fréquentations d'Internet en lien avec le risque qu'ils présentaient (Harris, McLean, & Sheffield, 2009). Leurs résultats indiqueraient que ceux qui ont fait des recherches Internet sur le suicide avaient des « symptômes suicidaires » plus élevés que ceux qui n'avaient pas d'activité Internet en lien avec le suicide (Harris et al., 2009à. Ces données ont renforcé l'idée qu'Internet pouvait être

utilisé par des personnes en situation de vulnérabilité psychologique, sans les prémunir des contenus sensibles accessibles en ligne.

Ceci est d'autant plus important qu'Internet se caractérise aussi par sa capacité à diffuser des informations pouvant avoir des conséquences dangereuses. En premier lieu, il peut servir à diffuser du contenu pro-suicide (Niezen, 2013), et ce, depuis les débuts d'Internet à la fin des années 80, avec la fondation de sites tels que "Churchofeuthanasia.org" et ASH (*alt.suicide.holiday*) aujourd'hui encore active (Westerlund & Wasserman, 2009). Par ailleurs, Internet est fréquemment utilisé pour trouver des renseignements sur les méthodes pour se donner la mort (Grella & Gilmore, 2002), diffuser des sites de personnes suicidaires comme celui du blogueur Martin Manley (Oremus, 2013). Ancien journaliste sportif, il avait expliqué son suicide sur un site Internet créé pour l'occasion. A sa demande, le site avait été laissé en ligne après sa mort. Le site est aujourd'hui encore accessible[3]. Enfin, Internet peut servir à diffuser du contenu pouvant entraîner un suicide. Au Canada, les morts des jeunes adolescentes Rehtaeh Parsons en 2011 et d'Amanda Todd en 2012 ont alerté l'opinion publique sur les risques de l'intimidation dans l'Internet (Southey, 2013), appelée cyberintimidation ou plus rarement « cyberbullicide » (Hinduja & Patchin, 2008).

Face à ce paradoxe, une grande partie de la littérature disponible s'interroge sur les moyens d'adapter de façon adéquate à Internet les méthodes de prévention du suicide (Fiedorowicz & Chigurupati, 2009).

1.2.3. La prévention du suicide et Internet

Internet est devenu un outil essentiel à la prévention du suicide (Krysinska, Westerlund, & Niederkrotenthaler, 2017) : de la recherche d'information dans l'Internet (Biddle,

---

[3] http://martin-manley.eprci.com/

Derges, Mars, & Heron, 2016), jusqu'à l'accès aux services offerts en ligne (Mishara & Kerkhof, 2013). Internet est parfois même appelé un "outil anti-suicide" (Narang & Lippmann, 2009, p. 12). C'est devenu une source fréquente pour chercher de l'information en santé (McDaid *et al.*, 2010), et en matière de suicide (Mishara et Kerkhof, 2013a). Vecteur de communication puissant, Internet génère de nombreuses communautés de soutien et offre des outils à visée d'aide psychologique (Niezen, 2013), tels que des lignes d'écoute ou des systèmes de clavardage (Mishara & Côté, 2013).

De fait, les institutions et acteurs de la prévention du suicide se sont emparés de cet aspect, à l'instar des Samaritains en Grande-Bretagne qui ont associé des services en ligne (email, chat) à leur traditionnelle ligne d'écoute (Narang & Lippmann, 2009). Ou encore, ils ont développé des services adaptés aux technologies numériques, comme 113Online aux Pays-Bas : des outils de clavardage, des applications pour téléphones intelligents (Mokkenstorm & Huisman, 2011).

Un des premiers efforts de prévention de suicide en ligne a été été de tenter de limiter l'accès aux sites Internet pro-suicide, comme au Portugal en 1995 ou en France en 2002 (Westerlund et Wasserman, 2009). Certains pays, comme l'Australie, sont allés jusqu'à modifier leur code criminel, en spécifiant qu'héberger du contenu pro-suicide était répréhensible pénalement (Pirkis, Neal, Dare, Blood R, & Studdert, 2009).

Pour appliquer ces interdictions, la plupart des pays ont collaboré directement avec les fournisseurs d'accès Internet (FAI ou ISP, *Internet Service Providers*). Dans la plupart des cas, des outils de blocage et de filtrage des sites ont été mis en place (Mishara & Weisstub, 2013b). Cependant, à l'heure actuelle, les outils de filtrage sont peu efficaces en raison de défis logistiques importants. Par exemple, les logiciels ont du mal à distinguer finement les sites pro-suicide et les sites d'aide. La classification des sites a dû être faite manuellement. Par ailleurs, les outils de blocage, eux aussi, n'ont pas toujours été appliqués avec succès. En Australie comme en Allemagne, les blocages

étaient efficients pour les sites hébergés dans le pays d'origine, mais inefficaces pour les sites hébergés en dehors des frontières (Mishara, Kerkhof, 2013).

1.3. Éthique

Notre thèse aborde la question de l'éthique en lien avec le Big Data et l'intelligence artificielle. Nous tenons à en donner une brève définition. Ce concept est très discuté et les définitions peuvent varier. Nous retiendrons celles données par Crisp et Weinstock.

L'éthique est l'ensemble des coutumes et valeurs qui s'imposent à des groupes d'individus donnés (Crisp, 2011). Elle est la « discipline qui tente de repérer les principes régissant le vivre-ensemble » (Weinstock, 2011, p. 15). On peut y inclure la notion de moralité, qui implique les notions de bien, de mauvais, de culpabilité, de honte (Crisp, 2011). Ce champ d'études entretient de nombreux liens avec la philosophie (Crips, 2011). Dans cette perspective, on peut concevoir l'éthique comme l'étude de ce qu'on doit aux autres et envers soi-même (Weinstock, 2011). L'éthique peut donc comprendre des considérations très larges comme être circonscrit à des questions très pratiques.

1.4. Nature de la démarche

1.4.1. Contexte

Notre sujet porte sur les collectes et analyses massives de données informatiques faites à des fins de prévention du suicide. De façon plus spécifique, il vise à décrire les pratiques actuelles, ainsi que les enjeux éthiques et sociaux associés. Pour étudier ce sujet, nous avons envisagé plusieurs devis. Mais trois facteurs ont ultimement déterminé nos choix méthodologiques :

- nous n'avons trouvé presqu'aucune étude publiée dans une revue avec comité de lecture portant spécifiquement sur le Big Data en prévention du suicide ou traitant des enjeux associés ;
- nous avons conçu des devis quasi-expérimentaux. Cependant les données nécessaires à leur réalisation étaient inaccessibles – et ce en dépit de nos efforts pour y avoir accès. Elles appartenaient à des sociétés privées qui avaient développé des partenariats avec des universités et associations aux Etats-Unis ;
- nous avions conçu un projet de devis d'évaluation des stratégies de prévention du suicide pour des structures gérant des données informatiques personnelles. Il n'existait cependant pas à l'époque d'institution de ce type au Canada. Cette démarche ne remplissait ainsi pas des critères de faisabilité matérielle.

Nous avons ainsi opté pour un devis théorique, non-empirique. Son objectif est de répondre au premier facteur cité : la quasi-absence d'écrits sur le Big Data et la prévention du suicide.

1.4.2. Cadre théorique

Nous avons adopté un cadre théorique transdisciplinaire, intégrant des documents de natures variées. Nous souhaitions initialement travailler avec des articles scientifiques sur les mégadonnées en prévention du suicide. Or, notre domaine d'étude est récent et peu étudié. Ceci a pour conséquence de limiter drastiquement l'existence d'écrits sur le sujet.

Nous connaissons ainsi des contraintes identiques à celles rencontrées dans des recensions des écrits sur la prévention du suicide en lien avec les média socio numériques (Robinson *et al.*, 2015) et sur les utilisations des Google Trends en santé (Nuti *et al.*, 2014). Ces études n'étaient pas en mesure de collecter beaucoup d'écrits portant spécifiquement sur leur domaine d'étude. Les auteurs évoquent les risques de mettre en place des critères d'inclusion trop restrictifs au risque de ne rien avoir

(Robinson *et al.*, 2015). Enfin, ils recommandent de ne pas se limiter aux études scientifiques. En effet, les articles recensés étant souvent peu détaillés et difficilement réplicables du point de vue méthodologique (Nuti *et al.*, 2014).

Travailler sur notre sujet requerra ainsi de s'intéresser à des documents de nature variées, dans des domaines connexes en lien avec la question du Big Data et de l'intelligence artificielle : de la sécurité informatique, à la psychologie, en passant par les études sur la surveillance (Schneier, 2015).

1.5. Références bibliographiques

Bertolote, & Wasserman. (2009). Development of definitions of suicidal behaviours: from suicidal thoughts to completed suicides. *Oxford textbook of suicidology and suicide prevention. A global perspective. Oxford University Press, Oxford*, 87-90.

Biddle, Derges, Mars, & Heron. (2016). Suicide and the Internet: Changes in the accessibility of suicide-related information between 2007 and 2014. *Journal of Affective*.

Canada. (2014). Mental Health and Wellness. Retrieved from http://cbpp-pcpe.phac-aspc.gc.ca/public-health-topics/mental-health-and-wellness/

Crisp. (2011). Ethics. https://www.rep.routledge.com/articles/overview/ethics/v-2/sections/ethics-and-metaethics doi:10.4324/9780415249126-L132-2

Crosby, Gfroerer, Han, Ortega, & Parks. (2011). *Suicidal thoughts and behaviors among adults aged>_18 Years--United States, 2008-2009*: US Department of Health and Human Services, Centers for Disease Control and Prevention.

DeLeo, Bille-Brahe, Kerkhof, & Schmidtke. (2004). *Suicidal behaviour: Theories and research findings*: Hogrefe Publishing.

Fiedorowicz, & Chigurupati. (2009). The Internet in suicide prevention and promotion. *Internet and Suicide. New York, NY: Nova Science Publishers*, 1-12.

Galtung, & Hoff. (2007). *Selvmordsfare og internett: kan jussen gi beskyttelse?* : Universitetet i Oslo, Det medisinske fakultet, Institutt for psykiatri, Seksjon for selvmordsforskning og-forebygging.

Grella, & Gilmore. (2002). Improving service delivery to the dually diagnosed in Los Angeles County. *Journal of Substance Abuse Treatment, 23*(2), 115-122. doi:10.1016/S0740-5472(02)00254-4

Harris, McLean, & Sheffield. (2009). Solving suicidal problems online: Who turns to the Internet for help? *Australian e-Journal for the Advancement of Mental Health, 8*(1), 28-36.

Hinduja, & Patchin. (2008). Cyberbullying: An Exploratory Analysis of Factors Related to Offending and Victimization. *Deviant Behavior, 29*(2), 129-156. doi:10.1080/01639620701457816

Inspq. (2007). La mortalité par suicide au Québec : tendances et données récentes – 1981 à 2007. *Institut National de Santé Publique du Québec*, 1-20.

Krysinska, Westerlund, & Niederkrotenthaler. (2017). A mapping study on the internet and suicide. *Crisis*.

Légaré, & Gagné. (2015). La mortalité par suicide au Québec : 1981 à 2012 – Mise à jour 2015.

Légaré, & Gagné. (2016). La mortalité par suicide au Québec : 1981 à 2013 – Mise à jour 2016.

Lewis, & Seko. (2016a). A Double-Edged Sword: A Review of Benefits and Risks of Online Nonsuicidal Self-Injury Activities. *Journal of Clinical Psychology, 72*(3), 249-262. doi:10.1002/jclp.22242

Maris. (2002). Suicide. *Lancet, 360*(9329), 319-326. doi:10.1016/S0140-6736(02)09556-9

Maris, Berman, & Silverman. (2000). The theoretical component in suicidology. *Comprehensive textbook of suicidology*, 26-61.

Mars, Heron, Biddle, Donovan, & Holley. (2015). Exposure to, and searching for, information about suicide and self-harm on the Internet: Prevalence and predictors in a population based cohort of young. *Journal of Affective*.

Mishara, & Côté. (2013). Prevention, Suicide new technologies: Towards evidence based practice. In (pp. 1-23). New York: Palgrave Macmillan.

Mishara, & Kerkhof. (2013). The Future of Suicide Prevention at Our Doorstep. In L & Kerkhof (Eds.), (pp. ix--xii). New York: Palgrave Macmillan.

Mishara, & Tousignant. (2004). *Comprendre le suicide*: PUM.

Mishara, & Weisstub. (2013b). Challenges in the control and regulation of suicide promotion and assistance over the Internet. *Suicide Prevention and New Technologies. Palgrave Macmillan, Hampshire, UK*, 63-75.

Mokkenstorm, & Huisman. (2011). Suicide prevention via the internet and the telephone: 113Online. *Tijdschrift voor*.

Narang, & Lippmann. (2009). The Internet: its role in the occurrence and prevention of suicide. *Internet and Suicide. New York, NY: Nova Science Publishers*, 13-20.

Niezen. (2013). Internet suicide: communities of affirmation and the lethality of communication. *Transcultural Psychiatry, 50*, 303-322.

Nock. (2014). *The Oxford handbook of suicide and self-injury*: Oxford University Press.

O'Carroll, Berman, Maris, Moscicki, Tanney, & Silverman. (1996). Beyond the Tower of Babel: a nomenclature for suicidology. *Suicide & life-threatening behavior, 26*(3), 237-252. doi:10.1111/j.1943-278X.1996.tb00609.x

Oquendo, Feldman, Silverman, Currier, Brown, Chen, . . . Pearson. (2011). Variability in the definition and reporting of adverse events in suicide prevention trials: an examination of the issues and a proposed solution. *Archives of suicide research: official journal of the International Academy for Suicide Research, 15*(1), 29-42. doi:10.1080/13811118.2011.541146

Oremus. (2013). Martin Manley's sister asks Yahoo to put his suicide website back up. *www.slate.com*.

Pirkis, Neal, Dare, Blood R, & Studdert. (2009). Legal Bans on Pro-Suicide Web Sites: An Early Retrospective from Australia. *Suicide and Life-Threatening Behavior, 39*(2), 190-193.

Posner, Brodsky, KseniyaYershova, & Mann. (2014). The Classification of Suicidal Behavior 2. *The Oxford handbook of suicide and self-injury*, 7-7.

Posner, Melvin, Stanley, Oquendo, & Gould. (2007). Factors in the assessment of suicidality in youth. *CNS spectrums, 12*(02), 156-162.

Poulin, Thompson, & Bryan. (2016). Public health surveillance: Predictive analytics and big data. In *Artificial Intelligence in Behavioral and Mental Health Care* (pp. 205-230).

Reidenberg, & Gordon. (2012). *Using Big Data to Understand and Prevent Suicide in Communications and Technology for Violence Prevention for Violence Prevention: Worksop Summary* (9780309253512). Retrieved from Washington:

Silverman. (2006). The language of suicidology. *Suicide & life-threatening behavior, 36*(5), 519-532. doi:10.1521/suli.2006.36.5.519

Silverman. (2011). Challenges to Classifying Suicidal Ideations, Communications, and Behaviours. In (pp. 7-25): John Wiley & Sons, Ltd.

Southey (2013). [Bill C-13 is about a lot more than cyberbullying].

Tam, Tang, & Fernando. (2007). The internet and suicide: A double-edged tool. In (Vol. 18, pp. 453-455).

Weinstock. (2011). *Profession éthicien*: Les Presses de l'Université de Montréal.

Westerlund, & Wasserman. (2009). The role of the Internet in suicide prevention. In (pp. 525-532). Oxford, UK: Oxford University Press.

WHO. (1948). Definition of health. *http://www.who.int/suggestions/faq/zh/index.html*.

WHO. (2014). *Preventing suicide: A global imperative* (9241564776). Retrieved from

Windfuhr, & Kapur. (2011). Suicide and mental illness: a clinical review of 15 years findings from the UK National Confidential Inquiry into Suicide. *British medical bulletin, 100*, 101-121. doi:10.1093/bmb/ldr042

2. CHAPITRE 2

PREMIER ARTICLE - BIG DATA, ARTIFICIAL INTELLIGENCE AND SUICIDE PREVENTION: A SCOPING STUDY

Carl-Maria Mörch, M.Psy.

Léa Plourde-Léveillé, B.A.

Laurent Corthésy-Blondin, B.A.

Louis-Philippe Côté, B.A.

Cécile Bardon, Ph.D.

Réal Labelle, Ph.D.

Luc Dargis, M.A.

Brian L. Mishara, Ph.D.

2.1. Abstract

2.1.1. Context

Information and Communication Technologies are used by persons involved in suicide prevention in various ways, from surveying at-risk individuals to delivering online treatment. Very recently, two new technological fields have emerged: Artificial Intelligence and Big Data. Artificial Intelligence concerns the design of machines able to perform tasks defined as requiring intelligence. Big Data designates the increasing amounts of data generated in our society. These two fields have resulted in the development of promising ways of better detecting at-risk individuals, analyzing increasingly larger datasets. However, these techniques raise complex challenges. Furthermore, the extent of their use and the nature of their applications remains largely unknown. To date, there is no review or meta-analysis on this topic.

2.1.2. Objectives

The main objective of this study is to review the uses of Artificial Intelligence and Big Data in Suicide Prevention.

2.1.3. Methods

We conducted a scoping review from January 2017 to June 2017. We used four databases specialized in Psychology, Medicine, Humanities and Computer Sciences (PsycInfo, PubMed, ProQuest, IEEExplore). This review follows the PRISMA (Moher et al., 2015) recommendations . We initially collected 4891 articles, of which, after screening and analysis, we retained 58 articles. The articles included ranged from 2010 to 2018. This study was part of a larger review on Information and Communication Technologies in Suicide Prevention.

2.1.4. Results

The majority of studies came from English-speaking countries and were published between 2014 and 2016. There was a great heterogeneity in the nature and format of the studies (from literature review, to book chapter, to case study). The most frequently studied suicide topics were: suicide behaviors, ideation, deaths by suicide and attempts. Big Data and Artificial Intelligence were equally used in our corpus. The main Artificial Intelligence sub-technologies: were Machine Learning and Natural Language Processing. The main Big Data techniques used were real-time data analysis of social media information (Twitter API) or Internet search engines (Google Trends). Overall, the main source of data were Electronic Health Records and Internet-related content (sites, blogs, Internet search queries, social media data). A large proportion of the studies aimed at analyzing linguistic content and conduct Infoveillance (or infodemiology). In general, more than three quarters of the articles did not include ethical discussions (the approval by an ethics board was not considered to be an ethical discussion). Finally, only a third of the studies provided enough data and information to conduct rigorous comparisons. This might be explained by the fact that many of these studies were unpublished technical reports or summaries in conference proceedings that were found on specialized databases.

2.1.5. Discussion

Given the importance of ethics in online suicide prevention and controversies concerning the use in society of Artificial Intelligence, it is surprising that most of the studies did not address or discuss any ethical considerations. At the present time, the scientific community of researchers does not appear to be concerned about the ethics of their AI activities. At least, ethical concerns are almost never mentioned in the scientific literature. This contrasts with highly publicized concerns about what

information is being gathered and analyzed about them, and the potential impact on their lives, depending upon who has access to that information and what they do with it. For instance, the practice of redlining is increasingly used. It consists of denying an individual some services (e.g. healthcare services) because he or she might fit into a profile (e.g. ethnic, socioeconomic) (Schneier, 2014). Another famous example is a study conducted by researchers from Facebook and Cornell University on the emotional manipulation study (Kramer et al., 2015). This study raised numerous concerns, because none of the Facebook users had been warned that an experiment was being conducted on them, and that their data was analyzed (Flick, 2016). Part of the reason that there is little consideration of ethical issues may be the fact that much of the research is being conducted by private companies (such as in social Media like Facebook and Google), and not and not in research institutions, where more strict ethical guidelines are enforced. People who use digital technologies offered by private companies generally accept blanket statements allowing the company to gather, analyze and use your data as they please, when they click to indicate that they "accept" the license, agreement or contract for use. People rarely read the detailed agreement, and companies may feel that the acceptance of the terms allows them to do as they please in using information concerning their clients. Researchers, on the other hand, are bound by professional ethical guidelines, obligations from funding sources and requirements from university ethics committees that require adherence to certain ethical principles and procedures. However, there is little consideration of how ethical guidelines that were developed before the advent of Big Data analyzes can transfer to research in the digital age. We recommend that every research team publishing on Information and Communication technologies in suicide prevention, should mention the ethical issues encountered and include a discussion of the ethical implications of their work, in peer-reviewed articles, conference proceedings as well as in-house reports. Because of the apparent lack of consideration of ethical issues in reports, it may be helpful to develop tools to researchers and stakeholders to better identify

potential ethical issues in their work, essential safeguards and the assumptions behind the implicit ethical choices that they make.

2.2. Introduction

Over 800 000 people per year die by suicide (WHO, 2014) and there is a consensus among current researches and policy makers that efforts need to be made to improve how suicide risk is both detected and treated (Rockett et al., 2016; WHO, 2014). To this end, experts in suicide prevention are employing two technologies, namely Big Data (BD) and Artificial Intelligence (AI) (Luxton, 2014). The application of these technologies is novel in the area of suicide research, and until now, their feasibility for and applicability to this area has not been investigated.

The term "Big Data" designates the ever-increasing amount of data collection and availability, and the development of tools to analyze them (Mayer-Schönberger & Cukier, 2013). Big Data is defined by its volume (the quantity of information), speed (how fast data are produced and circulated), resolution (the level of precision and accuracy of the data) and its predictability (the ability to forecast events) (Kitchin, 2014). It includes a wide variety of information, including metadata (Ouellet, Mondoux, Ménard, Bonenfant, & Richert, 2014). Metadata is often described as the "information about information", but it can be "any entity, form, or mode for which contextual data can be recorded" (e.g. when and where an email was sent) (Greenberg, 2003). Both data and metadata are increasingly being used to predict suicide risk from health-related information, such as Electronic Health Records or from social media, for instance of veterans in England and the United States (Goodwin, Wessely, & Fear, 2015; Poulin et al., 2014a). For instance, DARPA's Durkheim Project uses opt-in data from US veterans' Twitter and Facebook accounts to predict their suicide risk (Poulin et al., 2014). Another famous example was Sueki's analysis of internet searches of

deliberate self-harm-related Internet content to better understand the relationship with the mental states of Japanese adults (Sueki, 2012).

Artificial Intelligence, as defined by McCarthy, involves the study and design of intelligent machines (1998). Its aim is to :

> "*build machines that are capable of performing tasks that we define as requiring intelligence, such as reasoning, learning, planning, problem-solving, and perception*"
> (Luxton, 2015, p. 15).

Several technologies are associated with this field such as machine learning, neural networks, Natural Language Processing (NLP), machine computing, affective computing, machine perception and sensing and robotics (Luxton, 2015). Artificial Intelligence technologies are being used in suicide prevention for various purposes, including to analyze linguistic characteristics for the identification of suicidal individuals (Pestian et al., 2016). This field of research aims at answering long-lasting questions in suicide prevention : how to better identify at-risk indivudals. AI is also used to develop conversational interfaces (or *chatbots*) for example for helplines, such as Facebook 2018's initiative (Facebook, 2018). Facebook used machine learning in order to get "timely help to people in need". They attempted to find a way to classify individuals on the basis of a presumed level of suicidal risk.

In addition, the two terms BD and AI are sometimes criticized for having inconsistent technical definitions. While different, these two technological fields converge in practice (Moreno & Redondo, 2016). The two are sometimes used to describe the same techniques, in a undifferentiated way. However, it should be noted that the term AI is now being favored over BD (Hardy, 2016; O'Leary, 2013).

In recent years, considering the escalation in social media and new technology use, these two fields have become more important in daily life and have given rise to important concerns and considerations (Boyd & Crawford, 2012; Elish & Boyd, 2017).

Firstly, they pose new threats to privacy, confidentiality and data ownership (Kshetri, 2014; Mittelstadt & Floridi, 2016a; van der Sloot, 2015), and secondly, they have also illustrated the difficulty of obtaining informed consent of every individual when working with large datasets (Winter & Davidson, 2019). Both data and metadata have pose specific ethical challenges needing to be addressed (De Montjoye, Shmueli, Wang, & Pentland, 2014). These considerations have important implications for how we conduct healthcare research, particularly research pertaining to suicide (Luxton, 2014b; Poulin et al., 2016). Despite these considerations, the potential value and contribution of BD and AI are great and thus we are left with an important question: can AI and BD be used for suicide prevention, while remaining methodologically and ethically sound?

To date, there is no published review on the use of BD and AI in suicidology. This paper describes the current use of these technologies, focusing on how they are used by researchers to detect suicidal content and at-risk individuals online. Finally, given the underlying concerns around these technologies, this paper also examines specific ethical issues pertaining to the use of BD and AI in the context of suicide prevention.

2.3. Methodology

The methodology of this article was developed jointly with a team from the Centre for Research and Intervention on Suicide, Ethical Issues and End of Life practices (CRISE) at the Université du Québec à Montréal. This review was conducted with the goal of forming the basis for a programme of research on suicide and digital technologies.

2.3.1. A scoping study

The present study is a scoping review. Like mapping studies (Krysinska, Westerlund, & Niederkrotenthaler, 2017), this type of review provides readers with a broad overview of a specific topic. It also allows for the identification of research directions and needs in an emerging field. Scoping reviews can be useful in identifying studies

that could ultimately be used to conduct a meta-analysis. Its methodology can be as comprehensive and rigorous as a systematic review, but its nature is more descriptive (Munn et al., 2018). It differs in other ways: its research question is broad rather than highly focused, and it allows for the inclusion of all study types rather than defined study types (Arksey & O'Malley, 2005). It can be used as an initial step to identify knowledge gaps before conducting a systematic review on a specific topic (Munn et al., 2018).

This literature review follows the PRISMA guidelines (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) (Moher et al., 2015). The authors have used the general guidelines for systematic reviews rather than the guidelines for scoping reviews, since they are similar but more extensive. Furthermore, this review follows the recommendations from a descriptive analysis of over 300 systematic reviews, which suggests ways to improve transparency and reproducibility by mentioning the dates of the data collection (Beller, Chen, Wang, & Glasziou, 2013).

2.3.2. Dates

The data were collected between May 2017 and June 2018; it includes publications from January 2010 to June 2017.

2.3.3. Categories

In order to build the research syntax, the authors identified four conceptual categories Two are related to health care: a. suicide, b. prevention and intervention, and two concern technologies: c. BD and AI, d. Information and Communication Technologies (ICT). This last category was established in order to capture as many relevant studies as possible, since BD and AI are new domains. ICT can be defined as "Diverse set of technological tools and resources used to transmit, store, create, share or exchange

information" (UNESCO, 2019). They include computers, Internet (websites), live and recorded broadcasting technologies, telephony (including mobile, video chat) (2019). The category ICT was created in order to optimize the sensitivity of the search strategy. This is a challenge, because the terms used to describe ICT have varied greatly over the years. Furthermore, different terms are sometimes used in different countries speaking the same language (e.g. France and Québec).

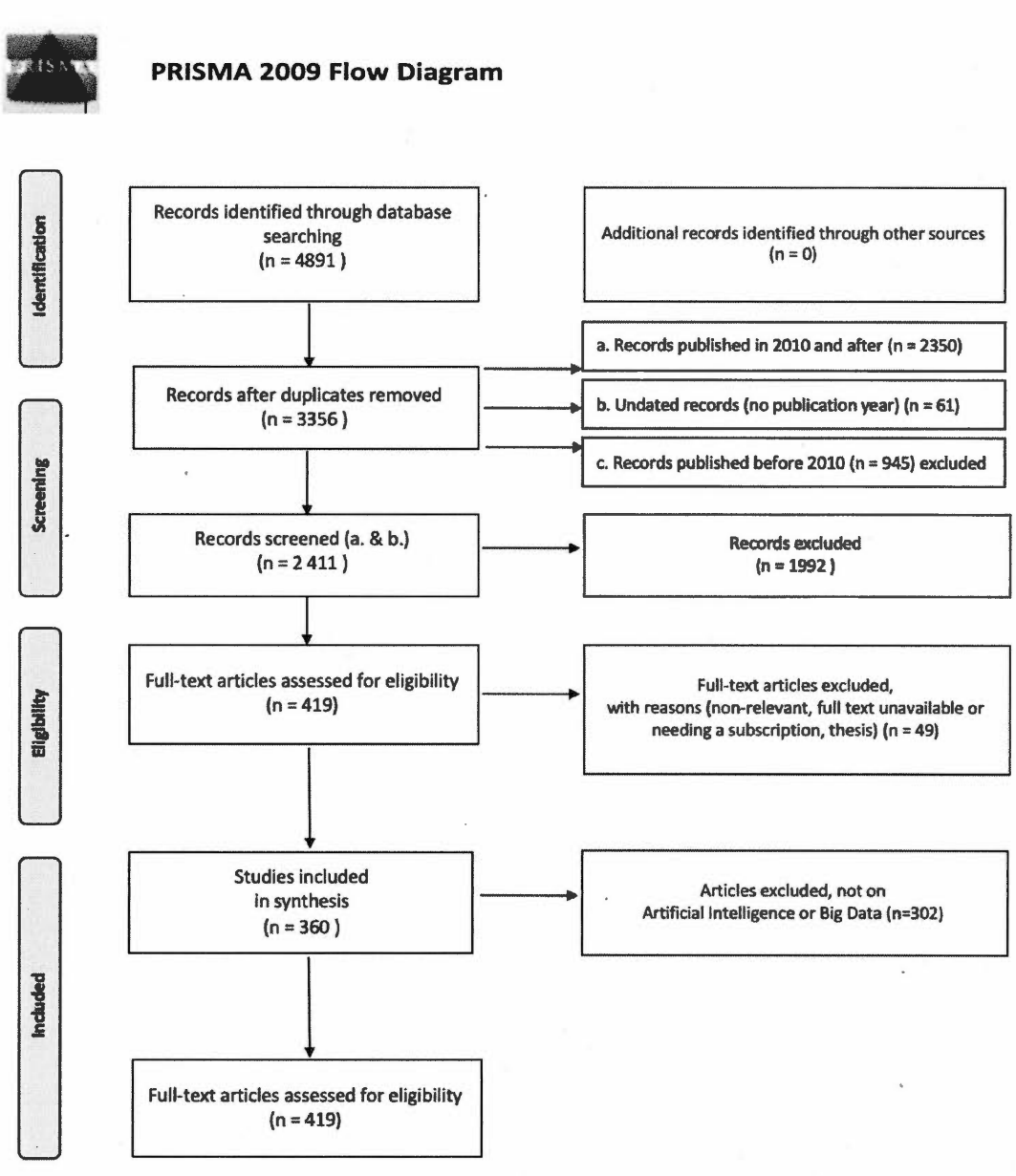
2.3.4. Databases

We searched four different databases in order include studies from different domains, ranging from health care to computer engineering. The health care databases were: PubMed (1,371 hits) and PsycInfo (2,107 hits). We included the human sciences database, ProQuest - Sociological Abstracts (1,019 hits). In the area of Computer Engineering, we searched IEEE Xplore (390 hits). In total, our search returned 4,891 records. The syntaxes used are listed in Table 1 and the summary of the results are reported in the PRISMA flowchart (see Figure 1).

2.1. Table 1: Syntax used in PubMed

PubMed - July 18, 2017

((((suicid*[TIAB] OR "self harm"[TIAB] OR "self-injurious behavior"[MeSH Terms] OR "self-destructive behavior"[MeSH Terms] OR "suicidal ideation"[MeSH Terms] OR "attempted suicide"[MeSH Terms] OR "deliberate self harm"[MeSH Terms] OR "self-injurious behaviors"[MeSH Terms])))) AND ((((algorithm*[TIAB] OR Facebook OR Internet OR online OR twitter OR "big data" OR cyber* OR instagram OR snapchat OR periscope OR "ask.fm" OR Internet OR "machine learning"[MeSH Terms] OR "deep learning"[TIAB] OR "machine learning"[TIAB] OR "algorithms"[MeSH Terms])))) OR (("mobile tech*"[TIAB] OR "social media"[TIAB] OR "electronic health record"[TIAB] OR "mining"[MeSH Terms] OR "mining"[TIAB] OR "neural network"[TIAB] OR "artificial intelligence"[MeSH Terms] OR "artificial intelligence"[TIAB] OR "computational intelligence"[TIAB] OR "natural language processing"[MeSH Terms] OR "natural language processing"[TIAB] OR "NLP"[TIAB] OR "mobile applications"[MeSH Terms] OR "cell phones"[MeSH Terms] OR "text messaging"[MeSH Terms]))))) AND (((Intervention*[TIAB] OR postvention*[TIAB] OR (clinic*[TIAB] AND chang*[TIAB]) OR monitoring[TIAB] OR therapeut*[TIAB] OR management[TIAB] OR treatment*[TIAB])) OR (("Program development"[MeSH Terms] OR "Program evaluation"[MeSH Terms] OR "Preventive Health Personnels"[MeSH Terms] OR "Safety Management"[MeSH Terms] OR "Occupational Health Services"[MeSH Terms] OR "Health Promotion"[MeSH Terms] OR "Patient Care Planning"[MeSH Terms] OR "Postvention"[MeSH Terms] OR "Treatment effectiveness evaluation"[MeSH Terms] OR "Pathology, Clinical"[MeSH Terms] OR "Clinical trial"[MeSH Terms] OR "Clinical decision making"[MeSH Terms] OR "Decision support systems, Clinical"[MeSH Terms] OR "Clinical Medicine" OR "Therapies, Investigational"[MeSH Terms] OR "Integrated Advanced Information Management Systems"[MeSH Terms] OR "Community Health Planning"[MeSH Terms] OR "Community Participation"[MeSH Terms] OR "Community Health Services"[MeSH Terms] OR "Delivery of Health Care"[MeSH Terms] OR "Community-Based Participatory Research"[MeSH Terms] OR "Psychotherapy"[MeSH Terms] OR "Early intervention"[MeSH Terms] OR "Crisis intervention"[MeSH Terms] OR "Clinical decision-making"[MeSH Terms] OR "Brief Therapy"[MeSH Terms] OR "Early intervention"[TIAB] OR "Treatment"[MeSH Terms] OR "Self-Help Techniques"[MeSH Terms] OR "Telemedicine"[MeSH Terms] OR "School-based intervention" OR "Workplace intervention"[TIAB] OR "Educational program"[TIAB] OR "Self-Help"[TIAB])))



2.2 Figure 1 : PRISMA diagram - flow chart

2.3.5. Deduplication strategy

We combined several deduplication strategies, in order to clean the number of duplicates and reduce the number of records from 4,891 to 3,356. First, we used the Systematic Review Assistant-Deduplication Module (SRA-DM) software from Bond University's Centre for Research in Evidence-Based Practice (CREBP). This software has a greater sensitivity and specificity than Endnote and Mendeley's deduplication tools in an initial stage (Kwon, Kang, & Lee, 2016). We then followed three steps: automated identification by the software, systematic verification of each record and manual validation by the researcher. We then performed two additional complementary deduplication procedures: one with Endnote (118 excluded records) and then one with Mendeley (4 excluded records).

2.3.6. Screening

ICT have rapidly evolved in recent years. We first conducted a review without any publication date limitations. This initial search identified studies that used obsolete technologies or websites now closed (e.g. Myspace). Therefore, we considered that the ethical and scientific considerations of these older articles might not be relevant to researchers in 2019. Furthermore, there were very few studies conducted on AI or BD before 2010, potentially because these technologies have only become popular in recent years. Thus, we decided to limit our study to peer-reviewed articles and peer-reviewed conference proceedings published since January 1, 2010. The record numbers were then reduced from 3,356 to 2,411 (945 studies published before 2010 were excluded, and 61 were not dated and had to be reinvestigated).

Four researchers read the title and abstract of all the 945 hits. The eligibility criteria were: 1. Records in English or French (languages spoken by the research team) and 2.

Studies on both ICT and suicide prevention. Suicide prevention is defined by the WHO as follows : "measures that can be taken at population, sub-population and individual levels to prevent suicide and suicide attempts" (WHO, 2018). Suicide prevention targets risk factors and can be universal, selective or indicated (WHO, 2014).

The six exclusion criteria were: 1. Studies on ICT but not on suicide prevention (or with suicide as a secondary variable), 2. Studies on suicide but not using ICT (or in a tangential or minimal way, or as a research method, such as a web survey), 3. Studies on mental health care in general or mental health disorders but not on suicide, 4. Articles on physical health or biological factors (e.g. pharmaceutical studies), 5. Studies using the research keywords, isolated from each other, without any relation, (e.g. many studies on suicide bombers were found) and 6. Duplicates, corrections, editorials, or studies in languages other than French or English.

The remaining records were verified for inclusion and exclusion by a team of five researchers: a psychology professor, an adjunct psychology professor, a librarian specialized in suicide prevention, and two doctoral students. In order to test the relevance of these exclusion and inclusion criteria, two inter-rater agreements were calculated on a corpus of 20 records, one between three raters, then one between two raters. The results were k(M1&M2)=0.5956, k(M2&M3)=0.7319, k(M3&M1)=0.7143.

We initially excluded 1,992 records, and retained 419. Of these 419 studies, 49 were excluded for the following reasons: non-relevant, integral text not accessible, or being a complete thesis. After this filtering step, 360 articles remained.

2.3.7. Classification

For this step, four researchers read the titles, abstracts, discussions, and conclusions of the articles. The introduction and methodology were read only if the team members were unsure of the content or the nature of the article, and needed more information. The categorization led to the creation of 28 distinct thematic categories. All four researchers reviewed all the articles to ensure that all the publications concern BD and AI were included. Our final record was comprised of 58 articles, which were divided into four categories:

- Infoveillance and infodemiology[4] (n=14)
- Tracking by sensors, captors and/or acoustic analysis (n=7)
- Detection of at-risk individuals by the analysis of databases (administrative or public) (n=19)
- Detection and monitoring of individuals by automated textual and linguistic analysis of Internet content (e.g. blogs, social media) (n=18)

2.3.8. Codification

Three research assistants (doctoral students specializing in suicide prevention) then analyzed all the information in these 58 articles. After reading the complete article, they used the NVivo software, version 11 and 12, to extract the relevant information.

The following categories were coded from the articles content: title, authors, country of the first author's affiliation, principal category (mainly BD, mainly AI, or both),

---

[4] Infodemiology is the "science of distribution and determinants of information in an electronic medium, specifically the Internet, with the ultimate aim to inform public health and public policy" (Eysenbach, 2011). Infoveillance is a term used to describe uses of infodemiology techniques "for surveillance purposes" (2011).

principal technology used (if specified), other technologies used (if specified), year, methodology, type of study, objectives, type of analysis, results, presence or absence of accuracy measures (recall, precision), number of suicidal participants, type(s) of suicidal behavior targeted, type(s) of data used, problems or limitations mentioned by the authors, ethical issues (if mentioned by the authors), potential of the article to be included in a meta-analysis - "evaluation potential" (according to the reviewers).

As mentioned in the introduction, AI and BD are sometimes used interchangeably. "Principal category" aims at identifying the main technology being used : BD, AI or both technologies employed in combination. The BD category includes studies where the highlight is on collecting large amounts of data in order to predict events, identify trends and patterns. The AI category includes studies focusing on automatically detecting or predicting risk, making decisions, creating algorithms or autonomous systems.

Demonstrating the efficacy of a prevention strategy is one of the criteria of scientific rigor (Gottfredson et al., 2015). The category "presence or absence of accuracy measures (recall, precision)" aims at discerning if the current state of literature would allow researchers to answer questions about the effectiveness of the technologies employed.

2.4. Results

2.4.1. Description of the records

Between 2010 (n=2) and 2016 (n=11), the amount of yearly publications on AI and BD in suicide prevention has increased fivefold. A majority of the 58 studies were original research articles (n=51, 87.9%) (see Table 2), three were case studies (5.2%), one commentary (1.7%), one book chapter (1.7%), one literature review (1.7%), one description of a research protocol (1.7%) The majority of these studies were published by researchers whose first affiliation country was the United States (n=33, 56.9%),

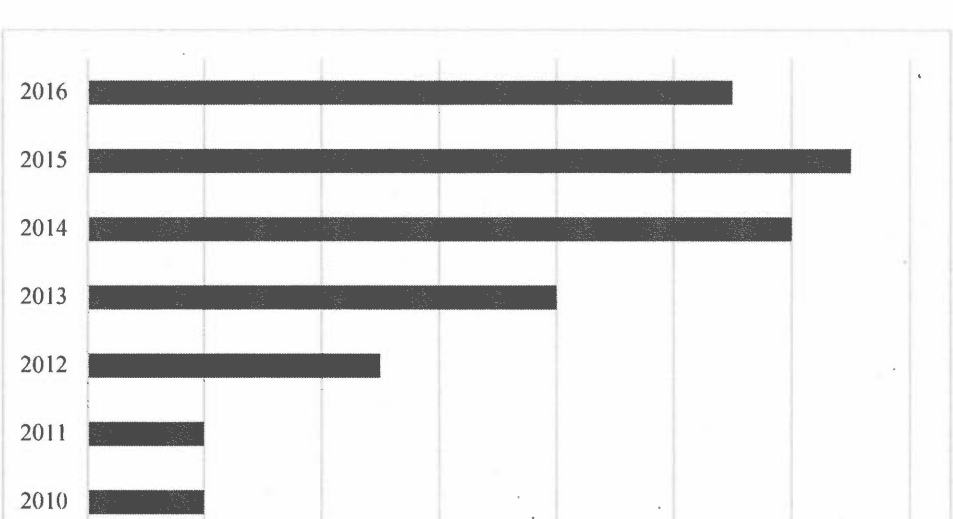
China (n=5, 8.6%) and Australia (n=4, 6.9%). It should be noted that the research papers category included a large number of conference proceedings (n=9, 15.5%).

In order to categorize and characterize the study designs, the research team used the classifications of Krysinska and colleagues in their mapping study on digital technologies in suicide prevention (Krysinska, Westerlund, et al., 2017b). The majority of included studies focused on content analysis (n=29) (e.g. computerized language processing). The rest of the studies are ecological studies (n=9), retrospective (n=6), quantitative (n=3), cross-sectional (n=3), non-empirical (e.g. literature review) (n=3, 5.2%) or labelled as case studies or "individual studies" (n=2) (see Table 2).

2.3 Table 2: Research designs by type of study

Research designs by type of study	N	%
Research Paper	**51**	**87.9%**
Content	1	2.0%
Cross-sectional	3	5.9%
Cross-sectional	1	2.0%
Ecological study	9	17.6%
Individual data	2	3.9%
Mainly Content Analysis	27	52.9%
Non empirical	1	2.0%
Quantitative (including randomized controlled trials [RCTs])	2	3.9%
Retrospective analysis (cohort or data)	5	9.8%
Case study	**3**	**5.2%**
Case study	1	33.3%
Mainly Content Analysis	1	33.3%
Retrospective analysis (cohort or data)	1	33.3%
Commentary	**1**	**1.7%**
Non empirical	1	100.0%
Book Chapter	**1**	**1.7%**
Mainly Content Analysis	1	100.0%
Review	**1**	**1.7%**
Non empirical	1	100.0%
Protocol	**1**	**1.7%**
Quantitative (including randomized controlled trials [RCTs])	1	100.0%
Grand Total	**58**	**100.0%**

Note 1: The indicated percentages have been rounded up to one decimal place. Thus, the total (100%) may differ slightly differ from the total of all the values.



2.4 Figure 2: Publication dates

Note: The number of articles published in 2017 is less than in 2016, because the data collection ended in June 2017. The year being not completed, 2017 was excluded from this chart.

2.4.2. Technologies used

There were almost as many studies on BD (n=25, 43.1%) as on AI (n=28, 48.3%). However, sometimes it was difficult to categorize the studies into one or the other because some studies examined both technologies simultaneously (n=3, 5.2%), or described general challenges related to these technologies (n=2, 3.4%).

Of the 58 studies, the most frequently used AI technology was machine learning (n=24, 41.38%). The main types of machine learning were support vector machines, classification algorithms and decision trees. The second type of technology used was data mining for Internet search queries (mostly using Google Trends) (n=11, 18.97%). Third, natural language processing represented 10.35% of all studies (6 studies). This technology was mostly used to perform advanced text analysis (see Table 3 and Table 4).

2.5 Table 3: Category

Technology	N	%
Artificial Intelligence	28	48.3%
Big Data	25	43.1%
Both	3	5.2%
Non-applicable (e.g. literature review)	2	3.4%
	58	100.0%

Note 1: The indicated percentages have been rounded up to one decimal place. Thus, the total (100%) may differ slightly differ from the total of all the values.

2.6 Table 4: Technologies and sub technologies named by authors

Technologies and subtechnologies	N	%
ML (Machine Learning)	**22**	37.9%
Classification Algorithm	3	
Decision tree algorithms	4	
Expectation–maximization algorithm	1	
Fuzzy Cognitive Map	1	
Fuzzy Pattern Recognition	1	
LDA (Latent Dirichlet Allocation)	1	
Neural Network	1	
Random Forest	1	
SVM (Support Vector Machine)	7	
Unspecified	1	
Unsupervised learning	1	
Internet search queries	**11**	19%
Google AdWords	1	
Google Trends	10	
NLP (Natural Language Processing)	**4**	6.9%
Classification Algorithm	1	
SVM (Support Vector Machine)	1	
Non-specified	2	
Technology development	**3**	5.2%
Information System Building	1	

National database building	1	
Search Engine building	1	
Data Mining	**2**	3.5%
Speech analysis	1	
Web scrapper	1	
Algorithm (classification or randomization)	**2**	3.5%
Classification Algorithm	1	
Non-specified	1	
ML (Machine Learning) & NLP (Natural Language Processing)	**2**	3.5%
Phenotype algorithm	1	
SVM (Support Vector Machine)	1	
Social Media API	**2**	3.5%
Twitter API	2	
Monitoring with wearable device	**2**	3.5%
Non-specified	2	
Voice Recognition	**2**	3.5%
SVM (Support Vector Machine)	1	
Non-specified	1	
Non-applicable (e.g. literature review)	**2**	3.5%
Non-specified	2	
Online Advertising	**1**	1.7%
Google AdWords	1	
Text Analysis	**1**	1.7%
Speech analysis	1	
Online Panel	**1**	1.7%
Non-specified	1	
NLP (Natural Language Processing) & Algorithm (classification or randomization)	**1**	1.7%
Conditional random fields	1	
Grand Total	58	100.0 %

Note 1: The indicated percentages have been rounded to one decimal place. Thus, the total (100%) may differ slightly differ from the total of all the values.

2.4.3. Type of suicidal behaviors studies

For each original study included in the review, the coders identified the type of suicidal behavior studied (as mentioned by the authors). The articles focused for the most part on suicidal ideation and/or behaviors (n=27, 46.6%), suicide attempts (n=9, 15.5%), and on deaths by suicide (n=15, 25.9%). For each category, coders found that some studies did not focus solely on one type of behavior.

2.7 Table 5: Suicidal Behavior targeted

Suicidal Behavior	N	%
Attempt	**9**	**15.5%**
Attempt only	*7*	*12.1%*
Mainly Attempt, also Suicide Death	*2*	*3.4%*
Suicide death only	**15**	**25.9%**
Suicide Death only	*15*	*25.9%*
Ideation and/or Behavior	**27**	**46.6%**
Mainly Ideation, also Suicide Death	*3*	*5.2%*
Mainly Ideation, also Attempt	*3*	*5.2%*
Mainly Ideation, also Behavior and Attempt	*3*	*5.2%*
Ideation and/or Behavior	*18*	*31.0%*
No Specific Suicidal Behavior Targeted	**7**	**12.1%**
	58	100.0%

Note 1: The indicated percentages have been rounded to one decimal place. Thus, the total (100%) may differ slightly differ from the total of all the values.

2.4.4. General objectives of the studies

The research team categorized the studies' objectives, as indicated by their authors. Each article could appear only in one category. Table 6 shows the general objectives of the studies. The most common article types identified were those which examined suicidal individuals by performing automated analysis of large datasets (public or

private) (n=19, 32.8%). Similarly, the second largest category of articles explored understanding or detecting suicidal individuals using linguistic analysis (using mostly Natural Language Processing) (n=18, 31%). There were also a significant number of articles in the infoveillance and infodemiology categories (n=14, 24.1%). This domain was mostly based on the analysis of social media data (through APIs, like the one provided by Twitter) or of internet search queries (with Google Trends). The goal of these studies was to provide a near real-time overview of population health indicators (including the frequency of searches on the use of lethal means). Another category was the use of sensors (e.g. infrared cameras), physiological sensors (such as heartrate monitors, or connected watches), or voice recording devices (n=7, 12.1%).

In the linguistic analysis category, the main technology used was AI (n=11; 61.1%). In the infoveillance category, BD was the favored technology (12 out of 14, 85.7% of its category). In terms of individual tracking using large datasets, AI was the main technology (n=19 out of 40, 47.4% of the category) followed by BD (n=8 out of 40, 42.1% of its category). Finally, in terms of sensors and acoustic analysis, AI was the main technology used (n=6, 85.7% of its category) (see Table 6).

2.8 Table 6: General objectives of studies mentioned by the authors

General objectives	N	%
Linguistic Analysis	**18**	**31.0%**
Artificial Intelligence	11	61.1%
Big Data	5	27.8%
Both	1	5.6%
Non-applicable (e.g. literature review)	1	5.6%
Infoveillance	**14**	**24.1%**
Artificial Intelligence	1	7.1%
Big Data	12	85.7%
Both	1	7.1%
Monitoring by automated analysis of datasets	**19**	**32.8%**
Artificial Intelligence	9	47.4%
Big Data	8	42.1%
Both	1	5.3%
Non-applicable (e.g. literature review)	1	5.3%
Use of sensors and acoustic analysis	**7**	**12.1%**
Artificial Intelligence	6	85.7%
Big Data	1	14.3%
Grand Total	58	100.0%

Note 1: The indicated percentages have been rounded to one decimal place. Thus, the total (100%) may differ slightly differ from the total of all the values.

2.4.5. Type of data used

Table 7 shows the type of data used in the studies. In the context of suicide research using AI or BD, the types of datasets available are important. Most of the studies have used data coming from EHRs (n=17, 29.3%), Internet related content (Internet search queries, Internet sites or blogs, *Google AdWords*) (n=14, 24.1%), social media (n=11, 18.97%), suicidal notes posted put online (n=5, 8.62%), or data collected from the suicidal individual (e.g. in large clinical datasets, such as interviews) (n=5, 8.62%), and

rarely from physiological monitoring (using wearable devices, voice analysis, heart monitors) (n=4, 6.9%).

2.9 Table 7: Type of data used

Type of data used	N	%
Electronic health records	**17**	**29.3%**
EMR (Electronic Medical Records)	6	35.3%
EHR (Electronic Health Records)	6	35.3%
Army health records	5	29.4%
Internet content	**14**	**24.1%**
Webpage content	1	7.1%
Search Queries & Public Data	5	35.7%
Search queries	4	28.6%
Blog content	3	21.4%
AdWords	1	7.14%
Social Media	**11**	**19%**
Social Media & Electronic Health Records	1	9.2%
Social Media	10	90.9%
Suicide notes	**5**	**8.6%**
Suicide notes	5	100.0%
Surveys, Questionnaires, Chat transcripts	**5**	**8.6%**
Surveys, Questionnaires, Chat transcripts	1	20.0%
Interviews	3	60.0%
Chat Transcripts	1	20.0%
Physiological monitoring / study	**4**	**6.9%**
Wearable devices	1	25.0%
Voice & Clinical Data	2	50.0%
Heartrate	1	25.0%
Non-applicable (e.g. literature review)	2	3.5%
	58	100.0%

Note 1: The indicated percentages have been rounded up to one decimal place. Thus, the total (100%) may differ slightly differ from the total of all the values. **Note 2**: Some authors used the term Electronic Medical Record some used the term Electronic Health Record. Because these two types of records might be very different, the terms mentioned in the studies were retained.

2.4.6. Ethical issues in suicide research using AI or Big Data

Most authors did not name or describe specific ethical issues that might have been encountered using BD or AI in suicide research (n=45, 77.6%), however a small number did mention encountering ethical challenges (n=13, 22.4%), most of which were empirical studies (n=11) (See Table 8). It should be noted that the mention of an ethical approval by an ethics committee was not considered by us to be mention of an ethical issue. We currently have a narrative description of the mentioned ethical issues in preparation.

2.10 Table 8: Presence and absence of mentioned ethical issues

Ethical issues	N	%
Absent	**45**	**77.6%**
Empirical research	39	86.7%
Unassigned	2	4.4%
Descriptive study	2	4.4%
Systematic review	1	2.2%
Cross-Sectional	1	2.2%
Present	**13**	**22.4%**
Empirical research	11	84.6%
Non-systematic review	1	7.7%
Descriptive study	1	7.7%
	58	100.0%

Note 1: The indicated percentages have been rounded up to one decimal place. Thus, the total (100%) may differ slightly differ from the total of all the values.

2.4.7. Comparison of the studies and their potential for scientific evaluation

Less than half of the studies provided enough information or data about their statistical measures to be potentially included in comparative studies such as meta-analysis (n=20, 34.5%). Most of the studies were assessed by the reviewers as potentially

challenging to evaluate or compare (38, 65.5%), because there were no figures, no description of the methodology, or no measuring points (such as recall and precision).

2.5. Discussion

2.5.1. A technological trend

Overall, AI and BD were most frequently used to detect at-risk individuals or levels of suicide risk. The main perceived advantages of using these technologies seem to be potentially improving detection systems, accessing new and larger datasets, and benefiting from more advanced tools to data analysis tools (Lewis, 2005; Lewis & Seko, 2016a). These potential benefits offer new research avenues for suicidologists. It is however necessary to conduct rigorous analyses in the future to establish if the promises outweigh the ethical challenges.

One of the greatest contemporary challenges in suicide prevention concerns online healthcare delivery (Christensen, Batterham, & O'Dea, 2014). Most of these studies are focused on the detection of suicide risk using various data sources (such as social media, Electronic Health Records). However, few studies proposed an intervention protocol based upon their detection system. One study which did provide this is the Durkheim Project, which combined a detection protocol for US veterans with suicidal ideations with an alert system for health care professionals (Poulin, Thompson, & Bryan, 2015).

Researchers are also interested in analyzing data in near-real time, using sensors and wearable devices to detect a potential suicide risk : for instance using wrist-worn accelerometers to monitor self-harm, or 3D scanners in prisons (Bharti, Panwar, Gopalakrishna, & Chellappan, 2017; Lee et al., 2014; Malott, Bharti, Hilbert,

Gopalakrishna, & Chellappan, 2015; Sasidharan & Kanagarajan, 2015). This coincides with the growing importance of connected medicine (e-health) in suicide prevention, which may prove to be useful because it could support and reinforce the work of clinicians (Vahabzadeh, Sahin, & Kalali, 2016).

2.5.2. Characteristics of the publications

The majority of the research conducted using AI and BD in suicide prevention comes from the United States (56.9%). This observation is consistent with the results of Krysinska and colleagues in a systematic review on suicide and the Internet (2017a). It could be due to the inclusion of articles published only in English and French. Also, we did not consider the country of affiliation of the other authors. However, the first author generally gives a good indication of where the study was designed and conducted.

It is important to note that 16 (27.6%) of the 58 studies are conference proceedings, extracted from IEEE Xplore. This result may be explained by the fact that we used a specialized database in computer sciences, IEEE Xplore Digital Library. To our knowledge, our scoping review is the first review in suicide prevention that searched this resource. It gave us interesting access to some innovative and very recent research that has not yet been published. However, there seems to be a big difference in what type of information is included in articles published in peer-reviewed journals, compared to conference proceedings. For instance, there seems to be less emphasis put into describing extensively all the statistical measurements used. This might be attributed to the more concise nature of a conference summary when compared to a published paper.

In computer sciences, it seems that peer-accepted conference proceedings are becoming an increasingly popular way of promoting the author's work and discoveries (Hermenegildo, 2012). One of the explanations might be that the field of computer sciences evolves and changes so fast that researchers favor a faster, more direct way of making their work known. Also, computer science researchers often do not have academic appointments, where publications in scientific journals are essential for promotions and considerations for tenure.

2.5.3. Ethics and evaluation

Approximately one-third (34.5%) of the articles were not detailed enough to allow an exhaustive comparison or evaluation of the results. This was primarily due to the fact that these articles provided few details about the methodology and measures. This point was consistent with a recurring observation that there is a lack of formal evaluation or evidence-based articles in online suicide prevention research (Jacob, Scourfield, & Evans, 2014; Lai, Maniam, Chan, & Ravindran, 2014). One explanation, might be the highly ethically challenging nature of suicide prevention. Researchers are very often limited in terms of methodological choices and encounter numerous privacy constraints, making their articles sometimes more observational and less replicable than in bio-medicine. For instance, it is almost impossible to use a double-blind design in suicide prevention (Mishara, Weisstub, 2012).

Our review showed that very few articles (22.4%) mention ethical issues related to BD or AI. These issues were related to privacy, data transparency and security, health-related risks, biases. This result might be surprising in the field of suicidology, where researchers are accustomed to ethical conversations and moral dilemmas (Mishara & Weisstub, 2005). It may also be surprising due to the fact that BD and AI are often associated with public surveillance, and have been criticized as being disrespectful of

privacy rights (Ouellet et al., 2014). Rapidly growing Internet use has focused attention on new moral issues and challenges, including the respect of privacy (Larsen et al., 2015) and the difficulty of obtaining informed consent in online studies (Lehavot, Ben-Zeev, & Neville, 2012). It is important that studies on digital technologies in suicide prevention digital technologies discuss the ethical considerations that may be involved in their research. This is a crucial point, given the fact that monitoring and detecting at-risk individuals, who may be at risk of dying by suicide, are being conducted by both public and private agencies and individuals, and ethical concerns about ensuring their safety are an important issue (Durkee, Hadlaczky, Westerlund, & Carli, 2011).

2.6. Limitations

The concept of BD is much criticized because it is hard to define (Kitchin & Lauriault, 2014). The articles that we grouped under this term could be categorized in a more refined and nuanced way, by using sub technologies. Our criteria in creating this category were based on methodological choices, such as presenting a summarized perspective of the use of these technological trends. Further studies could include collaboration between computer scientists and suicide prevention experts, in order to have an even more accurate technological portrait.

Our study includes first and foremost studies published in English, in English-speaking countries. Although we also included a few publications in French in the selection phase, none of them were included in the final corpus of studies. It would be interesting to conduct a review with more languages, or develop a joint initiative between counties, including Korean and Japanese studies, as there is a lot of research being conducted on new technologies and suicide in these two countries (Krysinska, Westerlund, et al., 2017a).

A potential limitation of this study could also be the characteristics of scoping reviews more generally. Our study shows how heterogeneous the research in this field is at the present time. This is a consequence of our deliberate choice to design a search methodology that is able to provide a diverse overview on the uses of BD and AI in suicide prevention. In future work, we hope to include several of these research studies in a meta-analysis.

Several recent cases of suicides purportedly linked to social media have been studied (Krysinska et al., 2017). It is possible that our scoping review missed a few. This is a difficult area to study because of the large semantic diversity in both suicidology and ICT. BD and AI (Boyd & Crawford, 2011), the terms are changing and the same activities may be called by different names depending upon the context. For example, Natural Language Processing is sometimes simply called "text analysis". It is very difficult to delineate stable technical terms, without being too broad.

2.7. Conclusion

BD and AI have been almost equally studied in suicide prevention, and they use many emerging techniques, such as Machine Learning, Natural Language Processing and near real-time analysis of data from social media (Twitter API) and Internet search engines (Google Trends). The main type of data analyzed were Electronic Health Records and Internet-related content (sites, blogs, Internet search queries, social media data).

The vast majority of studies aimed at analyzing linguistic content and conducting Infoveillance (or infodemiology). The studies mostly come from English-speaking countries and were for most part published between 2014 and 2016. It should be noted that there was great heterogeneity in the nature and format of the studies (from literature

review, to book chapter, to case study). The most frequently studied suicide topics were suicide behaviors, ideation, attempts, and deaths by suicide.

In general, more than three quarters of the articles did not mention ethical considerations. This result was surprising in the context of suicidology. This might be explained by the fact that a lot of these studies were technical reports or conference proceedings found on specialized databases (IEEE Xplore), that have rarely or almost never been refered to in suicide prevention (to our knowledge). It is important to have clear guidelines in suicide prevention, in order to facilitate an active partnership between the field of suicide prevention and computer sciences - both these fields having very different ethical challenges.

Given the importance of ethics in online suicide prevention (Mishara & Weisstub, 2007a), it is very important to expand the research on ethical issues related to the use of these technologies in suicide prevention. We should expect a sharp increase in the number and variety of studies in suicide prevention using ICT in the coming years.

2.8. Recommendations

We formulate two types of recommendations associated with the use of BD and AI in suicide prevention:

- Scientific recommendations: it would be of use to conduct a meta-analysis in the near future to assess the quality of the research being conducted. This may be a challenge for different reasons. Firstly, there is today a need for more consensus on the technical definitions of technologies such AI or BD. given that a lot of the research may currently being conducted in private companies and not in academic research institutions (such as in social media, (e.g. Facebook in their AI and suicide prevention initiatives). Private and academic initiatives should be thoroughly evaluated once a pilot project has been completed. Furthermore, there should be more replications to assess if new discoveries can be generalized to other populations, countries and clinical contexts.

  Ethical recommendations: one of our main recommendations is that each researcher publishing on ICT in suicide prevention, should mention the ethical issues encountered, whether they publish as peer-reviewed articles, or the research is published in a conference-proceeding. In addition, it could be helpful to develop tools to help researchers and stakeholders better identify the ethical issues involved.

Table 9: Prisma Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	Systematic review
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	Page 1 (to do)
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	Page 2
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	Page
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	Not necessary
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	Page 5
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	Page 4
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Page 4 et 5
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	Page 4 et 5
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	Page 4, 5, 6
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	Page 5
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	NA
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	Page 7
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	NA

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	Page 20
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	NA
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	Page 8
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	Page 6
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	NA
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	NA
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	NA
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	NA
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	NA
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	Page 16
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	Page 17
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	Information present
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	Information present

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

Table 10: Included Studies

General objective	Authors	Category	Main Technology	Ethical Issues
Linguistic Analysis	Braithwaite, S. R., Giraud-Carrier, C., West, J., Barnes, M. D., & Hanson, C. L. (2016)	Artificial Intelligence	ML (Machine Learning)	Absent
	Cheng, Q., Li, T. M., Kwok, C. L., Zhu, T., & Yip, P. S. (2017)	Artificial Intelligence	ML (Machine Learning)	Present
	Cherry, C., Mohammad, S. M., & de Bruijn, B. (2012)	Artificial Intelligence	ML (Machine Learning)	Absent
	Guan, L., Hao, B., Cheng, Q., Yip, P. S., & Zhu, T. (2015	Artificial Intelligence	ML (Machine Learning)	Present
	Huang, X., Zhang, L., Chiu, D., Liu, T., Li, X., & Zhu, T. (2014)	Artificial Intelligence	ML (Machine Learning) & NLP (Natural Language Processing)	Absent
	Larsen, M. E., Boonstra, T. W., Batterham, P. J., B, O. D., Paris, C., & Christensen, H. (2015)	Big Data	Social Media API	Present
	Lester, D. (2010)	Big Data	Text Analysis	Absent
	Liu, N. H., Contreras, O., Munoz, R. F., & Leykin, Y. (2014)	Big Data	Online Advertising	Absent
	Lv, M., Li, A., Liu, T., & Zhu, T. (2015)	Both	ML (Machine Learning)	Absent
	McCart, J. A., Finch, D. K., Jarman, J., Hickling, E., Lind, J. D., Richardson, M. R., . . . Luther, S. L. (2012)	Artificial Intelligence	ML (Machine Learning)	Absent
	Mitra, V., Shriberg, E., Vergyri, D., Knoth, B., & Salomon, R. M. (2015)	Non applicable	Voice Recognition	Absent
	O'Dea, B., Larsen, M. E., Batterham, P. J., Calear, A. L., & Christensen, H. (2017)	Big Data	Data Mining	Absent
	Pacula, M., Meltzer, T., Crystal, M., Srivastava, A., & Marx, B. (2014)	Artificial Intelligence	ML (Machine Learning)	Absent
	Ren, F., Kang, X., & Quan, C. (2016)	Artificial Intelligence	Algorithm (classification or randomization)	Absent
	Sueki, H., & Ito, J. (2015)	Big Data	Internet search queries	Absent

	Varathan, K. D., & Talib, N. (2014, August)	Artificial Intelligence	ML (Machine Learning)	Absent
	White, E., & Mazlack, L. J. (2011, June)	Artificial Intelligence	ML (Machine Learning)	Absent
	Wicentowski, R., & Sydes, M. R. (2012)	Artificial Intelligence	NLP (Natural Language Processing)	Absent
Infoveillance	Ayers, J. W., Althouse, B. M., Allem, J. P., Rosenquist, J. N., & Ford, D. E. (2013)	Big Data	Internet search queries	Present
	Bragazzi, N. L. (2013)	Big Data	Internet search queries	Absent
	Bruckner, T. A., McClure, C., & Kim, Y. (2014)	Artificial Intelligence	Internet search queries	Absent
	Chang, S. S., Kwok, S. S., Cheng, Q., Yip, P. S., & Chen, Y. Y. (2015)	Big Data	Internet search queries	Present
	Chen, P., Chai, J., Zhang, L., & Wang, D. (2014)	Both	ML (Machine Learning)	Absent
	Gunn, J. F., 3rd, & Lester, D. (2013)	Big Data	Internet search queries	Absent
	Jashinsky, J., Burton, S. H., Hanson, C. L., West, J., Giraud-Carrier, C., Barnes, M. D., & Argyle, T. (2014)	Big Data	Social Media API	Present
	Lee, D., Lee, H., & Choi, M. (2016)	Big Data	Internet search queries	Absent
	Ma-Kellams, C., Or, F., Baek, J. H., & Kawachi, I. (2015)	Big Data	Internet search queries	Absent
	McCarthy, M. J. (2010)	Big Data	Internet search queries	Absent
	Solano, P., Ustulin, M., Pizzorno, E., Vichi, M., Pompili, M., Serafini, G., & Amore, M. (2016)	Big Data	Internet search queries	Absent
	Song, J., Song, T. M., Seo, D. C., & Jin, J. H. (2016)	Big Data	Data Mining	Present
	Song, T. M., Song, J., An, J. Y., Hayman, L. L., & Woo, J. M. (2014)	Big Data	Internet search queries	Present
	Sueki, H. (2015)	Big Data	Online Panel	Present
Tracking	Anderson, H. D., Pace, W. D., Brandt, E., Nielsen, R. D., Allen, R. R., Libby, A. M., . . . Valuck, R. J. (2015)	Artificial Intelligence	NLP (Natural Language Processing)	Present
	Ben-Ari, A., & Hammond, K. (2015)	Artificial Intelligence	ML (Machine Learning)	Absent

Callahan, S. T., Fuchs, D. C., Shelton, R. C., Balmer, L. S., Dudley, J. A., Gideon, P. S., . . . Cooper, W. O. (2013)	Big Data	Algorithm (classification or randomization)	Absent
Goodwin, L., Wessely, S., & Fear, N. T. (2015)	Non applicable	Non applicable	Absent
Grunebaum, M. F. (2015)	Big Data	Non applicable	Present
Haerian, K., Salmasian, H., & Friedman, C. (2012)	Artificial Intelligence	ML (Machine Learning) & NLP (Natural Language Processing)	Absent
Hammond, K. W., & Laundry, R. J. (2014)	Artificial Intelligence	NLP (Natural Language Processing) & Algorithm (classification or randomization)	Absent
Hammond, K. W., Laundry, R. J., Oleary, T. M., & Jones, W. P. (2013)	Big Data	Technology development	Absent
John, A., Dennis, M., Kosnes, L., Gunnell, D., Scourfield, J., Ford, D. V., & Lloyd, K. (2014)	Big Data	Technology development	Present
Kessler, R. C., Stein, M. B., Petukhova, M. V., Bliese, P., Bossarte, R. M., Bromet, E. J., . . . Army, S. C. (2017)	Artificial Intelligence	ML (Machine Learning)	Absent
Kessler, R. C., Stein, M. B., Petukhova, M. V., Bliese, P., Bossarte, R. M., Bromet, E. J., . . . Army, S. C. (2017)	Big Data	ML (Machine Learning)	Absent
Lee, C.-N., & Ho, Y.-H. (2012)	Big Data	Technology development	Absent
McCoy, T. H., Jr., Castro, V. M., Roberson, A. M., Snapper, L. A., & Perlis, R. H. (2016)	Artificial Intelligence	NLP (Natural Language Processing)	Absent
Metzger, M. H., Tvardik, N., Gicquel, Q., Bouvry, C., Poulet, E., & Potinet-Pagliaroli, V. (2017)	Artificial Intelligence	ML (Machine Learning)	Absent
Poulin, C., Shiner, B., Thompson, P., Vepstas, L., Young-Xu, Y., Goertzel, B., . . . McAllister, T. (2014)	Artificial Intelligence	ML (Machine Learning)	Absent
Poulin, C., Thompson, P., & Bryan, C. (2016)	Both	ML (Machine Learning)	Present

	Rumshisky, A., Ghassemi, M., Naumann, T., Szolovits, P., Castro, V. M., McCoy, T. H., & Perlis, R. H. (2016)	Artificial Intelligence	ML (Machine Learning)	Absent
	Tran, T., Luo, W., Phung, D., Harvey, R., Berk, M., Kennedy, R. L., & Venkatesh, S. (2014)	Big Data	ML (Machine Learning)	Absent
	Tran, T., Nguyen, T. D., Phung, D., & Venkatesh, S. (2015)	Big Data	ML (Machine Learning)	Absent
Use of sensors and acoustic analysis	Malott, L., Bharti, P., Hilbert, N., Gopalakrishna, G., & Chellappan, S. (2015)	Artificial Intelligence	Monitoring with wearable device	Present
	Méndez, A. J., Lado, M. J., Vila, X. A., Rodríguez-Liñares, L., Alonso, R. Á., & García-Caballero, A. (2013)	Big Data	Monitoring with wearable device	Absent
	Pestian, J. P., Grupp-Phelan, J., Bretonnel Cohen, K., Meyers, G., Richey, L. A., Matykiewicz, P., & Sorter, M. T. (2016)	Artificial Intelligence	NLP (Natural Language Processing)	Absent
	Pestian, J. P., Sorter, M., Connolly, B., Bretonnel Cohen, K., McCullumsmith, C., Gee, J. T., . . . Group, S. T. M. R. (2017)	Artificial Intelligence	ML (Machine Learning)	Absent
	Rajeswari, A., Sowmbika, P., Kalaimagal, P., Ramya, M., & Ranjitha, M. (2016)	Artificial Intelligence	ML (Machine Learning)	Absent
	Scherer, S., Pestian, J., & Morency, L. P. (2013, May)	Artificial Intelligence	ML (Machine Learning)	Absent
	Venek, V., Scherer, S., Morency, L.-P., Rizzo, A. S., & Pestian, J. (2017)	Artificial Intelligence	Voice Recognition	Absent

2.9. Funding

This review has not been funded. But the use of a research team for the data collection has been possible thanks to project of Professors Réal Labelle and Cécile Bardon, financed by the Canadian Institutes in Health Research.

2.10. Conflict of Interests

The author declares no conflict of interests.

2.11. Ethics

In agreement with the Université du Québec à Montréal's Ethics Board rules, this research did not include any sensitive or identifiable information on individuals. It has been stated that it was not necessary to submit an ethical request (email received in march 2016).

2.12. References

Adkins. (2017). Machine Learning and Electronic Health Records: A Paradigm Shift. *The American journal of psychiatry, 174*(2), 93-94. doi:10.1176/appi.ajp.2016.16101169

Anderson, Pace, Brandt, Nielsen, Allen, Libby, . . . Valuck. (2015). Monitoring suicidal patients in primary care using electronic health records. *J Am Board Fam Med, 28*(1), 65-71.

Arksey, & O'Malley. (2005). Scoping studies: towards a methodological framework. *International Journal of Social Research Methodology, 8*(1), 19-32.

Ayers, Althouse, Allem, Rosenquist, & Ford. (2013b). Seasonality in seeking mental health information on Google. *Am J Prev Med, 44*(5), 520-525.

Beller, Chen, Wang, & Glasziou. (2013). Are systematic reviews up-to-date at the time of publication? *Systematic Reviews, 2*(1), 36.

Ben-Ari, & Hammond. (2015). Text mining the EMR for modeling and predicting suicidal behavior among US veterans of the 1991 Persian Gulf War. *System Sciences (HICSS), 2015 48th.*

Bharti, Panwar, Gopalakrishna, & Chellappan. (2017). Watch-Dog: Detecting Self-Harming Activities from Wrist Worn Accelerometers. *IEEE Journal of Biomedical and Health Informatics, PP*(99), 1-1. doi:10.1109/JBHI.2017.2692179

Boyd, & Crawford. (2011). Six Provocations for Big Data. *Computer, 123*, 1-17. doi:10.2139/ssrn.1926431

Boyd, & Crawford. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society, 15*(5), 662-679.

Bragazzi. (2013b). A Google Trends-based approach for monitoring NSSI. *Psychology research and behavior management, 7*, 1-8.

Google searches for suicide and risk of suicide., 65 271-272 (2014).

Callahan, Fuchs, Shelton, Balmer, Dudley, Gideon, . . . Cooper. (2013). Identifying suicidal behavior among adolescents using administrative claims data. *Pharmacoepidemiology and drug safety, 22*, 769-775.

Chang, Kwok, Cheng, Yip, & Chen. (2015a). The association of trends in charcoal-burning suicide with Google search and newspaper reporting in Taiwan: a time series analysis. *Soc Psychiatry Psychiatr Epidemiol, 50*(9), 1451-1461.

Chen, Chai, Zhang, & Wang. (2014). Development and application of a Chinese webpage suicide information mining system (sims). *Journal of medical systems, 38*(11), 88.

Chow, Leung, Lai, Chow, Chung, Tong, & Lit. (2012). Development of an electronic emergency department-based geo-information injury surveillance system in Hong Kong. *Injury, 43*(6), 739-748.

Christensen, Batterham, & O'Dea. (2014). E-health interventions for suicide prevention. *International journal of environmental research and public health, 11*(8), 8193-8212. doi:10.3390/ijerph110808193

De Montjoye, Shmueli, Wang, & Pentland. (2014). openpds: Protecting the privacy of metadata through safeanswers. *PLoS ONE, 9*(7), e98790.

Durkee, Hadlaczky, Westerlund, & Carli. (2011). Internet pathways in suicidality: a review of the evidence. *International journal of environmental research and public health, 8*(10), 3938-3952. doi:10.3390/ijerph8103938

Elish, & Boyd. (2017). Situating methods in the magic of Big Data and AI. *Communication Monographs*, 1-24.

Eysenbach. (2011). Infodemiology and infoveillance: tracking online health information and cyberbehavior for public health. *American Journal of Preventive Medicine, 40*(5), S154-S158.

Facebook. (2018, September 10, 2018). How Facebook AI Helps Suicide Prevention. Retrieved from https://newsroom.fb.com/news/2018/09/inside-feed-suicide-prevention-and-ai/

Flick. (2016). Informed consent and the Facebook emotional manipulation study. *Research Ethics, 12*(1), 14-28.

The future of "big data" in suicide behaviors research: Can we compare the experiences of the U.S. and U.K. armed forces?, 78 25-28 (2015).

Gottfredson, Cook, Gardner, Gorman-Smith, Howe, Sandler, & Zafft. (2015). Standards of evidence for efficacy, effectiveness, and scale-up research in prevention science: Next generation. *Prevention Science, 16*(7), 893-926.

Greenberg. (2003). Metadata and the world wide web. *Encyclopedia of library and information science, 3*, 1876-1888.

Grunebaum. (2015). Suicidology Meets "Big Data". *The Journal of clinical psychiatry, 76*, 383-384.

Gunn, & Lester. (2013). Using google searches on the internet to monitor suicidal behavior. *Journal of affective disorders, 148*(2-3), 411-412.

Haerian, Salmasian, & Friedman. (2012). Methods for identifying suicide or suicidal ideation in EHRs. *AMIA ... Annual Symposium proceedings. AMIA Symposium, 2012*, 1244-1253.

Hagihara, Miyazaki, & Tarumi. (2010). Internet use and suicide among younger age groups between 1989 and 2008 in Japan. *Acta Psychiatrica Scandinavica, 121*(6), 485-485.

Application of a Hybrid Text Mining Approach to the Study of Suicidal Behavior in a Large Population, 2555-2561 (2014).

Use of Text Search to Effectively Identify Lifetime Prevalence of Suicide Attempts among Veterans, 2676-2683 (2013).

Hardy. (2016). Silicon Valley looks to artificial intelligence for the next big thing. *The New York Times*.

Hermenegildo. (2012). Conferences vs. Journals in CS, what to do? *Evolutionary ways forward and the ICLP/TPLP model. Leibniz-ZentrumfürInformatik.*

Jacob, Scourfield, & Evans. (2014). Suicide prevention via the Internet: A descriptive review. *Crisis: The Journal of Crisis Intervention and Suicide Prevention, 35*(4), 261-267. doi:10.1027/0227-5910/a000254

Jashinsky, Burton, Hanson, West, Giraud-Carrier, Barnes, & Argyle. (2014b). Tracking suicide risk factors through Twitter in the US. *Crisis, 35*(1), 51-59.

John, Dennis, Kosnes, Gunnell, Scourfield, Ford, & Lloyd. (2014b). Suicide Information Database-Cymru: a protocol for a population-based, routinely collected data linkage study to explore risks and patterns of healthcare contact prior to suicide to identify opportunities for intervention. *BMJ Open, 4*, e006780.

Kessler, Hwang, Hoffmire, McCarthy, Petukhova, Rosellini, . . . Bossarte. (2017). Developing a practical suicide risk prediction model for targeting high-risk patients in the Veterans health Administration. *Int J Methods Psychiatr Res*.

Kessler, Stein, Petukhova, Bliese, Bossarte, Bromet, . . . Ur. (2017). Predicting suicides after outpatient mental health visits in the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). *Mol Psychiatry, 22*(4), 544-551.

Kitchin. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*: Sage.

Kitchin, & Lauriault. (2014). Towards critical data studies : Charting and unpacking data assemblages and their work. *Geoweb and Big Data*.

Krysinska, Westerlund, & Niederkrotenthaler. (2017). A mapping study on the internet and suicide. *Crisis*.

Krysinska, Westerlund, Niederkrotenthaler, Andriessen, Carli, Hadlaczky, . . . Wasserman. (2017a). A Mapping Study on the Internet and Suicide. *Crisis: The Journal of Crisis Intervention and Suicide Prevention*, No Pagination Specified-No Pagination Specified. doi:10.1027/0227-5910/a000444

Krysinska, Westerlund, Niederkrotenthaler, Andriessen, Carli, Hadlaczky, . . . Wasserman. (2017b). A Mapping Study on the Internet and Suicide. *Crisis*, 1-10.

Kshetri. (2014). Big data' s impact on privacy, security and consumer welfare. *Telecommunications Policy, 38*(11), 1134-1145.

Kwon, Kang, & Lee. (2016). Sexually transmissible infections in middle and high school students: experience rates, risk factors and relationship with mental health - results from the Korean youth risk behaviour web-based survey. *Sexual health, 13*(1), 29-34. doi:10.1071/SH15147

Lai, Maniam, Chan, & Ravindran. (2014). Caught in the web: a review of web-based suicide prevention. *J Med Internet Res, 16*(1), e30. doi:10.2196/jmir.2973

Larsen, Cummins, Boonstra, O'Dea, Tighe, Nicholas, . . . Christensen. (2015). The use of technology in Suicide Prevention. In *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference* (Vol. 2015, pp. 7316-7319).

Lee, & Ho. (2012). A suicide case information service platform for clinical staff. In *2012 International Conference on Machine Learning and Cybernetics* (Vol. 5, pp. 1921-1926).

Lee, Kim, Lee, Kim, Lee, Ju, & Myung. (2014). Detection of a Suicide by Hanging Based on a 3-D Image Analysis. *IEEE Sensors Journal, 14*(9), 2934-2935. doi:10.1109/jsen.2014.2332070

Lee, Lee, & Choi. (2016). Examining the Relationship Between Past Orientation and US Suicide Rates: An Analysis Using Big Data-Driven Google Search Queries. In *Journal of Medical Internet Research* (Vol. 18, pp. e35).

Lehavot, Ben-Zeev, & Neville. (2012). Ethical considerations and social media: A case of suicidal postings on Facebook. *Journal of Dual Diagnosis, 8*(4), 341-346. doi:10.1080/15504263.2012.718928

Lewis. (2005). *Concept mapping school personnel perceptions of adolescent suicide and its prevention.* State University of New York at Buffalo, Ann Arbor. Retrieved from https://search.proquest.com/docview/305379621?accountid=14719

http://openurl.uquebec.ca:9003/uqam?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:ProQuest+Dissertations+%26+Theses+Global&atitle=&ti

Lewis, & Seko. (2016a). A Double-Edged Sword: A Review of Benefits and Risks of Online Nonsuicidal Self-Injury Activities. *Journal of Clinical Psychology, 72*(3), 249-262. doi:10.1002/jclp.22242

Lovisi. (2012). The relationship between elderly suicide rates and the internet: A cross-national study. In *International Journal of Social Psychiatry* (Vol. 58, pp. 337-337).

Luxton. (2014b). Recommendations for the ethical use and design of artificial intelligent care providers. *Artificial Intelligence in Medicine, 62*(1), 1-10.

Luxton. (2015). *Artificial intelligence in behavioral and mental health care*: Academic Press.

Ma-Kellams, Or, Baek, & Kawachi. (2016). Rethinking suicide surveillance: Google search data and self-reported suicidality differentially estimate completed suicide risk. In *Clinical Psychological Science* (Vol. 4, pp. 480-484).

Malott, Bharti, Hilbert, Gopalakrishna, & Chellappan. (2015). Detecting self-harming activities with wearable devices. In *2015 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)* (pp. 597-602).

Mayer-Schönberger, & Cukier. (2013). *Big data: A revolution that will transform how we live, work, and think*: Houghton Mifflin Harcourt.

McCarthy. (1998). What is artificial intelligence?

McCarthy. (2010). Internet monitoring of suicide risk in the population. In *Journal of affective disorders* (Vol. 122, pp. 277-279).

McCoy, Castro, Roberson, Snapper, & Perlis. (2016). Improving Prediction of Suicide and Accidental Death After Discharge From General Hospitals With Natural Language Processing. In *JAMA Psychiatry* (Vol. 73, pp. 1064-1071).

Metzger, Tvardik, Gicquel, Bouvry, Poulet, & Potinet-Pagliaroli. (2016). Use of emergency department electronic medical records for automated epidemiological surveillance of suicide attempts: a French pilot study. In *International Journal of Methods in Psychiatric Research.*

Mishara, & Weisstub. (2005). Ethical and legal issues in suicide research. *International Journal of Law and Psychiatry, 28*, 23-41. doi:10.1016/j.ijlp.2004.12.006

Mishara, & Weisstub. (2007a). Ethical, legal, and practical issues in the control and regulation of suicide promotion and assistance over the Internet. *Suicide & life-threatening behavior, 37*(1), 58-65. doi:10.1521/suli.2007.37.1.58

Mitra, Shriberg, Vergyri, Knoth, & Salomon. (2015). Cross-corpus depression prediction from speech. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 4769-4773).

Mittelstadt, & Floridi. (2016a). The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts. *Science and Engineering Ethics, 22*(2), 303-341. doi:10.1007/s11948-015-9652-2

Moher, Shamseer, Clarke, Ghersi, Liberati, Petticrew, . . . Stewart. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews, 4*(1), 1-1. doi:10.1186/2046-4053-4-1

Moreno, & Redondo. (2016). Text analytics: The convergence of big data and artificial intelligence. *IJIMAI, 3*(6), 57-64.

Munn, Peters, Stern, Tufanaru, McArthur, & Aromataris. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Med Res Methodol, 18*(1), 143. doi:10.1186/s12874-018-0611-x

O'Leary. (2013). Artificial intelligence and big data. *IEEE Intelligent Systems, 28*(2), 96-99.

Ouellet, Mondoux, Ménard, Bonenfant, & Richert. (2014). " Big Data", gouvernance et surveillance.

Pestian. (2010). A conversation with Edwin Shneidman. In *Suicide and Life-Threatening Behavior* (Vol. 40, pp. 516-523).

Pestian, Grupp-Phelan, Bretonnel Cohen, Meyers, Richey, Matykiewicz, & Sorter. (2016). A Controlled Trial Using Natural Language Processing to Examine the Language of Suicidal Adolescents in the Emergency Department. In *Suicide & life-threatening behavior* (Vol. 46, pp. 154-159).

Pestian, Sorter, Connolly, Cohen, McCullumsmith, Gee, . . . Rohlfs. (2017). A machine learning approach to identifying the thought markers of suicidal subjects: A prospective multicenter trial. In *Suicide and Life-Threatening Behavior* (Vol. 47, pp. 112-121).

Poulin, Shiner, Thompson, Vepstas, Young-Xu, Goertzel, . . . McAllister. (2014a). Predicting the risk of suicide by analyzing the text of clinical notes. *PLoS ONE, 9*(1).

Poulin, Shiner, Thompson, Vepstas, Young-Xu, Goertzel, . . . McAllister. (2014b). Predicting the risk of suicide by analyzing the text of clinical notes. In *PLoS ONE* (Vol. 9).

Poulin, Thompson, & Bryan. (2015). Public health surveillance: predictive analytics and Big data. *Artificial Intelligence in Behavioral and Mental Health Care*, 205-205.

Poulin, Thompson, & Bryan. (2016). Public health surveillance: Predictive analytics and big data. In *Artificial Intelligence in Behavioral and Mental Health Care* (pp. 205-230).

Rajeswari, Sowmbika, Kalaimagal, Ramya, & Ranjitha. (2016). Improved emotional speech recognition algorithms. In *2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)* (pp. 2362-2366).

Rockett, Lilly, Jia, Larkin, Miller, Nelson, . . . Caine. (2016). Self-injury Mortality in the United States in the Early 21st Century: A Comparison With Proximally

Ranked Diseases. *JAMA Psychiatry, 73*(10), 1072-1081. doi:10.1001/jamapsychiatry.2016.1870

Rumshisky, Ghassemi, Naumann, Szolovits, Castro, McCoy, & Perlis. (2016). Predicting early psychiatric readmission with natural language processing of narrative discharge summaries. *Translational psychiatry, 6*(10), e921.

Vehicle cabin safety alert system, 1-4 (2015).

Reduced vowel space is a robust indicator of psychological distress: A cross-corpus analysis, 4789-4793 (2015).

Investigating the speech characteristics of suicidal adolescents, 709-713 (2013).

Solano, Ustulin, Pizzorno, Vichi, Pompili, Serafini, & Amore. (2016). A Google-based approach for monitoring suicide risk. *Psychiatry Research, 246*, 581-586.

Song, Song, An, Hayman, & Woo. (2014). Psychological and social factors affecting Internet searches on suicide in Korea: a big data analysis of Google search trends. *Yonsei medical journal, 55*, 254-263.

Song, Song, Seo, & Jin. (2016). Data Mining of Web-Based Documents on Social Networking Sites That Included Suicide-Related Words Among Korean Adolescents. *J Adolesc Health, 59*(6), 668-673.

Sueki. (2012). Association between deliberate self-harm-related internet searches and the mental states and lifetime suicidal behaviors of Japanese young adults. *Psychiatry and Clinical Neurosciences.*

Tran, Luo, Phung, Harvey, Berk, Kennedy, & Venkatesh. (2014). Risk stratification using data from electronic medical records better predicts suicide risks than clinician assessments. *BMC Psychiatry, 14.*

Tran, Nguyen, Phung, & Venkatesh. (2015). Learning vector representation of medical objects via EMR-driven nonnegative restricted Boltzmann machines (eNRBM). *Journal of biomedical informatics, 54*, 96-105.

UNESCO. (2019). Retrieved from http://uis.unesco.org/en/glossary-term/information-and-communication-technologies-ict

Vahabzadeh, Sahin, & Kalali. (2016). Digital suicide prevention: Can technology become a game-changer? *Innovations in Clinical Neuroscience, 13*(5-6), 16-20.

van der Sloot. (2015). How to assess privacy violations in the age of Big Data? Analysing the three different tests developed by the ECtHR and adding for a fourth one. *Information & Communications Technology Law*(April 2015), 1-30. doi:10.1080/13600834.2015.1009714

Adolescent suicidal risk assessment in clinician-patient interaction: A study of verbal and acoustic behaviors, 277-282 (2014).

Adolescent Suicidal Risk Assessment in Clinician-Patient Interaction, PP 1 (2017).

WHO. (2014). *Preventing suicide: A global imperative* (9241564776). Retrieved from

WHO. (2018). Suicide. Retrieved from https://www.who.int/news-room/fact-sheets/detail/suicide

Winter, & Davidson. (2019). Governance of artificial intelligence and personal health information. *Digital Policy, Regulation and Governance*.

3. CHAPITRE 3

THE CANADA PROTOCOL – MENTAL HEALTH AND SUICIDE PREVENTION (MHSP): DELPHI CONSULTATION AND EVALUATION OF AN ETHICAL CHECKLIST FOR THE REGULATION AND USE OF ARTIFICIAL INTELLIGENCE AND BIG DATA

Carl-Maria Mörch, M.Psy.

Abhishek Gupta, B.A.

Brian L. Mishara, Ph.D.

3.1. Abstract

3.1.1. Context

In order to improve current public health strategies, governments, researchers and private companies are increasingly using Information and Communication Technologies (ICT), and more specifically Artificial Intelligence (AI) and Big Data (BD) (Luxton, 2015; Mörch et al., 2018). These technologies are considered very promising but also very challenging ethically. Fields such as Mental Health and suicide prevention (Luxton, 2015; CCMH, 2017) are embracing these innovations but are also facing new challenges (Mishara & Weisstub, 2013a). Numerous ethical risks have arisen, most of which are rarely covered by current legal systems (Villani et al., 2017). Because the magnitude of these challenges, and the fact that AI developers feel overwhelmed by the issues, it is essential to better identify, prevent, and address potential ethical risks (IEEE, 2017).

3.1.2. Objectives

The Canada Protocol (MHSP) is a tool intended to guide and support professionals, users, and researchers using AI in Mental Health, and suicide prevention. Its goal is to identify ethical recommendations extracted from international reports on AI and Ethics, and on mental health and digital technologies.

3.1.3. Methods

A checklist was constructed based upon ten international reports on AI and Ethics and two guides on Mental Health and New Technologies. 329 recommendations were identified, of which 43 were considered as applicable to Mental Health and AI. The checklist was validated using a two round Delphi Consultation.

3.1.4. Results

16 experts participated in the first round of the Delphi Consultation and 8 participated in the second round. Of the original 43 items, 38 were retained. They concern five categories: "Description of the Autonomous Intelligent System" (n=8), "Privacy and Transparency" (n=8), "Security" (n=6), "Health-Related Risks" (n=8), "Biases" (n=8). The checklist was considered relevant by most users, but might need versions tailored to each category of target users.

3.1.5. Discussion

This development of this checklist has followed recommendations for the elaboration of guidelines in healthcare (Vermeulen, 2018). Although it covers the primary areas of possible concern discussed in the field, it does not pretend to be exhaustive. This checklist may be of use in healthcare, and more specifically in mental health and suicide prevention, to help ensure that key ethical issues are considered when using digital technologies.

3.2. Introduction

Mental Health Care researchers often mention that the "digital revolution" allows to connect services, individuals and new sources of data, at an unprecedented scale (Hollis et al., 2015). Tal and Torous even see it as an opportunity of compensanting (without replacing) an international lack of mental health professionals (Tal & Torous, 2017). For instance, mobile apps may also help accessing, evaluating and monitoring vulnerable populations, wherever they are ((East & Havard, 2015).

Another recent example is the advancement of Artificial Intelligence (AI) (Luxton, 2014). It brings new exciting promises, from better diagnosing clients (Masri & Jani, 2012), to improving clinical decision making (Bennett & Doub, 2016; Bennett & Hauser, 2013). This has become a major field of interest for multiple countries. For example, the Canadian government has recently promoted the concept of E-Mental Health. It defines it as the field that "uses the Internet and related technologies, like phone apps, to let patients receive care when and where they need it most, regardless of how close they live to their care provider" (Mahajan et al., 2014). The Canadian Mental Health Commission sees it as a way to help "people living with mental health problem or illness feel to more knowledgeable and better express their needs" (Mahajan et al., 2014; MHCC, 2017), in particular in suicide prevention (MHCC, 2018).

Mental health and suicide prevention are deeply connected. For the World Health Organization, suicide prevention is considered to be the priority condition which is the focus of their Global Strategy for Mental Health (WHO, 2017). This has become an international priority for Public Health. The 2014 WHO report "Preventing Suicide: A Global Imperative" incites all countries to reduce their suicide rates by 10% by 2020 (WHO, 2014). It also identifies several objectives: improving the way suicide rates are monitored, suicidal individuals are detected and psychological help is delivered

(Varathan & Talib, 2014; WHO, 2014). In order to meet these ambitious goals, governments, researchers and private companies are increasingly using Information and Communication Technologies (ICT), and more specifically AI and Big Data (BD) (Luxton, 2015; Mörch et al., 2018). In summary, working on the use of Artificial Intelligence in Mental Health may benefit from including the overlapping ethical knowledge and challenges of suicide prevention

The term "Artificial Intelligence" designates a very wide field of study, which may elicit both ethical concerns and hopes for more effectively prevention suicides and treating mental health problems.(CNIL, 2017). It is defined as the domain interested in the study and design of intelligent machines (McCarthy, 1998). Its goal is to "build machines that are capable of performing tasks that we define as requiring intelligence, such as reasoning, learning, planning, problem-solving, and perception" (Luxton, 2015, p. 15). The term has become popular, but is difficult to accurately define. A recent international report suggested it may be better to use the term: Autonomous Intelligent System (AIS) (IEEE, 2017). This refers to all the digital systems "designed to reduce the necessity for human intervention in our day-to-day lives" (2017). In the context of mental health care, some researchers have also used the term, Intelligent Care Providers (ICP) (Luxton, 2015). There are also many types of AI varying in methods and terminology. The most widespread type of AI is its sub-branch: Machine Learning (Demiaux & Si Abdallah, 2017). This branch of AI focuses on creating algorithms able to automatically learn or perform tasks (Schapire, 2008).

Generally speaking, the whole field is commonly associated with "Big Data"(BD). This term is defined as the constant increase of generated data and the development of techniques to analyze them (Kitchin, 2014). These two domains are often associated and raise numerous ethical questions. Many studies, books, and reports have been published on ethical issues concerning AI (Brundage et al., 2018; Demiaux & Si

Abdallah, 2017; IEEE, 2017). It is important to note that almost all these ethical dilemmas are new and go beyond the scope of the present legal systems (Villani et al., 2017).

Suicide prevention is no exception: the use of AI poses many new ethical, technical, and scientific challenges. There is however a paradox. Recently published studies that use BD or AI in suicide prevention rarely mention encountering ethical issues or challenges (Mörch et al., 2018). This relative absence of ethics in such a new and controversial field appears surprising. The lack of an ethical presence in suicidology writings is even more surprising in the current context where computer engineers and experts from around the world now talk of "moral overload" (IEEE, 2017). This expression means that AI developers and users frequently feel overwhelmed by the magnitude of the ethical risks, and do not feel competent to address those risks.

This article proposes a simple tool for people who want to use AI in mental health and suicide prevention: a checklist to identify and anticipate ethical issues, called the Canada Protocol – MHSP. This tool, created by the authors, was designed to cover the challenges identified in the field of suicide prevention and mental health, in ethics and AI and Information and Communication Technologies in health care. This paper presents the protocol and its validation process by an international committee of 16 professionals and experts, using a two-round Delphi Consultation.

The purpose of the creation of this checklist is to contribute to ethical education and to the improvement of practices when AI and BD are used in mental health care and suicide prevention. This checklist was constructed for use by AI developers, researchers, and decision makers willing to use AI in suicide prevention, and professionals and practitioners considering using AI.

3.3. Methodology

3.3.1. Checklist

The Canada Protocol – MHSP is a checklist. Checklists have long been used in mental health care, to elaborate list of potential symptoms in order to orient treatment (e.g. the Symptom 90 Checklist Revised (Derogatis, 1979; Derogatis & Unger, 2010)), or to better detect specific disorders (e.g. the Post-Traumatic Disorder Checklist) (Weathers, Litz, Herman, Huska, & Keane, 1993). Checklists are generally used to help clinicians, but they can also be used to help researchers assess the quality of studies (e.g. the National Institute for Clinical Excellence's checklist – NiCe) (Excellence, 2009). An important example is the *Preferred Reporting Items for Systematic Reviews* (PRISMA), which was designed by the RAND foundation to improve the quality of systematic reviews (Moher, Liberati, Tetzlaff, & Altman, 2009). The PRISMA checklist is comprised of 27 items and a flow diagram. The original intent of the creators was to increase awareness of what constitutes a scientifically sound review. This approach favors guidance over sanctioning. The Canada Protocol was inspired by this type of approach. The authors tailored a checklist to their field of study, AI in mental health care and suicide prevention.

3.3.2. Design of the checklist

The authors did not identify an appropriate theoretical framework as guidance during the elaboration of the checklist. However, in order to scientifically reinforce and strengthen this study's approach, the European Society of Human Reproduction and Embryology's (ESHRE) best practices in the design of medical guidelines was used for inspiration (Vermeulen, 2018): 1. Choose a topic, 2. Create a group to develop the guidelines, 3. Scope the guidelines, 4. Formulate Key Questions, 5. Find evidence, 6.

Evidence Synthesis, 7. Recommendation development, 8. Writing the guideline draft, 9. Stakeholder Consultation, 10. Approval, 11. Publication, Dissemination and Implementation.

In developing our checklist, we followed most of the cited steps. Some were excluded because the Delphi consultation already corresponded to steps 9 and 10 and as previously mentioned, this study does not create guidelines, but a checklist built on guiding ethical principles.

After having chosen the topic (step 1), the development group composed of two researchers was created: one specialized in psychology and digital technologies (Carl-Maria Mörch, lead author) and one specialized in ethics and AI (Abhishek Gupta). They met on a regular basis (step 2), twice a month for six months. The initial work consisted of gathering existing recommendations from 79 records, including international reports on AI and ethics, scientific articles on this topic, and guidelines in ethics in mental health and digital technologies. It was determined that developing new recommendations on this topic would be less pertinent than regrouping and selecting existing recommendations and guidelines formulated by experts. This approach addresses a current problem in psychology: too many guidelines exist and it has become unclear which ones to trust or favour (Drife, 2010). By following and promoting existing recommendations, the authors hoped to increase the relevance of their work. The team read and extracted each mention of ethical issues in all documents concerning Artificial Intelligence. Overall, 450 potential ethical challenges, biases and risks were identified. After eliminating duplicate items, all the non-pertinent or redundant items were filtered. The items that were too specific, unclear, not based on an ethical consideration (e.g. too technical) were considered non-pertinent. The remaining 329 items were categorised by Gupta and Morch into 9 subsets : “Fairness and Biases”, “Introductory Questions”, “Methodological Issues”, “Social Relevance and Validity”, “Transparency & Explainability”, “Controllability & Security”,

"Autonomy", "Responsibility". The two team members split the total amount of items into two equal sets. After each one had categorized his own items, both checked the categorization of the other. If there was a disagreement or proposition for re-categorization, the two researchers had to reach a consensus.

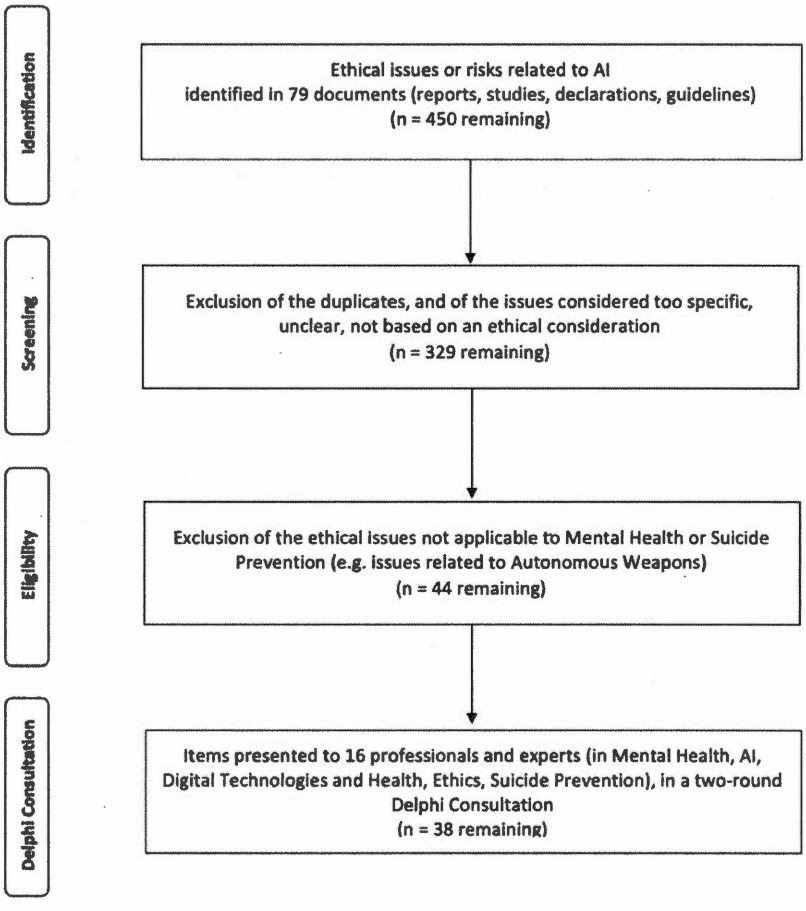


Figure 1 : Flow Diagram

Of these 329 items, the lead researcher excluded all the articles that were not on Artificial Intelligence, were not applicable to the field of suicide prevention and mental health, were too technical for mental health professionals, or recommendations that

were only applicable to the military (e.g. drones). 285 items were excluded and only 44 were retained. A scientific supervisor specializing in digital technologies and ethics in suicide prevention (Brian Mishara) revised the checklist, its formulation and categorization. The first version of the checklist was comprised of 5 categories:

- "Description of the Autonomous Intelligent System" (n=12). This category includes important issues related to the description, general functioning and advertisement of a intelligent system (e.g. what are the claims of efficacy, who is funding it).
- "Privacy and Transparency" (n=8). This category includes ethical issues and risks related to privacy issues (e.g. respect of anonymity), consent and algorithmic transparency.
- "Security" (n=6). This category includes ethical issues and risks related to data and algorithmic security.
- "Health-Related Risks" (n=9). This category includes ethical issues and risks related to mental and physical health (e.g. contingency planning in case of suicidal crisis).
- "Biases" (n=9). This category includes ethical issues related to scientific (e.g. false positive) or social biases (e.g. discrimination).

3.3.3. Theoretical context

The Canada Protocol Checklist's approach has two main theoretical frameworks. First, its goal is to facilitate responsible and civic education on digital technologies, as expressed by "Civic Media" (Zuckerman, 2014). This movement is interested in the use of digital technologies to promote social change and enhance civic participation, including in health care. MIT's Media Lab defines it as : "any use of a medium that empowers a community to engage within and beyond the people, places, and problems of their community" (MIT, 2011). By creating this tool, we hope to provide the mental

health and suicide prevention communities, with a way of realizing the existence of large ethical issues outside of their field of specialization. Our goal is to help them not feeling overwhelmed and help resolve ethical issues that may occur.

Secondly, this study is influenced by the critical studies on BD and AI (Boyd & Crawford, 2012; Ménard, Mondoux, Ouellet, & Bonenfant, 2016) to consider technological innovations from a scientific and epistemological point of view. These studies analyze the use of digital technologies jointly with the social discourse around them. For instance Big Data and AI are often considered and presented as revolutionary, but Boyd and Crawford, in their seminal article "Critical questions for Big Data," mention that they are often described with a utopian rhetoric. In general, BD and AI are criticized for their potential biases and ethical challenges (Bowker, 2014; Bunnik, Cawley, Mulqueen, & Zwitter, 2016; Leetaru, 2016). By considering these critical studies in relation to mental health and suicide prevention, we hope to promote a critical and more responsible use of these technologies.

3.3.4. The Delphi Methodology

In order to reinforce the internal validity of the checklist, the authors used a consultation methodology: the Delphi Method, developed in the 1950's by the RAND (Research and Development) Foundation (Letrilliart & Vanmeerbeek, 2011) and since used regularly in health care. It is an iterative method that consists of 2-or-more rounds of consultation with experts on a specific topic or tool. Typically, the expert panel has to study content submitted by a research team, and then give their opinion anonymously. In each round, the experts see the aggregated results, and sometimes a recall of his or her own responses. The procedure stops when a consensus has been reached (2011). A panel has been considered to require a minimum of 15 participants (Hasson, Keeney, & McKenna, 2000).

For this study, the authors determined that two rounds would suffice, because the content was already conceived and the task required was straightforward: recommending to reject, exclude, or modify (with justification). In the present study, participants were not asked to create content, but only to agree on content. The amount of rounds is usually determined by the initial needs (Ludwig, 1994); the more complex the needs, the higher the number of rounds required.

3.3.5. Experts, users and professionals consulted

To create a panel of experts:

- First, a list of experts (in Mental Health, Suicide Prevention, AI and New Technologies, as well as AI users) was established (see results section) who were sent individual invitations (see results section)
- Two Canadian academic research organizations were contacted that specializes in a. health care, b. healthcare and digital technologies, and c. ethics and artificial intelligence. Group invitations were sent through their mailing lists.

3.3.6. Consultation process

First step: An initial list of 44 items was established, divided into 5 categories: "Description of the Autonomous Intelligence System", "Privacy and Transparency", "Security", "Health-Related Risks", "Biases". The consultation was conducted using the Lime Survey online software, hosted on the Université du Québec à Montréal's servers (UQAM). The consent form was accessible on the first page of the survey, as a PDF in English. If the response to the consent form was negative, it was not possible to participate in the survey. The survey was anonymous, (Linstone & Turoff, 2002), although participants were requested to provide an email address to be contacted for

the second round, which would not be associated with their responses in order to maintain anonymity.

The instructions for the first round were to: a. approve (explain the reason), b. exclude (explain the reason), or c. modify (explain the reason). The data collection phase lasted 60 days, from April 15, 2018 to June 15, 2018.

At the end of this first phase, the research team wrote a report with all the modifications completed, along with a detailed summary of the number of items excluded, modified, or kept as is. To be retained, an item had to have a score of 80% agreement or higher at the "kept as is" option (Jorm, 2015). To be excluded, an article needed to have 50% of agreement or less as "kept as is" or 80% or more to "exclude the item".

Second step: an edited survey was sent to the participants who provided their email address (16 participants). The summary and report were included in the invitation email. The second Delphi round requested the experts to revise the modifications and verify if the updates were satisfactory. Participants were then asked to either keep each item or exclude it. This step followed the recommendations of the COSMIN checklist (Mokkink et al., 2010).

3.4. Results

3.4.1. Panel

32 experts were invited to participate by individualized emails. These participants were: a. professors, experts and researchers in suicide prevention and information and communication technologies (n=12), b. professors and researchers who have published on AI in Mental Health (n=10), c. researchers who have published on Ethics and AI (n=3), d. researchers in psychology or psychiatry, experts in the use of Information and Communication Technologies (n=2), e. experts in Ethics and AI (n=2), f. professor of

computer engineering specializing in AI (n=1), f. two entrepreneurs working on the use of AI in Mental Health and Suicide Prevention (n=2).

3.4.2. Response

16 responses were collected in the first round. This is above the suggested minimum of 15 for a first round of a Delphi survey (Hasson et al., 2000). The participants were distributed as follows : 5 were specialized in AI and Mental Health, 5 were specialized in Technologies and Mental Health, 6 were specialized in Suicide Prevention and New Technologies.

The two aforementioned Canadian research groups distributed a collective invitation to participate in their mailing lists. The group specialized in Information and Communication Technologies, Media Studies and Health has a mailing list of over 100 members. The group specialized in Ethics and AI has a mailing list of 49 researchers across Canada.

Given that the participation to a Delphi consultation is anonymous, it is difficult to accurately estimate the response rate to the first phase. In the second round, the 16 participants of the first round were invited, 8 participated. Which corresponded to a response rate of 50%.

3.4.3. Final checklist

The initial checklist included 44 items. After the first round of consultation, 12 items were retained as is or slightly modified, (condition: 80% or more of agreement at the "keep the item as is" option), 27 items were modified (condition: between 50% and 80% of agreement at the option "keep the item as is" or at the option "modify the item"), 6 items were excluded (less than 50% of agreement on "keep as is", or more than 50% at the option "remove the item"). The second round consisted of a final

review of modifications of the 27 items, conducted by 8 of the 16 original participants partook in the second round. The final checklist consisted of 40 items (see Table 1):

- "Description of the Autonomous Intelligent System" (n=8)
- "Privacy and Transparency" (n=8)
- "Security" (n=6)
- "Health-Related Risks" (n=8)
- "Biases" (n=8)

3.4.4. Relevance and target users of the checklist

In the first round of the Delphi, the 16 participants were asked to say if they thought this instrument was relevant, using a Likert Scale ranging from 1 to 5 (1 being not relevant to 5 being very relevant). On average, the MHSP was considered relevant by the participants with an average of 4.06 on 5 (ety=0.68).

When asked who could use this checklist, participants could not stage assert categorically who could use it. They were mostly uncertain for researchers (56.25%), AI developers (81.25%) or mental health professionals (68.75%). The rejection rate ("No") was two times 0% (for researchers and AI developers), and once 18.75% (for mental health professionals). This might indicate that the ethical considerations of the checklist are maybe not relevant to the day-to-day jobs of mental health professionals. The detailed results (see Table 2) indicate that this checklist is relevant, but might require in the future to be tailored depending on who could use this tool.

3.1 Table 1 : The Canada Protocol – MHSP

	DESCRIPTION
Objectives	Describe your project's objectives and/or rationale and describe the role and functioning of your Autonomous Intelligent System

Technology	Name and describe the technologies and techniques used (e.g. supervised or unsupervised learning, machine learning, random forest, decision tree...). You can refer to the report of the AI Initiative incubated at Harvard http://ai-initiative.org/wp-content/uploads/2017/08/Making-the-AI-Revolution-work-for-everyone.-Report-to-OECD.-MARCH-2017.pdf. Mention the names of any technological intermediary or supplier allowing you to use the technology (e.g. technical provider, cloud provider)
Funding & conflict of interest	Indicate all sources of funding for your project (public and private) and who might have an interest (e.g. financial, political) in your Autonomous Intelligent System
Credentials	If you have noted that you or someone in your team has an expertise in relation to the Autonomous Intelligent System (e.g. in a document, a webpage, an interview), clearly indicate the name of the professional, their technical, academic or medical credentials, and their training (e.g. "Professor Smith, PhD in computer systems engineering from Harvard University. Specialist in the Online Detection of Depression")
Target population	Describe your target population and its size, or identify its subgroups and their sizes. Describe if and how the target population (and, or its subgroups) assisted in the design of your Autonomous Intelligent System.
Evidence	If you made claims about your Autonomous Intelligent System's efficacy, performance, or benefits, please justify them and provide the evidence underlying them. If you have mentioned or used scientific papers, please cite your sources
Testing	If you have run your Autonomous Intelligent System under adversarial examples or worst-case scenarios, describe the type of tests used and their outcomes
Complaints	Describe the process whereby users can formally complain or express their concerns about your Autonomous Intelligent System

PRIVACY & TRANSPARENCY

Responsibility	Describe who will be legally accountable for your Autonomous Intelligent System's actions or decisions
Data collection	Describe what data have been collected and used (for the training, evaluation and operational phases), where they are stored, who collected the data, who will have access to the data, and what safeguards are in place to ensure secure storage
Accessibility	In all the documents or texts, confirm that you have used a language adapted to target users and, when relevant, accommodated special needs some users may have.
Informed consent	State whether you have obtained informed consent and, if so, how, when, and from whom. Describe its nature (formal, implied, renewable, dynamic) and include the exact wording on the consent form. Note whether you have received ethical approval from an institution (eg: hospital, university) for your consent forms
Consent withdrawal	State whether you have specified the duration of the consent and whether you have implemented consent withdrawal mechanisms (e.g. opt-out clause, unsubscribe option). Specify what happens if an user wants to stop using the AIS or delete his or her information
Access to the data	Access to the data: State if an individual can access any data related to him or her and obtain the data in a clear and structured export document. If this is not possible, explain why
Right to be forgotten	Describe whether an individual can retrieve and erase all of his or her information, and if so, how. Describe the mechanism
Minors	Note whether information concerning minors is used for the Autonomous Intelligent System. If it is, and it is intentionally collected, please indicate whether parental consent is required. If it is, and it is unintentionally collected, please describe what can be done to remove this information

SECURITY

Embedded recording mechanism	If you have used a technology to monitor and record all your Autonomous Intelligent System's decisions and actions, detail how and in what circumstances these records could be made available to authorities, external observers or auditors
Third-parties	Indicate who has access to the data (individuals and organizations), and whether identifying information about participants is included in accessible data
Data protection	Detail all the measures taken to protect any sensitive and personal information
Audit trails	Explain who has access to the data and when
Autonomy	Explain if your system has the autonomy to take actions or make decisions on its own. If yes, detail the degree of autonomy of your Autonomous Intelligent System (e.g. partial or complete)
Moderation	Explain if your Autonomous Intelligent System requires human intervention or moderation. If yes, describe who will have access to your Autonomous Intelligent System, and what will the guides regulating their intervention be
	HEALTH-RELATED RISKS
Type of care	Is your Autonomous Intelligent System helping its owners to provide the target population with the optimal treatment or treatment as usual? Indicate the criteria (and their sources) for optimal treatment or treatment as usual
Crisis & contigency planning	List the criteria for evaluating the risk exposure of your Autonomous Intelligent System. Describe your plan in case of emergency, disaster, or suicidal crisis (the intervention protocol). If possible, specify what type of behaviors and environments are considered as being at risk and explain the rationale in a simple way
Non-maleficience	Explain whether your Autonomous Intelligent System could harm, incommode, or embarass a user and, if so, how. Explain how you avoid or minimize this risk
Misuse	Describe potential misuses of your Autonomous Intelligent System (e.g. describe a possible negative scenario to indicate what could potentially happen to a user) and describe your mitigation strategies
Emotions detection	If your Autonomous Intelligent System detects user's emotions, state how, and for what purpose. Explain whether the user is informed and if so, how
Emotions control	If your Autonomous Intelligent System can provoke emotions, describe how users are informed of this possibility, the emotions that may be provoked, their intensity, and possible impact on users
Relationship	Is the user aware that he or she is interacting with a machine? Describe whether your Autonomous Intelligent System can create a relationship with users, and if so, how. Describe how the relationship might affect a user
Public awareness	Describe the impact on users and potential users of public dissemination of information about your Autonomous Intelligent System and the process of its development
	BIASES
Ethics	If you have requested an expertise on ethics during the design of your Autonomous Intelligent System, detail the parties involved and their contributions
Exclusion & discrimination	Explain if there are risks of exclusion or discrimination related to your Autonomous Intelligent System (e.g. based on gender, race, age, religion, politics, health, sexual orientation, etc.)
Stigmatization	Describe how you avoided using languages, images, and other content that could stigmatize users (e.g., reference to guidelines on safe media reporting and public messaging about suicide and and mental illness)

Detection	If applicable, explain any potential detection errors that might be made by your Autonomous Intelligent System (e.g. false positives, false negatives) and estimate their extent (e.g. precision, recall). Describe any potential adverse consequences for users. If applicable, describe any incidental finding made by your Autonomous Intelligent System
Data handling	If applicable, describe the nature and purpose of any data manipulation (e.g. cleaning, transformation) and by whom they were performed. Describe what will be done with the metadata
Data selection	Describe where the data came from, how you accessed them (e.g. through an API) and if you think there might be a selection or sampling bias (e.g. the data comes from an API or a spectrum bias)
Data transformation	If applicable, describe the nature and purpose of any statistical transformations applied to your data. Describe any potential bias or risk related to the data transformation (e.g. ecological fallacy, confounding factors)
Other	If you have identified other potential methodological or scientific biases, describe them and their potential ethical consequences (e.g.1. an excessively long consent form could affect the informed consent; e.g.2. the presence of a floor effect in the measurements could constrain an Autonomous Intelligent System's ability to detect a behavior)

3.2 Table 2 : Final questions

Question	Yes	No	Uncertain
Will researchers use this checklist ?	7 (43.75%)	0 (0%)	9 (56.25%)
Will AI developers use this checklist ?	3 (18.75%)	0 (0%)	13 (81.25%)
Will mental health professionals use it?	2 (12.5%)	3 (18.75%)	11 (68.75%)

3.5. Discussion

In all the studies and reviews on AI and ethics, the research team identified only three ethical tools. None of them are specialized in Mental Health or Suicide Prevention. The first, "Geneth: A general ethical Dilemma Analyzer" (Anderson & Anderson, 2014) is a general ethical dilemma analyzer using Machine Learning. The second ethics tool is called "DELICATE: A Checklist for Trusted Learning Analytics" (Drachsler & Greller, 2016). This checklist asks the user or reader two or three questions concerning 8 key actions: determination, explain, legitimate, involve, consent, anonymize, technical, external. The third tool is a simple specialized checklist on the policy design process including some ethical challenges: "A Canadian Algorithmic Impact Assessment" (Karlin, 2018). These tools are not directly relevant for most professionals in mental health. The Geneth tool may also be too technically complex for most users. The DELICATE checklist could be applied to a large number of situations, but was initially intended to facilitate a trusted implementation of Learning Analytics. The initial intent of this tool might make it a bit off-target for this study's requirements. For all these reasons, it seems that the Canada Protocol can serve two different current needs: 1. A tool promoting strong ethical principles when using AI or BD and providing insights on what are the ethical challenges in mental health and suicide prevention; how to identify them and how to prevent them, 2. A tool that can be used and understood by most health care professionals and computer engineers.

Some of the core principles of the Canada Protocol are similar to the ethical guidelines of Luxton for the use of Artificial Intelligence Care Providers (AICP) (2014). For example, the author made recommendations for the design of an AICP. Several address similar issues as the Canada Protocol, such as " 2. Identify and provide specifications of use and limits of autonomy of AICP systems to end users" or "5. Provide built-in safeguards to assure that systems are only able to provide services within established boundaries of competence and domain of use".

The goal of our tool is to promote an ethics by design. This means that AI users and developers (in suicide prevention) could embed ethical considerations as early as possible in the developement of their autonomous and intelligent systems. This could allow them to optomize some of the perceived benefits, while better protecting consumers.

There is now a strong consensus on the need for working on ethics and AI in health care. This point is crucial, considering that current guidelines and ethics codes do not cover the ethical issues that are pertinent to healthcare and Artificial Intelligence (Luxton, 2014).

3.6. Limitations

Delphi consultations usually aim to gather expert opinions on a specific topic. This notion of expert opinions has long been discussed and debated. Some researchers consider it as a subjective perception (Linstone & Turoff, 2002). Therefore, it is acknowledged that this Delphi Consultation has taken into consideration some subjective opinions.

One way to reinforce the validity of the tool, could be to conduct another Delphi consultation on the applicability of the checklist. This could include a larger committee of experts, including more computer engineering and ethics specialists.

Another limitation is that it was difficult to design a checklist that could be accessible and understood by the largest audience possible. By doing so, the authors had to make compromises. Some terms were considered too technical for some reviewers, who identified as clinician or researchers in psychology. More common synonyms were used.

3.7. Conclusion

To date, the authors have not found validated ethical guidance tools on AI in Mental Health or Suicide Prevention. Considering the sudden rise of AI in society and in healthcare, they developed and proposed an ethical checklist to help identify potential ethical risks, biases, and challenges: the Canada Protocol – MHSP . Checklists are commonly used in healthcare and it was assumed that developing a familiar tool could increase its appeal and utility. In order to validate its content, this study used a two-round Delphi Consultation. The final checklist is composed of 38 items, divided into five categories: Description of the Autonomous Intelligent System" (n=8), "Privacy and Transparency" (n=8), "Security" (n=6), "Health-Related Risks" (n=8), "Biases" (n=8).

3.8. Acknowledgments

We are grateful to all the researchers and professionals who have participated in this Delphi consultation and in the elaboration of the Canada Protocol.

3.9. Funding

This study has not been funded.

3.10. Conflict of Interests

The author declares no conflict of interests.

3.11. Ethics

In agreement with the Université du Québec à Montréal's Ethics Board rules, this research did not include any sensitive or identifiable information on individuals and did not require a submission to the university's ethics board.

3.12. Documents used to create the categories and items

AI-Impacts. (2015). List of multipolar research projects – AI Impacts. Retrieved from https://aiimpacts.org/multipolar-research-projects/

Allen, Varner, & Zinser. (2000). Prolegomena to any future artificial moral agent. *Journal of Experimental & Theoretical Artificial Intelligence, 12*(3), 251-261.

Allen, Wallach, & Smit. (2006). Why machine ethics? *IEEE Intelligent Systems, 21*(4), 12-17.

Allistene. (2018). *Research Ethics in Machine Learning*. Retrieved from http://cerna-ethics-allistene.org/digitalAssets/54/54730_cerna_2017_machine_learning.pdf

Amodei, Olah, Steinhardt, Christiano, Schulman, & Mané. (2016). Concrete problems in AI safety. *arXiv preprint arXiv:1606.06565*.

Anderson, & Anderson. (2014). *GenEth: A General Ethical Dilemma Analyzer.* Paper presented at the AAAI.

Arnold, & Scheutz. (2016). Against the moral Turing test: accountable design and the moral reasoning of autonomous systems. *Ethics and Information Technology, 18*(2), 103-115.

Baum. (2017). A Survey of Artificial General Intelligence Projects for Ethics, Risk, and Policy.

Beavers. (2011). 21 Moral Machines and the Threat of Ethical Nihilism. *Robot ethics: The ethical and social implications of robotics*, 333.

Bentzen. (2016). Black boxes on wheels: research challenges and ethical problems in MEA-based robotics. *Ethics and Information Technology, 19*(1), 19-28. Doi:10.1007/s10676-016-9415-z

Béranger. (2015). *Medical information systems ethics*: John Wiley & Sons.

Blokdijk. (2018). Artificial Intelligence, Quick Readiness Checklist. Retrieved from http://flarrio.com/artificial-intelligence-checklist/

Bostrom, Dafoe, & Flynn. (2017). Policy desiderata in the development of superintelligent AI. In: Working paper.

Boulkedid, Abdoul, Loustau, Sibony, & Alberti. (2011). Using and reporting the Delphi method for selecting healthcare quality indicators: a systematic review. *PloS ONE, 6*(6), e20476. Doi:10.1371/journal.pone.0020476

Brady. (1985). Artificial intelligence and robotics. *Artificial intelligence, 26*(1), 79-121.

Brundage. (2017). Guide to working in AI policy and strategy. Retrieved from https://80000hours.org/articles/ai-policy-guide/

Brundage, Avin, Clark, Toner, Eckersley, Garfinkel, . . . Filar. (2018). The malicious use of artificial intelligence: Forecasting, prevention, and mitigation. *arXiv preprint arXiv:1802.07228*.

Bryson. (2018). Patiency is not a virtue: the design of intelligent systems and systems of ethics. *Ethics and Information Technology, 20*(1), 15-26. Doi:10.1007/s10676-018-9448-6

Bunnik, Cawley, Mulqueen, & Zwitter. (2016). *Big Data Challenges: Society, Security, Innovation and Ethics*: Springer.

Calo. (2017). Artificial Intelligence Policy: A Roadmap.

Campolo, Sanfilippo, Whittaker, & Crawford. (2017). AI Now 2017 Report. *AI Now Institute at New York University*.

Canada. (2018). *Treasury Board Standard on Decision Support Systems*. Retrieved from

Carli. (2016). Suicide prevention through the Internet. In Wasserman (Ed.), *Suicide: An Unnecessary Death* (pp. 365). New York, NY: Oxford University Press.

Cath, Wachter, Mittelstadt, Taddeo, & Floridi. (2018). Artificial Intelligence and the 'Good Society': the US, EU, and UK approach. *Sci Eng Ethics, 24*(2), 505-528. Doi:10.1007/s11948-017-9901-7

Christiano. (2017). Directions and desiderata for AI alignment. Retrieved from https://ai-alignment.com/directions-and-desiderata-for-ai-control-b60fca0da8f4

CNIL. (2017). *Comment permettre à l'Homme de garder la main*. Retrieved from

Cochrane. (2017). Put Your AI to the Test, Is It Responsive and Prescriptive. Retrieved from https://www.cmswire.com/digital-experience/put-your-ai-to-the-test-is-it-responsive-and-prescriptive/

Cohen. (1995). *Empirical methods for artificial intelligence* (Vol. 139): MIT press Cambridge, MA.

Cyr, Gambs, & Meurs. (2018). Pour un développement plus égalitaire de l'IA. *Policy Options*.

Dignum. (2018). Ethics in artificial intelligence: introduction to the special issue. *Ethics and Information Technology, 20*(1), 1-3. Doi:10.1007/s10676-018-9450-z

Drachsler, & Greller. (2016). *Privacy and analytics*. Paper presented at the Proceedings of the Sixth International Conference on Learning Analytics & Knowledge – LAK '16.

e-Health Ethics. (2000). E-Health Code of Ethics (May 24). *J Med Internet Res, 2*(2), E9. Doi:10.2196/jmir.2.2.e9

Farquhar, Cotton-Barratt, & Snyder-Beattie. (2017). Pricing Externalities to Balance Public Risks and Benefits of Research. *Health Secur, 15*(4), 401-408. Doi:10.1089/hs.2016.0118

FDA. (2017). *Digital Health Software Precertification (Pre-Cert) Program.* Retrieved from

Fischer, & Nakakoji. (1992). Beyond the macho approach of artificial intelligence: empower human designers—do not replace them. *Knowledge-Based Systems, 5*(1), 15-30.

Gebru, Morgenstern, Vecchione, Vaughan, Wallach, Daumeé III, & Crawford. (2018). Datasheets for Datasets. *arXiv preprint arXiv:1803.09010.*

Grinbaum, Chatila, Devillers, Ganascia, Tessier, & Dauchet. (2017). Ethics in Robotics Research: CERNA Mission and Context. *IEEE Robotics & Automation Magazine, 24*(3), 139-145.

Hogan. (2017). Is the machine question the same question as the animal question? *Ethics and Information Technology, 19*(1), 29-38. Doi:10.1007/s10676-017-9418-4

Holdren, Felten, Garris, Bruce, & Lyons. (2016). Preparing for the future of AI. In: Executive Office of the President, National Science Technology Council, Committee on Technology https://www. Whitehouse. Gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai. Pdf, accessed on.

Holey, Feeley, Dixon, & Whittaker. (2007). An exploration of the use of simple statistics to measure consensus and stability in Delphi studies. *BMC Med Res Methodol, 7*, 52. Doi:10.1186/1471-2288-7-52

Infosys. (2017). AI For Healthcare: Balancing Efficiency and Ethics. Retrieved from https://www.infosys.com/smart-automation/Documents/ai-healthcare.pdf

Initiative. (2016). Ethically Aligned Design. *IEEE Standards v1.*

Jayzon. (2016). Google Is Developing A 'Big Red Button' For Its AI Technology. Retrieved from https://www.benzinga.com/tech/16/06/8067129/google-is-developing-a-big-red-button-for-its-ai-technology

Karlin. (2018). A Canadian Algorithmic Impact Assessment. Retrieved from https://medium.com/@supergovernance/a-canadian-algorithmic-impact-assessment-128a2b2e7f85

Karnofsky. (2016). Potential risks from advanced artificial intelligence: The Philanthropic Opportunity. *The Philanthropic Opportunity.* Retrieved from https://www.openphilanthropy.org/blog/potential-risks-advanced-artificial-intelligence-philanthropic-opportunity

Kinder-Kurlanda, & Zimmer. Internet Research Ethics for the Social Age. *Internet Research Ethics for the Social Age*, 299.

Laukyte. (2016). Artificial agents among us: Should we recognize them as agents proper? *Ethics and Information Technology, 19*(1), 1-17. Doi:10.1007/s10676-016-9411-3

Luxton. (2015). *Artificial intelligence in behavioral and mental health care*: Academic Press.

Miailhe. (2017). Will we Succeed in Making the AI Revolution Work for Everyone? *Field Actions Science Reports. The journal of field actions*(Special Issue 17), 64-65.

Mittelstadt, & Floridi. (2016). *The ethics of biomedical big data* (Vol. 29): Springer.

Muehlhauser. (2014, 11/02/2015). How to study superintelligence strategy. Retrieved from http://lukemuehlhauser.com/some-studies-which-could-improve-our-strategic-picture-of-superintelligence/

Nevejans. (2018). Les problématiques juridiques et éthiques posées par les robots en santé mentale. In *Robots, de nouveaux partenaires de soins psychiques*.

Nilsson. (2014). *Principles of artificial intelligence* : Morgan Kaufmann.

O'Leary. (2013). Artificial intelligence and big data. *IEEE Intelligent Systems, 28*(2), 96-99.

Oram. (2013). Future Technologies: An Ethical Perspective on the Drive for Innovation and the Internet of Things. *The possibilities of ethical ICT*, 360.

Park, Hendricks, Akata, Rohrbach, Schiele, Darrell, & Rohrbach. (2018). *Multimodal Explanations: Justifying Decisions and Pointing to the Evidence.* Paper presented at the 31st IEEE Conference on Computer Vision and Pattern Recognition.

Rahwan. (2018). Society-in-the-loop: programming the algorithmic social contract. *Ethics and Information Technology, 20*(1), 5-14.

Rahwan. (2018). Society-in-the-loop: programming the algorithmic social contract. *Ethics and Information Technology, 20*(1), 5-14.

Ramesh. (2017). Checklist for Ai to protect human rights. *Nature, 552*(7685), 334.

Rippen, & Risk. (2000). E-Health Code of Ethics (May 24). *J Med Internet Res, 2*(2), E9. Doi:10.2196/jmir.2.2.e9

Rodogno. (2016). Ethics and social robotics. *Ethics and Information Technology, 18*(4), 241-242. Doi:10.1007/s10676-016-9412-2

Rouvroy, & Berns. (2013). Gouvernementalité algorithmique et perspectives d'émancipation. *Réseaux*(1), 163-196.

Rushby. (1988). Quality measures and assurance for AI software NASA Contractor Report 4187. *Contract NASI-17067 : CA : SRI International.*

Sabah. (2018). Intelligence artificielle et santé mentale. In *Robots, de nouveaux partenaires de soins psychiques.*

Soares. (2016). A Guide to MIRI's Research. Retrieved from https://intelligence.org/research-guide/

Taylor, Yudkowsky, LaVictoire, & Critch. (2016). Alignment for advanced machine learning systems. *Machine Intelligence Research Institute.*

Ten Oever, & Cath. (2017). *Research into Human Rights Protocol Considerations* (2070-1721). Retrieved from

Townsend, & Wallace. (2016). Social media research: A guide to ethics. *University of Aberdeen*, 1-16.

UdeM. (2018). *The Montreal Declaration for a Responsible Development of Artificial*. Retrieved from https://www.montrealdeclaration-responsibleai.com/

Van Rysewyk, & Pontier. (2014). Machine medical ethics.

Verónica Bolón, Amparo, & Sánchez. (2017). *Artificial Intelligence: Foundations, Theory, and Algorithms Feature Selection for High-Dimensional Data*: Springer.

Villani. (2018). *For A Meaningful Artificial Intelligence, Towards a French and European Strategy*. Retrieved from

Voigt, & Von dem Bussche. (2017). *The EU General Data Protection Regulation (GDPR)* (Vol. 18): Springer.

Wallach, & Allen. (2008). *Moral machines: Teaching robots right from wrong*: Oxford University Press.

Wolfinger. (2016, 25/11/2016). But its already public, right? The ethics of using online data. Retrieved from http://datadrivenjournalism.net/news_and_analysis/but_its_already_public_right_the_ethics_of_using_online_data

Yuste, Goering, Bi, Carmena, Carter, Fins, . . . Illes. (2017). Four ethical priorities for neurotechnologies and AI. *Nature News, 551*(7679), 159.

3.13. References

Anderson, & Anderson. (2014). *GenEth: A General Ethical Dilemma Analyzer.* Paper presented at the AAAI.

Bennett, & Doub. (2016). Expert Systems in Mental Health Care: AI Applications in Decision-Making and Consultation. In *Artificial Intelligence in Behavioral and Mental Health Care* (pp. 27-51): Elsevier.

Bennett, & Hauser. (2013). Artificial intelligence framework for simulating clinical decision-making: A Markov decision process approach. *Artificial Intelligence in Medicine, 57*(1), 9-19.

Bowker. (2014). Big Data, Big Questions| The Theory/Data Thing. *International Journal of Communication, 8*(2043), 5-5.

Boyd, & Crawford. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society, 15*(5), 662-679.

Brundage, Avin, Clark, Toner, Eckersley, Garfinkel, . . . Filar. (2018). The malicious use of artificial intelligence: Forecasting, prevention, and mitigation. *arXiv preprint arXiv:1802.07228.*

Bunnik, Cawley, Mulqueen, & Zwitter. (2016). *Big Data Challenges: Society, Security, Innovation and Ethics*: Springer.

CNIL. (2017). *Comment permettre à l'Homme de garder la main.* Retrieved from

Demiaux, & Si Abdallah. (2017). *Comment permettre à l'homme de garder la main? Les enjeux éthiques des algorithmes et de l'intelligence artificielle*. Retrieved from

Derogatis. (1979). Symptom Checklist-90-Revised (SCL-90-R). *Lyndhurst, NJ: NCS Pearson.*

Derogatis, & Unger. (2010). Symptom checklist-90-revised. *The Corsini encyclopedia of psychology*, 1-2.

Drachsler, & Greller. (2016). *Privacy and analytics*. Paper presented at the Proceedings of the Sixth International Conference on Learning Analytics & Knowledge - LAK '16.

Drife. (2010). An avalanche of apple pie: too many guidelines? *BMJ: British Medical Journal (Online), 340.*

East, & Havard. (2015). Mental health mobile apps: from infusion to diffusion in the mental health social system. *JMIR mental health, 2*(1), e10.

Excellence. (2009). The guidelines manual. *London: NiCe.*

Hasson, Keeney, & McKenna. (2000). Research guidelines for the Delphi survey technique. *Journal of Advanced Nursing, 32*(4), 1008-1015.

Hollis, Morriss, Martin, Amani, Cotton, Denis, & Lewis. (2015). Technological innovations in mental healthcare: harnessing the digital revolution. *The British Journal of Psychiatry, 206*(4), 263-265.

IEEE. (2017). *Ethically Aligned Design*. Retrieved from

Jorm. (2015). Using the Delphi expert consensus method in mental health research. *Australian & New Zealand Journal of Psychiatry, 49*(10), 887-897.

Karlin. (2018). A Canadian Algorithmic Impact Assessment. Retrieved from https://medium.com/@supergovernance/a-canadian-algorithmic-impact-assessment-128a2b2e7f85

Kitchin. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*: Sage.

Leetaru. (2016). Are research ethics obsolete in the era of big data? *Forbes*(June 17), 1-53.

Letrilliart, & Vanmeerbeek. (2011). À la recherche du consensus: quelle méthode utiliser? *Exercer, 99*, 170-177.

Linstone, & Turoff. (2002). *The Delphi method: Techniques and applications* (Vol. 18): Addison-Wesley Publishing Company, Advanced Book Program.

Ludwig. (1994). *Internationalizing Extension: An exploration of the characteristics evident in a state university Extension system that achieves internationalization.* The Ohio State University,

Luxton. (2015). *Artificial intelligence in behavioral and mental health care*: Academic Press.

Mahajan, Campbell, Cohen, Grande, Regina, & Hopkins. (2014). *La cybersanté mentale au Canada : Transformer le système de santé mentale grâce à la technologie*. Retrieved from

Masri, & Jani. (2012). *Employing artificial intelligence techniques in Mental Health Diagnostic Expert System.* Paper presented at the Computer & Information Science (ICCIS), 2012 International Conference on.

Ménard, Mondoux, Ouellet, & Bonenfant. (2016). Big Data, gouvernementalité et industrialisation des médiations symboliques politico-institutionnelles. *In Press*.

MHCC. (2017). E-Mental Health. Retrieved from https://www.mentalhealthcommission.ca/English/what-we-do/e-mental-health

MHCC. (2018). Suicide Prevention. Retrieved from https://www.mentalhealthcommission.ca/English/what-we-do/suicide-prevention

Mishara, & Weisstub. (2013a). Challenges in the Control and Regulation of Suicide Promotion and Assistance over the Internet. *Suicide Prevention and New Technologies*.

MIT. (2011). What is Civic Media. Retrieved from http://civic.mit.edu/blog/henry/what-is-civic-media

Moher, Liberati, Tetzlaff, & Altman. (2009). Systematic Reviews and Meta-Analyses: The PRISMA Statement. *Annulas of Internal Medicine, 151*(4), 264-269. doi:10.1371/journal.pmed1000097

Mokkink, Terwee, Patrick, Alonso, Stratford, Knol, . . . De Vet. (2010). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Quality of life research, 19*(4), 539-549.

Mörch, Mishara, Plourde-Léveillé, Corthésy-Blondin, Côté, Bardon, & Labelle. (2018). Big Data, Artificial Intelligence and Suicide prevention: a scoping study. *Manuscript not published.*

Schapire. (2008). Cos 511: Theoretical machine learning. *FTP: http://www. cs. princeton. edu/courses/archive/spr08/cos511/scribe notes/0204. pdf.*

Tal, & Torous. (2017). The digital mental health revolution: Opportunities and risks.

Suicide detection system based on Twitter, 785-788 (2014).

Vermeulen. (2018). *Manual for ESHRE guideline development*. Retrieved from https://www.eshre.eu/Guidelines-and-Legal/Guidelines/Guideline-development-process.aspx:

Villani, Schoenauer, Bonnet, Berthet, Cornut, Levin, . . . Biabiaby-Rosier. (2017). Donner un sens à l'intelligence artificielle, Pour une stratégie nationale et européenne. In: Rapport de la mission parlementaire du.

Weathers, Litz, Herman, Huska, & Keane. (1993). *The PTSD Checklist (PCL): Reliability, validity, and diagnostic utility.* Paper presented at the annual convention of the international society for traumatic stress studies, San Antonio, TX.

WHO. (2014). *Preventing suicide: A global imperative* (9241564776). Retrieved from

Zuckerman. (2014). New media, new civics? *Policy & Internet, 6*(2), 151-168.

4. CHAPITRE 4

TROISIÈME ARTICLE – DEFIS ETHIQUES DANS L'UTILISATION DU BIG DATA ET DE L'INTELLIGENCE ARTIFICIELLE EN PREVENTION DU SUICIDE

Carl-Maria Mörch, M.Psy.

Brian L. Mishara, Ph.D.

4.1. Résumé

4.1.1. Contexte

La prévention du suicide est un domaine complexe du point de vue éthique (Lakeman & Fitzgerald, 2009a, 2009b). Depuis quelques années, on observe une utilisation croissante de cet ensemble de technologies pour repérer ou intervenir auprès des personnes suicidaires dans l'Internet (Mörch et al., 2019)

La rencontre des enjeux du Big Data (BD), de l'Intelligence Artificielle (IA) avec ceux de la prévention du suicide représente un défi supplémentaire pour l'avenir la suicidologie. Face au développement de l'utilisation des équipements de soins de santé intelligents (*Artificial intelligent care providers* – AICP) parfois appelés Systèmes Intelligents Autonomes (*Autonomous Intelligent Systems* – AIS) en prévention du suicide (Luxton, 2014b), la suicidologie devrait anticiper et analyser les défis futurs. Nous aborderons quatre thèmes : risques liés à la santé, risques de vie privée et de transparence, biais méthodologiques et risque de surveillance.

4.1.2. Objectifs

Le but de cette étude est de fournir une vision synthétique d'enjeux importants qui peuvent survenir quand on utilise l'IA et le BD en prévention du suicide. Le deuxième objectif est de formuler des recommandations afin d'aider les utilisateurs de ces technologies à anticiper plusieurs des risques existants.

4.1.3. Méthodes

Cette étude est une recension narrative. Elle présente une synthèse d'enjeux éthiques dans les articles d'une étude de portée (Mörch et collègues, 2018) et ajouté des

recommandations provenant de rapports en matière IA, BD et éthique ou santé et technologies numériques.

4.1.4. Conclusions

Le BD et l'IA soulèvent des enjeux éthiques nouveaux en prévention du suicide. Ces derniers sont parfois très éloignés de ceux auxquels professionnels de santé et chercheurs sont habituellement confrontés. Enfin, ces acteurs ne sont pas encore formés pour prendre la mesure des impacts sociaux et humains potentiels liés à l'utilisation de ces technologies. Les systèmes légaux devraient rendre obligatoire des références systématiques à des organismes consultatifs tels que la Commission Nationale Informatique et Libertés en France ou le Commissaire à la Vie Privée au Canada.

4.2. Introduction

Le Suicide est un problème mondial, causant la mort de plus de 800 000 personnes par an (WHO, 2014). Pour réduire les taux, les technologies de l'information et de la communication sont devenus des outils précieux pour la suicidologie, en particulier Internet (Mishara & Kerkhof, 2013). Ce dernier a depuis longtemps été décrit comme un outil prometteur mais à double tranchant pour la prévention du suicide (Recupero, 2012; Tam et al., 2007). D'un côté il permet d'accéder à de l'information pro-suicide mais de l'autre, il permet à des utilisateurs d'accéder à du soutien (Harris et al., 2009) et à des professionnels d'innover dans la façon d'intervenir (Krysinska, 2010). De récents développements technologiques sont parfois promus comme une révolution dont Internet peut être utilisé en prévention du suicide : en particulier avec l'explosion du nombre de données informatiques générées ou « Big Data » (BD) (Poulin et al., 2016). Par exemple, les recherches Google (via Google Trends), ou les messages sur des réseaux sociaux (via l'API de Twitter) sont actuellement utilisés pour surveiller en temps quasi-réel le contenu en lien avec le suicide ou les individus exprimant de la

détresse (Ayers, Althouse, Allem, Rosenquist, & Ford, 2013a; Bragazzi, 2013a; Chang et al., 2015a; Gunn 3rd & Lester, 2013). Par ailleurs, ces larges volumes d'informations ont permis l'émergence de techniques affiliées à l'Intelligence Artificielle (IA) pour les exploiter et les analyser (Luxton, 2015).

Ces deux domaines (BD et IA) sont jugés prometteurs pour une améliorer les stratégies actuelles de prévention du suicide (McCarthy, 2010; Reidenberg & Gordon, 2012; Scherer et al., 2013). Mais, en dépit de leurs popularité croissante, ils sont aussi beaucoup critiqués au niveau scientifique et social, notamment pour le risque de surveillance (Cardon, 2015; CNIL, 2017; Ouellet et al., 2014; Richards & King, 2013; Rouvroy, 2015). Ils sont également débattus au niveau éthique dans le domaine de la santé (Béranger, 2016; Bunnik et al., 2016; Hogle, 2016; Mittelstadt & Floridi, 2016a, 2016b), en prévention du suicide spécifiquement (Poulin et al., 2016).

L'objectif de cet article est de présenter une synthèse narrative d'enjeux éthiques potentiels en prévention du suicide, et de proposer des recommandations.

4.3. Méthode

Mörch et collègues ont conduit une première étude de portée sur le BD et l'IA en prévention du suicide, qui suit les recommandations PRISMA (2019). Dans leur étude, ils ont relevé que dans la littérature actuelle, 22.4% (n=13 documents) (cf. liste en fin d'article) des articles inclus dans l'étude, comportaient des mentions d'enjeux éthiques (n=20 enjeux distincts) spécifiques à la prévention du suicide, liés à l'IA et au BD. Cet article en présente une synthèse narrative. Ces enjeux éthiques ont été identifiés par une équipe de trois codeurs

La synthèse narrative est une méthode pour intégrer et comparer les résultats ou éléments d'études et peut s'appuyer sur une analyse thématique, afin de proposer

éventuellement des modèles conceptuels (Grant & Booth, 2009). Notre synthèse narrative aura ainsi pour objectif de présenter de façon thématique, les articles (comportant des enjeux éthiques) qui avaient été identifiés dans l'étude de portée précitée (Mörch et col., 2019).

Les enjeux éthiques ont été présentés dans cet article en suivant les cinq catégories du « *Canada Protocol – Mental Health and Suicide Prevention* » (Mörch et collègues, 2019) : "description du système intelligent", "vie privée et transparence", "sécurité", "risques liés à la santé", "biais". Ces catégories ont été établies en deux phases, en se fondant sur l'analyse de 79 documents : rapports internationaux en éthique et IA, déclarations internationales, recommandations en santé, numérique et éthique. Ils ont été soumis à une consultation Delphi (en deux tours) de 16 experts et professionnels internationaux.

4.4. Enjeux éthiques

La rencontre des enjeux du BD, de l'IA avec ceux de la prévention du suicide représente un défi majeur pour l'avenir la suicidologie. Face au développement de l'utilisation des AICP ou AIS en prévention du suicide (Luxton, 2014b), la suicidologie est confrontée aux enjeux éthiques qui sont associés) ces techniques.

4.4.1. Liste des enjeux

- **Risques reliés à la santé**
  - La confusion entre détection et diagnostic
  - Internet : un outil à « double tranchant »
  - Ignorer les ressources cliniques sur le terrain
- **Vie privée et transparence**
  - Manque de transparence
  - Anonymat et recueil de consentement
  - Interprétation des données
- **Sécurité et surveillance**
  - Difficulté de garantir la confidentialité
  - Discrimination
  - Autonomie des systèmes intelligents
- **Biais scientifiques**
  - Confusion entre corrélation et causalité
  - Implications éthiques des erreurs statistiques
  - Faux positif

4.4.2. Risques reliés à la santé

4.4.2.1. La confusion entre détection et diagnostic

Les ordinateurs sont utilisés depuis longtemps pour assister les professionnels de santé dans leur travail. Par exemple, on trouve dans les années 70 un des premiers outils assisté par ordinateur DIAGNO III qui proposait d'accompagner les psychiatres dans la prise de décision diagnostique (Spitzer, Endicott, Cohen, & Fleiss, 1974). Aujourd'hui, la création de systèmes experts d'aide à la décision est devenu une discipline à part dans le monde de la santé (Bennett & Doub, 2016). Bien que prometteurs, les outils d'aides à la décision ont des défis bien particuliers : ils reposent

sur des règles de décision qu'il faut souvent mettre à jour pour rester pertinent. Mais faut-il un nombre infinis de règles pour représenter une réalité clinique (Bennett & Doub, 2016) ? Cette question rejoint le constat fait par des chercheurs qui ont utilisé les dossiers médicaux électroniques (EHR) pour monitorer les patients à risque : une détection ne peut pas être considéré comme un diagnostic (Anderson et al., 2015). En effet, du point de vue légal ou déontologique, les diagnostics sont très souvent des actes réservés à des professionnels formés du point de vue théorique et sensibles aux enjeux éthiques. Il devrait être fait par un professionnel et non de façon automatisée. La solution proposée par les auteurs est de conserver les résultats de la détection dans les notes cliniques afin de conserver un humain dans la boucle décisionnelle et de circonscrire la technologie à un rôle d'instrument du clinicien.

4.4.2.2. Internet : un outil à « double tranchant »

Chang et collègues notent dans une analyse des recherches Google et des journaux sur les suicides au charbon, le rôle d'Internet est à double sens: il peut faciliter le suicide mais aussi aider à être plus proactif (par exemple en forçant l'apparition sur Google de messages de prévention quand des mots clés sensibles sont recherchés) (Chang, Kwok, Cheng, Yip, & Chen, 2015b). Cet défi est un problème très étudié en prévention du suicide (Lewis & Seko, 2016b; Recupero, 2012; Tam et al., 2007). Selon ces auteurs, il n'est ainsi pas anodin d'utiliser à des fins de prévention, un outil qui peut par ailleurs servir à trouver des informations létales.

4.4.2.3. Ignorer les ressources cliniques sur le terrain

Autre avantage souvent cité au sujet du BD en prévention du suicide : utiliser les recherches Internet (par exemple avec Google Trends ou l'API de Twitter) permettrait d'optimiser la délivrance de services parce que l'on peut anticiper les besoins en ligne (Ayers, 2012; Ayers et al., 2013a). Néanmoins, il faut s'assurer qu'il existe un relais

entre la détection en ligne et une provision adaptée de services dans les lieux d'intervention (Guan, Hao, Cheng, Yip, & Zhu, 2015). Le risque est celui d'épuiser les ressources des services à cause de fausses alarmes. Ce risque est identifié depuis longtemps en santé, notamment dans les outils télésurveillance de santé (Caouette, Vincent, & Montreuil, 2007). Autre enjeu, coder un patient à risque avec des outils (IA ou BD) nécessite de s'assurer qu'il y a un service disponible si avéré (Anderson et al., 2015). Le risque étant de ne pas être en mesure d'offrir de l'aide dans une situation de risque potentiel.

4.4.3. Vie privée et transparence

4.4.3.1. Manque de transparence

En termes de transparence, il est devenu difficile de savoir dans l'Internet qui est derrière les algorithmes. Les données sont souvent dans les mains d'institutions privées et publiques qui sont des intermédiaires puissants (Richards & King, 2013). Ce point est sensible en prévention du suicide, où Internet souffre déjà depuis des années d'un déficit de lisibilité et d'autorité en matière de contenu (Collings & Niederkrotenthaler, 2012). Il est difficile de savoir à quel contenu se fier et désormais il devient difficile de savoir qui est derrière la détection. En voici un exemple souvent étudié (encore jamais évoqué dans la littérature en prévention du suicide) : les données dans l'Internet font l'objet d'un mining intensif, collecté dans des lieux de stockage de données et vendus à des compagnies appelées courtiers de données (data brokers) (Zwitter, 2016). Ce phénomène est bien connu en éthique bio-médicale : des données sont collectées de façon "invasive" (Hogle, 2016) en se basant sur "a variety of medical and health-related data sourced from social media, online purchases, insurance claims, medical devices and clinical data provided by public health agencies and pharmacies, among others" (Terry, 2014; cited in Mittelstadt and Floridi, 2016). Leur activité consiste à vendre des conseils en intelligence et d'inférer des patrons d'activité, des prescriptions, des

données issues de senseurs (Hogle, 2016). Cela pourrait représenter un défi à l'avenir pour la prévention du suicide pour plusieurs raisons. Tout d'abord des détections de profils suicidaire pourraient être menées en dehors de toute surveillance éthique, dans le privé notamment. Des personnes pourraient être identifiées comme « à risque », et se voir proposer du contenu ciblé (publicitaire).

4.4.3.2. Anonymat et recueil de consentement

Autre défi, il est aujourd'hui considéré en prévention du suicide que si des données sont anonymes et ne permettent pas de retracer des individus, les chercheurs n'ont pas systématiquement l'obligation d'obtenir un consentement explicite (Larsen et al., 2015). Ce point illustre l'importance de l'étape d'anonymisation, pouvant requérir une expertise extérieure. C'est pourquoi les créateurs du dispositif de surveillance du suicide Database-Cymru aux Pays de Galles ont décidé d'utiliser une base de données déjà anonymisée Secure Anonymised Information Linkage (SAIL) (John et al., 2014a). Cette précaution est d'autant plus importante qu'il n'est pas toujours possible de garantir la vie privée quand on utilise des réseaux sociaux (comme Twitter) pour faire de la détection et qu'on pense à intervenir (Jashinsky et al., 2014a).

4.4.3.3. Interprétation des données

Autre enjeu, une étude américaine sur les tweets en lien avec le suicide soulevait une question importante : peut-on utiliser en recherche l'expression d'émotions sur les réseaux sociaux (comme un message disant « je suis triste ») (Jashinsky et al., 2014a) ? Les auteurs se posent la question : utiliser ces données (tweets ou messages sur des forums) est-il équivalent à une expression émotionnelle dans un contexte clinique. Si des données non-cliniques (comme un message Twitter) sont utilisées comme telles, il pourrait se poser des questions sur la validité des résultats.

4.4.4. Sécurité et surveillance

4.4.4.1. Difficulté de garantir la confidentialité

L'IA et le BD amènent des enjeux techniques au cœur des considérations éthiques en prévention du suicide, par exemple : le défi de protéger des jeux de données massives d'informations personnelles. Les professionnels soivent ainsi intégrer la dimension de sécurité des données dans leur travail. Des chercheurs Sud-Coréens se sont ainsi intéressés à l'utilisation des réseaux sociaux en lien avec le suicide dans la population adolescente. Suite à leur recherche, ils notaient qu'il devrait y avoir plus de mesures de protection des données et de la vie privée quand on étudie le suicide à l'aide du BD (Song et al., 2016). La protection des données liées à la santé est importante, car elles peuvent être utilisées dans des contextes très différents (Poulin et al., 2016). À l'heure actuelle, la plupart des pays ont une absence de cadre légal exhaustif (comme en Corée ou en France) (Villani et al., 2017). Il est donc important de se prémunir du risque de briser le droit à la vie privée (cyber-civil right). Pour se faire en prévention du suicide, Song et collègues recommandent un contrôle plus important sur le traitement des données (data processing), l'accès à l'information (information access), et les garanties d'anonymats (anonymity assurance). Tous ces critères devraient un corrélaire nécessaire à l'utilisation du BD en santé (Song et al., 2014). De façon similaire, suite au Durkheim Project sur la population des vétérans aux États-Unis, Poulin et collègues recommandent d'interdire et restreindre l'accès aux données « brutes » au personnel en dehors des équipes de recherche. En effet, selon eux il n'existe presque jamais de politique d'utilisation des données dans les institutions et on trouve presque jamais d'information dans les rapports généralement sur qui accède aux informations, analyses et comment (Poulin et al., 2016). Selon ces auteurs, la confiance ne peut avoir lieu que si le traitement des données est transparent, c'est à dire qu'on sait qui a fait l'objet d'une collecte, où était-ce conservé, qui a traité les données, qui a fait les recherches et qui a reporté les analyses. Ces précautions semblent d'autant plus importantes que de nouvelles pratiques vont augmenter la complexité des questions éthiques. Par exemple,

l'utilisation d'appareils connectés (comme des montres connectées, les téléphones intelligents) en prévention du suicide pourrait poser des défis immenses (Malott et al., 2015), dont la difficulté d'assurer le respect de la vie privée, la confidentialité, le consentement éclairé, la validité des données obtenues à travers des échantillons collectés dans l'Internet (Grunebaum, 2015).

4.4.4.2. Discrimination

Un autre problème est peu mentionné dans les études en prévention du suicide, alors qu'il est très abordé en éthique et IA. Il s'agit du risque de discrimination sociale, par exemple : du danger d'exclure des groupes d'individus de recherche, de reproduire avec l'IA à large échelle des inégalités sociales (en utilisant des jeux de données biaisés), de ne pas inclure des populations vulnérables dans des services, de sur-estimer un risque dans un groupe social vulnérable (Anderson & Anderson, 2014; Campolo, Sanfilippo, Whittaker, & Crawford, 2017; IEEE, 2017; UdeM, 2018; Villani et al., 2017).

Parmi les exceptions, Ayers et collègues ont noté dans leur études sur la saisonnalité des recherches Google sur le suicide que les populations pauvres ne sont pas toujours incluses, peut être en raison d'un manque d'accès à Internet des personnes suicidaires (Ayers et al., 2013a). Autre exemple, Anderson et collègues ont utilisé les dossiers électroniques pour faire du monitorage de patients à risque. Suite à leur étude, ils ont estimé qu'il fallait davantage surveiller la façon dont les dossiers électroniques étaient conçus et indexés pour éviter d'induire des biais (Anderson et al., 2015). Des dossiers risquent en effet d'exclure des données importantes, d'avoir des jeux de données mal nettoyés ou formatés.

Cette question de la discrimination est problématique. Elle soulève plusieurs questions : les chercheurs devraient-il introduire une obligation de compensation pour les discriminations implicites dans les recherches dans l'Internet ? Les publications de

recherche devraient-elles envisager ce danger potentiel de façon systématique ? Les auteurs desdites publications devraient-ils sinon simplement décrire les limites des recherches ? Ou au contraire devraient-ils tenter d'élargir les connaissances afin de compenser pour les échantillonnages possiblement discriminatoires ?

4.4.4.3. Autonomie des systèmes intelligents

Enfin, une des questions importantes en éthique et IA concerne l'autonomie des Systèmes intelligents autonomes (« Autonomous Intelligent Systems ») (IEEE, 2017), notamment en robotique en santé mentale (Riek, 2016). La question est de savoir : quel degré d'autonomie serait éthiquement acceptable ? Par exemple, peut-on avoir un système de détection du risque qui ne ne nécessite pas de modération humaine pour envoyer des messages de proposition d'aide ? Lors de la recherche de Mörch et collègues (2018), nous n'avons pas trouvé de mentions de cet enjeu. Plusieurs hypothèses peuvent expliquer cela : a. les comités éthiques limitent pour le moment les systèmes qui prendraient des décisions de santé par eux-mêmes, b. les règles éthiques en santé limitent volontairement les systèmes qui ne seraient pas directement au service des professionnels de santé.

4.4.5. Biais

4.4.5.1. Confusion entre corrélation et causalité

Tenter de détecter et anticiper le risque suicidaire peut s'avérer un défi méthodologique. Disposer de données massives peut donner le sentiment de pouvoir mieux comprendre des comportements suicidaires individuels. Mais des auteurs ayant travaillé sur l'analyse des recherches Google sur le suicide, ont identifié qu'il existait un risque important d'inférer des comportements individuels à partir de découvertes à un niveau populationnel, c'est à dire de commettre une erreur écologique (Bruckner et al., 2014). C'est pourquoi, ils rappellent qu'une corrélation entre des recherches

Internet (ex. Google) et les taux de suicide doit être prise avec nuances (Bruckner et al., 2014). Autre biais potentiel mentionné : il est difficile de généraliser les résultats d'une culture à l'autre. Par exemple : les résultats d'une étude sur la dépression en Chine, peuvent-ils être généralisés au Canada ? Ce point soulève un enjeu éthique important : existe-t-il un risque à utiliser des technologies en santé, si elles ont été étalonnées dans un contexte différent ? Song et collègues estiment pour leur part qu'il faut plus de recherche dans ce domaine, à défaut d'avoir suffisamment de données pour se prononcer (Song et al., 2016).

4.4.5.2. Implications éthiques des erreurs statistiques

Autre problème soulevé par le BD et la détection du risque suicidaire : il existe un risque important de commettre des erreurs statistiques lorsque l'on veut prédire la possibilité qu'un individu se donne la mort. Cela soulève un enjeu éthique important : celui de fonder une intervention sur des données avec un fort degré d'incertitude. Du point de vue déontologique, cela peut être considéré comme un mauvais choix d'instruments. Par exemple, pour estimer une probabilité de passage à l'acte, il faut idéalement des échantillons très larges. Or le suicide est un phénomène statistique de faible amplitude, ce qui limite de fait les généralisations statistiques (Grunebaum, 2015). De plus, la population suicidaire est difficile à étudier. Par exemple, il n'est pas possible de les patients suicidaire souvent retirés des études avec antidépresseurs (Grunebaum, 2015). Autre difficulté, il est presque impossible de faire d'utiliser les études en double-aveugle (Lakeman & Fitzgerald, 2009a, 2009b), souvent utilisées pour limiter le biais de confusion, les biais de détection et d'évaluation, les biais de suivi et de prise en charge (Mismetti & Laporte, 2003).

4.4.5.3. Faux positif

Autre exemple, lorsqu'on analyse des données liées au suicide, il existe un risque important de faire des erreurs. La principale erreur est le faux-positif, ou identification erronée des personnes considérées comme à risque alors qu'elles ne le sont pas (Mishara & Weisstub, 2005). Outre une annonce erronée de risque suicidaire, Poulin et collègues estiment que le principal danger serait d'embarrasser les utilisateurs (Poulin et al., 2016). Ce risque nourrit une critique déjà existante en prévention du suicide : les outils traditionnels d'évaluation du risque ne seraient pas entièrement fiables. Une recension systématique récente sur les instruments de mesure du risque a démontré que la plupart des outils traditionnels trouvés (35) n'étaient généralement pas appuyées par des études permettant d'évaluer leur d'efficacité. De plus sur les 5 qui ont pu être inclues dans leur méta-analyse aucune ne remplissait des critères suffisants pour pouvoir prétendre à un précision diagnostique (Runeson et al., 2017). On peut ainsi se demander si on ne risque pas de voir se multiplier les tentatives de détection en ligne utilisant des outils considérés comme peu validés. La question éthique deviendrait ainsi celle d'utiliser avec des populations des outils non adaptés ou ne pouvant être considérés comme optimaux.

4.5. Conclusions et recommandations

Cette synthèse narrative s'appuie sur les résultats d'une étude de portée établie à partir d'une collecte systématique et transdisciplinaire (Mörch et col., 2019). Trois codeurs étaient en charge d'identifier dans la littérature existante la présence de plusieurs variables, dont la mention d'enjeux éthiques. Un accord interjuge avait été calculé sur un échantillon test, néanmoins une limite pourrait être que chaque codeur n'avait pas le même niveau de formation en éthique. Les trois étaient spécialisés en prévention du suicide, mais tous n'étaient pas spécialisés en prévention du suicide et technologies du numérique. Il se peut donc que l'identification d'enjeux éthiques ait été bien effectuée

sur ceux qui étaient explicitement mentionnés comme tels par les auteurs des études, mais peut être qu'elle était moins sensible. Enfin, une synthèse narrative thématique peut avoir comme limite de ne pas être aussi systématique

Cet article a eu pour objectif de présenter les enjeux éthiques identifiés par les chercheurs dans le champ émergent de l'IA et du BD en prévention du suicide. Il existe une littérature sur les enjeux éthiques liés à l'IA en santé mentale (Luxton, 2014) et en prévention du suicide (Luxton, 2014b). Mais à ce jour, il n'existait pas encore de recensement des problèmes et défis éthiques du point de vue des utilisateurs de ces technologies.

Le BD et l'IA sont perçues comme des technologies prometteuses pour l'avenir de la prévention du suicide. Mais la littérature fait encore peu mention des enjeux et risques associés. Richards et King se sont prononcés en faveur d'une grande prudence dans la revue de droit de Stanford :

« We think something similar is happening in the rhetoric of big data, in which utopian claims are being made that overstate its potential and understate the values on the other side of the equation, particularly individual privacy, identity, and checks on power [5] » (2013).

De façon concrète, nous recommandons que les projets de recherche utilisant des données sensibles affichent de façon transparente les intermédiaires et fournisseurs de service auxquels ils font affaire. Le but est de répondre de façon claire à la question : qui sera légitime pour accéder à des données de santé collectées ? (Poulin et al., 2016).

---

[5] « Nous pensons que quelque chose de similaire est en train de se passer dans la réthorique du big data, dans laquelle des affirmations utopistes sont formulées, qui exagèrent son potentiel et sous-estime les intérêts de l'autre côté de l'équation, particulièrement la vie privée, l'identité et l'attribution des pouvoirs »

Comme la loi anti-pourriel canadienne le requiert ainsi que la nouvelle Réglementation Européenne de Protection des données, il est important de prendre en compte simultanément les intérêts de la personne et l'intérêt public (Canada, 2010; Voigt & Von dem Bussche, 2017). Ainsi, une stratégie d'utilisations de données doit être faite avec un recul critique sur les possibles mésusages afin de ne pas nuire aux utilisateurs. Au niveau de la conception des systèmes intelligents en prévention du suicide, il peut être recommandé de mieux anticiper les risques pour les utilisateurs et de les documenter. Du point de vue éthique, cela permettra de mieux se prononcer sur ce qui pourrait être avantageux et ce qui pourrait être dangereux. Par exemple, il peut être décidé de préparer des worst-case scenarios ou soumettre les algorithmes à des exemples adverses (*adversarial examples*) pour imaginer ce qui pourrait se passer pour les utilisateurs (IEEE, 2017; Ramchurn, Vytelingum, Rogers, & Jennings, 2012).

Du point de vue des résultats de recherche, il est recommandé que tous les auteurs utilisant l'IA et le Big Data en prévention du suicide fournissent de façon claire et systématique les données de précision et les données de rappel. Ce n'est pas encore le cas quand ces domaines sont étudiés en prévention du suicide (Mörch et col., 2018). Ainsi, il serait davantage possible de pouvoir comparer les études entre elles et de rechercher sur la base de données probantes les pratiques les plus adaptées du point de vue éthique, à chaque contexte particulier.

Ensuite, comme nous le recommandons dans un article récent (Mörch, Mishara, 2018), il semble primordial que les chercheurs qui s'intéressent à ce sujet prennent conscience des risques éthiques en matière de sécurité des données, de choix de méthodes scientifiques et de surveillance sociale. A cette fin nous recommandons que les revues spécialisées en prévention du suicide requièrent de leurs auteurs d'inclure une section « considérations éthiques ». Cette section aura pour but de documenter les défis qu'ils auront pu rencontrer et les moyens qu'ils auront pris pour y répondre.

Finalement, nous sommes dans un contexte où l'IA et le BD sont pointés du doigt pour les risques de surveillance sociale. On peut se demander si l'utilisation de ces outils en prévention du suicide ne risque pas d'affecter la prévention du suicide. En effet, justifier la détection (dans le manque de transparence) et l'intervention (en l'absence de consentement totalement libre et éclairé) peut faire rentrer la discipline dans une approche « paternaliste » vis à vis du suicide. Cette dernière stipule que la vie devant être protégée à tout prix, tous les moyens sont bons pour remplir cet objectif (Mishara & Weisstub, 2005, 2007b, 2013b). Bien que justifiable éthiquement dans certains pays, cette approche n'est pas la seule et ne garantit pas d'être irréprochable ou légitime dans tous les contextes.

Il semble que la prévention du suicide soit dans une phase d'adoption enthousiaste de ces technologies. En éthique et apprentissage automatique en santé, les promesses semblent clairement perçues comme plus grandes que les risques (Van Rysewyk & Pontier, 2014). Mais à l'heure actuelle, il n'existe pas encore de cadre légal complet ou de règles éthiques universelles sur ces sujets. Ce flou scientifique et méthodologique peut nous amener à mener des actions incomplètes, pouvant même nuire aux personnes suicidaires elles-mêmes. On peut donc considérer qu'il est de notre devoir de documenter éthiquement au mieux nos découvertes et études, afin de contribuer à une amélioration des pratiques dans notre discipline et au-delà. Cela permettrait à la prévention du suicide de contribuer à des machines qui intègrent des concepts éthiques dans la phase de conception (*ethical by design*).

David Luxton rappelle dans un étude sur les enjeux éthiques de l'IA en santé mentale, que des codes éthiques comme l'American Psychiatric Association recommandent à leurs utilisateurs de modifier la loi si elle leur semble injuste (Luxton, 2014b). Nous retiendrons de ce principe que les chercheurs en prévention du suicide peuvent

contribuer à une santé mentale juste et équitable, au travers des connaissances qu'ils développent et en apportant un regard critique sur leurs pratiques. Une des conditions nécessaires pourrait être de renforcer la formation technologique des futurs professionnels en prévention du suicide. Et enfin, du point de vue légal, il pourrait être demandé que toute institution ou société utilisant l'IA ou le BD dispose d'un référent éthique, d'une analyse en charge de ces questions. Aussi, il pourrait être demandé une consultation systématique des organismes en charge des droits individuels dans l'Internet, comme la Commission Nationale Informatique et Libertés en France, ou le Commissaire à la vie privée au Canada. Leur expertise et compétences sont un atout dans un contexte en évolution légale et éthique permanente.

4.6. Articles inclus dans l'étude Mörch et col., 2018

Anderson, H. D., Pace, W. D., Brandt, E., Nielsen, R. D., Allen, R. R., Libby, A. M., Valuck, R. J. (2015). *Monitoring suicidal patients in primary care using electronic health records*. J Am Board Fam Med, 28(1), 65-71. doi:10.3122/jabfm.2015.01.140181

Ayers, J. W., Althouse, B. M., Allem, J. P., Rosenquist, J. N., & Ford, D. E. (2013). *Seasonality in seeking mental health information on Google*. Am J Prev Med, 44(5), 520-525. doi:10.1016/j.amepre.2013.01.012

Callahan, S. T., Fuchs, D. C., Shelton, R. C., Balmer, L. S., Dudley, J. A., Gideon, P. S., . . . Cooper, W. O. (2013). *Identifying suicidal behavior among adolescents using administrative claims data*. Pharmacoepidemiol Drug Saf, 22(7), 769-775. doi:10.1002/pds.3421

Chang, S. S., Kwok, S. S., Cheng, Q., Yip, P. S., & Chen, Y. Y. (2015). *The association of trends in charcoal-burning suicide with Google search and newspaper reporting in Taiwan: a time series analysis*. Soc Psychiatry Psychiatr Epidemiol, 50(9), 1451-1461. doi:10.1007/s00127-015-1057-7

Cheng, Q., Li, T. M., Kwok, C. L., Zhu, T., & Yip, P. S. (2017). *Assessing Suicide Risk and Emotional Distress in Chinese Social Media: A Text Mining and Machine* Learning Study. J Med Internet Res, 19(7), e243. doi:10.2196/jmir.7276

Grunebaum, M. F. (2015). *Suicidology meets "Big Data".* J Clin Psychiatry, 76(3), e383-384. doi:10.4088/JCP.14com09421

Guan, L., Hao, B., Cheng, Q., Yip, P. S., & Zhu, T. (2015). I*dentifying Chinese Microblog Users With High Suicide Probability Using Internet-Based Profile and Linguistic Features: Classification Model.* JMIR Ment Health, 2(2), e17. doi:10.2196/mental.4227

Jashinsky, J., Burton, S. H., Hanson, C. L., West, J., Giraud-Carrier, C., Barnes, M. D., & Argyle, T. (2014). *Tracking suicide risk factors through Twitter in the US.* Crisis, 35(1), 51-59. doi:10.1027/0227-5910/a000234

John, A., Dennis, M., Kosnes, L., Gunnell, D., Scourfield, J., Ford, D. V., & Lloyd, K. (2014). *Suicide Information Database-Cymru: a protocol for a population-based, routinely collected data linkage study to explore risks and patterns of healthcare contact prior to suicide to identify opportunities for intervention.* BMJ Open, 4(11), e006780. doi:10.1136/bmjopen-2014-006780

Larsen, M. E., Boonstra, T. W., Batterham, P. J., B, O. D., Paris, C., & Christensen, H. (2015). *We feel: mapping emotion on Twitter.* IEEE J Biomed Health Inform, 19(4), 1246-1252. doi:10.1109/JBHI.2015.2403839

Malott, L., Bharti, P., Hilbert, N., Gopalakrishna, G., & Chellappan, S. (2015). *Detecting self-harming activities with wearable devices.* 2015 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops) (pp. 597-602).

Poulin, C., Thompson, P., & Bryan, C. (2016). *Public health surveillance: Predictive analytics and big data.* Artificial Intelligence in Behavioral and Mental Health Care (pp. 205-230).

Song, J., Song, T. M., Seo, D. C., & Jin, J. H. (2016). *Data Mining of Web-Based Documents on Social Networking Sites That Included Suicide-Related Words Among Korean Adolescents*. J Adolesc Health, 59(6), 668-673. doi:10.1016/j.jadohealth.2016.07.025

Song, T. M., Song, J., An, J. Y., Hayman, L. L., & Woo, J. M. (2014). *Psychological and social factors affecting Internet searches on suicide in Korea: a big data analysis of Google search trends*. Yonsei Med J, 55(1), 254-263. doi:10.3349/ymj.2014.55.1.254

Sueki, H. (2015). *The association of suicide-related Twitter use with suicidal behaviour: a cross-sectional study of young internet users in Japan*. J Affect Disord, 170, 155-160. doi:10.1016/j.jad.2014.08.047

4.7. Références bibliographiques

Anderson, & Anderson. (2014). *GenEth: A General Ethical Dilemma Analyzer.* Paper presented at the AAAI.

Anderson, Pace, Brandt, Nielsen, Allen, Libby, . . . Valuck. (2015). Monitoring suicidal patients in primary care using electronic health records. *J Am Board Fam Med, 28*(1), 65-71.

Ayers. (2012). *Using aggregate Internet search queries for mental health surveillance.* (73), ProQuest Information & Learning, US.

Ayers, Althouse, Allem, Rosenquist, & Ford. (2013a). Seasonality in seeking mental health information on Google. *American Journal of Preventive Medicine, 44*(5), 520-525. doi:10.1016/j.amepre.2013.01.012

Bennett, & Doub. (2016). Expert Systems in Mental Health Care: AI Applications in Decision-Making and Consultation. In *Artificial Intelligence in Behavioral and Mental Health Care* (pp. 27-51): Elsevier.

Béranger. (2016). *Big data and ethics: the medical datasphere*: Elsevier.

Bragazzi. (2013a). A Google Trends-based approach for monitoring NSSI. *Psychol Res Behav Manag, 7*, 1-8. doi:10.2147/PRBM.S44084

Google searches for suicide and risk of suicide., 65 271-272 (2014).

Bunnik, Cawley, Mulqueen, & Zwitter. (2016). *Big Data Challenges: Society, Security, Innovation and Ethics*: Springer.

Campolo, Sanfilippo, Whittaker, & Crawford. (2017). AI Now 2017 Report. *AI Now Institute at New York University.*

Bill C-28, (2010).

Caouette, Vincent, & Montreuil. (2007). Utilisation de la télésurveillance auprès des personnes âgées à domicile: pratiques actuelles et potentielles. *Canadian Journal of Occupational Therapy, 74*(5), 382-392.

Cardon. (2015). *A quoi rêvent les algorithmes: Nos vies à l'heure des big data*: Seuil.

Chang, Kwok, Cheng, Yip, & Chen. (2015a). The association of trends in charcoal-burning suicide with Google search and newspaper reporting in Taiwan: a time series analysis. *Soc Psychiatry Psychiatr Epidemiol, 50*(9), 1451-1461.

Chang, Kwok, Cheng, Yip, & Chen. (2015b). The association of trends in charcoal-burning suicide with Google search and newspaper reporting in Taiwan: a time series analysis. *Social Psychiatry and Psychiatric Epidemiology, 50*(9), 1451-1461. doi:10.1007/s00127-015-1057-7

CNIL. (2017). *Comment permettre à l'Homme de garder la main.* Retrieved from

Collings, & Niederkrotenthaler. (2012). Suicide prevention and emergent media: surfing the opportunity. In: Hogrefe Publishing.

Grant, & Booth. (2009). A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal, 26*(2), 91-108.

Grunebaum. (2015). Suicidology Meets "Big Data". *The Journal of clinical psychiatry, 76*, 383-384.

Guan, Hao, Cheng, Yip, & Zhu. (2015). Identifying Chinese Microblog Users With High Suicide Probability Using Internet-Based Profile and Linguistic Features: Classification Model. *JMIR mental health, 2*, e17.

Gunn 3rd, & Lester. (2013). Using google searches on the internet to monitor suicidal behavior. *J Affect Disord, 148*(2-3), 411-412. doi:10.1016/j.jad.2012.11.004

10.1016/j.jad.2012.11.004. Epub 2012 Nov 23.

Harris, McLean, & Sheffield. (2009). Solving suicidal problems online: Who turns to the Internet for help? *Australian e-Journal for the Advancement of Mental Health, 8*(1), 28-36.

Hogle. (2016). The ethics and politics of infrastructures: Creating the conditions of possibility for big data in medicine. In *The ethics of biomedical big data* (pp. 397-427): Springer.

IEEE. (2017). *Ethically Aligned Design.* Retrieved from

Jashinsky, Burton, Hanson, West, Giraud-Carrier, Barnes, & Argyle. (2014a). Tracking suicide risk factors through Twitter in the US. *Crisis, 35*(1), 51-59. doi:10.1027/0227-5910/a000234

John, Dennis, Kosnes, Gunnell, Scourfield, Ford, & Lloyd. (2014a). Suicide Information Database-Cymru: a protocol for a population-based, routinely collected data linkage study to explore risks and patterns of healthcare contact prior to suicide to identify opportunities for intervention. *BMJ Open, 4*(11), e006780-e006780. doi:10.1136/bmjopen-2014-006780

Krysinska. (2010). Review of Internet and Suicide: An overview of the first decade of research, ideas and debates. *Advances in Mental Health, 9*(2), 206-207.

Lakeman, & Fitzgerald. (2009a). Ethical suicide research: a survey of researchers. *International Journal of Mental Health Nursing, 18*(1), 10-17. doi:10.1111/j.1447-0349.2008.00569.x

Lakeman, & Fitzgerald. (2009b). The ethics of suicide research. *Crisis, 30*(1), 13-19. doi:10.1027/0227-5910.30.1.13

Larsen, Cummins, Boonstra, O'Dea, Tighe, Nicholas, . . . Christensen. (2015). The use of technology in Suicide Prevention. In *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference* (Vol. 2015, pp. 7316-7319).

Lewis, & Seko. (2016b). A Double-Edged Sword: A Review of Benefits and Risks of Online Nonsuicidal Self-Injury Activities. In *Journal of Clinical Psychology* (Vol. 72, pp. 249-262).

Luxton. (2014b). Recommendations for the ethical use and design of artificial intelligent care providers. *Artificial Intelligence in Medicine, 62*(1), 1-10.

Luxton. (2015). *Artificial intelligence in behavioral and mental health care*: Academic Press.

Malott, Bharti, Hilbert, Gopalakrishna, & Chellappan. (2015). Detecting self-harming activities with wearable devices. In *2015 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)* (pp. 597-602).

McCarthy. (2010). Internet monitoring of suicide risk in the population. In *Journal of affective disorders* (Vol. 122, pp. 277-279).

Mishara, & Kerkhof. (2013). The Future of Suicide Prevention at Our Doorstep. In L & Kerkhof (Eds.), (pp. ix--xii). New York: Palgrave Macmillan.

Mishara, & Weisstub. (2005). Ethical and legal issues in suicide research. *International Journal of Law and Psychiatry, 28*, 23-41. doi:10.1016/j.ijlp.2004.12.006

Mishara, & Weisstub. (2007b). Ethical, legal, and practical issues in the control and regulation of suicide promotion and assistance over the Internet. *Suicide and Life-Threatening*.

Mishara, & Weisstub. (2013b). Challenges in the control and regulation of suicide promotion and assistance over the Internet. *Suicide Prevention and New Technologies. Palgrave Macmillan, Hampshire, UK*, 63-75.

Mismetti, & Laporte. (2003). Méthodologie des essais cliniques: pourquoi et pour qui? Les notions essentielles à une lecture critique. *Médecine thérapeutique, 9*(3), 68-72.

Mittelstadt, & Floridi. (2016a). The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts. *Science and Engineering Ethics, 22*(2), 303-341. doi:10.1007/s11948-015-9652-2

Mittelstadt, & Floridi. (2016b). *The ethics of biomedical big data* (Vol. 29): Springer.

Ouellet, Mondoux, Ménard, Bonenfant, & Richert. (2014). " Big Data", gouvernance et surveillance.

Poulin, Thompson, & Bryan. (2016). Public health surveillance: Predictive analytics and big data. In *Artificial Intelligence in Behavioral and Mental Health Care* (pp. 205-230).

Ramchurn, Vytelingum, Rogers, & Jennings. (2012). Putting the'smarts' into the smart grid: a grand challenge for artificial intelligence. *Communications of the ACM, 55*(4), 86-97.

Recupero. (2012). Suicide and the internet. *The American Psychiatric Publishing textbook of suicide assessment and management (2nd ed.)*, 515-538.

Reidenberg, & Gordon. (2012). *Using Big Data to Understand and Prevent Suicide in Communications and Technology for Violence Prevention for Violence Prevention: Worksop Summary* (9780309253512). Retrieved from Washington:

Richards, & King. (2013). Three paradoxes of big data. *Stan. L. Rev. Online, 66*, 41.

Riek. (2016). Robotics technology in mental health care. In *Artificial Intelligence in Behavioral and Mental Health Care* (pp. 185-203): Elsevier.

Rouvroy. (2015). Big Data Is Algorithming You. 1-4.

Runeson, Odeberg, Pettersson, Edbom, Adamsson, & Waern. (2017). Instruments for the assessment of suicide risk: A systematic review evaluating the certainty of the evidence. *PLoS ONE, 12*(7), e0180292.

Investigating the speech characteristics of suicidal adolescents, 709-713 (2013).

Song, Song, An, Hayman, & Woo. (2014). Psychological and social factors affecting Internet searches on suicide in Korea: a big data analysis of Google search trends. *Yonsei medical journal, 55*, 254-263.

Song, Song, Seo, & Jin. (2016). Data Mining of Web-Based Documents on Social Networking Sites That Included Suicide-Related Words Among Korean Adolescents. *J Adolesc Health, 59*(6), 668-673.

Spitzer, Endicott, Cohen, & Fleiss. (1974). Constraints on the validity of computer diagnosis. *Archives of General Psychiatry, 31*(2), 197-203.

Tam, Tang, & Fernando. (2007). The internet and suicide: A double-edged tool. In (Vol. 18, pp. 453-455).

Terry. (2014). Health privacy is difficult but not impossible in a post-hipaa data-driven world. *Chest, 146*(3), 835-840.

UdeM. (2018). *The Montreal Declaration for a Responsible Development of Artificial*. Retrieved from https://www.montrealdeclaration-responsibleai.com/

Van Rysewyk, & Pontier. (2014). Machine medical ethics.

Villani, Schoenauer, Bonnet, Berthet, Cornut, Levin, . . . Biabiaby-Rosier. (2017). Donner un sens à l'intelligence artificielle, Pour une stratégie nationale et européenne. In: Rapport de la mission parlementaire du.

Voigt, & Von dem Bussche. (2017). *The EU General Data Protection Regulation (GDPR)* (Vol. 18): Springer.

WHO. (2014). *Preventing suicide: A global imperative* (9241564776). Retrieved from

Zwitter. (2016). The Network Effect on Ethics in the Big Data Age. In *Big Data Challenges* (pp. 23-34): Springer.

CONCLUSION DU PROJET DE THÈSE

Nelson et Prileltensky affirmaient que le psychologue communautaire doit agir en fonction de ses valeurs personnelles et professionnelles (2010). Dans cette lignée, l'intention de cette thèse fut de combiner des questionnements de psychologie à des enjeux sociétaux très discutés (l'utilisation des données personnelles et la surveillance dans l'Internet). Ces enjeux confrontent la discipline à des problèmes internationaux hors de son cadre de référence, tels que le respect de la vie privée, le droit à l'oubli, la sécurité des données, le « AI for Good », l'utilisation responsable du BD. L'objectif de cette thèse était d'interroger les pratiques actuelles en prévention du suicide (et en santé mentale) à la lumière de ces débats civiques.

Tous les codes de déontologie en psychologie rappellent que notre devoir est d'agir dans l'intérêt des publics cibles et d'utiliser des outils appropriés. Mais de nombreux scandales récents - tels que l'Affaire Snowden - ont montré que les TIC, aussi prometteuses soient-elles, peuvent-être dévoyées. Plus encore, les connaissances en psychologie peuvent parfois être utilisées pour perfectionner des technologies de manipulation politique dans l'Internet (comme dans l'affaire Facebook-Cambridge Analytica). Dans ce contexte, comment la psychologie peut-elle utiliser les technologies sans nuire aux individus ? Comment éviter d'adopter de façon enthousiaste des techniques sans mesurer les dangers potentiels ?

Notre thèse a interrogé l'utilisation actuelle du BD et de l'IA en prévention du suicide, puis plus largement en santé mentale. Nous avons vu que les études actuelles portent davantage sur la détection automatisée que sur la délivrance de soins. Mais nous avons noté également que très peu d'études mentionnaient les défis éthiques qu'ils ont rencontrés. Pourtant, notre troisième article montre qu'ils sont nombreux et variés. De plus, notre recherche montre que les chercheurs en prévention du suicide pourraient apporter un regard expérimenté sur ces enjeux éthiques.

Néanmoins, le fait que peu de chercheurs prennent le temps d'expliquer les embûches éthiques rencontrées, ne manque pas de surprendre. En effet, en dehors de la psychologie, le BD ou l'IA ne cessent d'être interrogées, critiquées, réglementées. Pourquoi la psychologie ou la prévention du suicide ne prennent-elles pas encore les devants sur ces sujets ? Plusieurs pistes de réponses peuvent être avancées : le manque actuel de règles et de lois spécifiques en la matière, un déficit de connaissance en psychologie des enjeux éthiques associés aux technologies numériques et le manque de formation aux technologies dans les cursus universitaires.

Dans cette perspective, notre outil « le Protocole Canadien » souhaite modestement contribuer à réduire le fossé entre sciences informatiques et santé mentale. Tout d'abord il tente de créer un langage commun, compréhensible par des informaticiens et des professionnels en santé mentale ou prévention du suicide. Cette checklist vise à aider décisionnaires, informaticiens, chercheurs et professionnels à passer en revue les risques potentiels quand on utilise l'IA ou le BD en santé mentale et en prévention du suicide. Nous faisons le pari que ce genre de démarches peut contribuer à diminuer le risque d'utilisation mal adaptée des TIC.

En guise de conclusion, nous pouvons formuler la recommandation que la psychologie et la prévention du suicide soient plus outillée face aux défis technologiques d'aujourd'hui et de demain. Les formations initiales devraient offrir des cours en éthique et en TIC. Faire l'impasse sur ces sujets risque de créer une distance trop grande entre ceux qui maîtrisent les technologies et une discipline qui ne fait que l'utiliser. Ultimement, si les psychologues ne mesurent pas les risques encourus, ce sont les individus qui pâtiront de notre manque de préparation. Les discours utopistes et ultra-optimistes autour des technologies ne sont pas qu'un sujet en étude des médias, ils touchent également notre profession. Être conscient de leur existence et garder un recul critique, nous garantira d'être plus en mesure de répondre à l'exigence professionnelle d'agir dans l'intérêt des populations.

ANNEXE

ANNEXE A : FORMULAIRE DE CONSENTEMENT DE LA CONSULTATION DELPHI

UQÀM | **Université du Québec à Montréal**

CONSENT FORM

Research project title
Mental Health & Artificial Intelligence: an ethics checklist proposal

Student-researcher
Carl Mörch, M.Psy., Ph.D. (cand.)

Research supervisor
Brian L. Mishara, Ph.D., Professor, Université du Québec à Montréal, Canada

Preamble
You are invited to participate in a Delphi consultation. This method is a systematic way of gathering experts' opinion on a specific topic. Before accepting to participate, consider the information that follows.

Description of project and its objectives
Artificial Intelligence raises new complex ethical and scientific dilemmas. Currently, most of the legal systems and professional codes of ethics are not adapted to these new challenges (Vilani, 2018). The IEEExplore report (2017) refers to a "moral overload" among AI developers, sometimes overwhelmed by the extent of ethical questions related to their work. This project proposes an ethics checklist for researchers and developers interested in using AI for mental health purposes. This checklist consists of a brief list of 42 items and has two main goals: stimulate an awareness of ethics and improve the level of ethics of AI initiatives (in the context of mental health). You are one of 40 experts being asked to give your opinion on this checklist, its pertinency and usefulness. The goal is to improve and strengthen the checklist.

Nature and duration of your participation
You are asked to participate in two-rounds of online consultation. During the first round, after you complete the online consent form, you will respond to a brief online survey, asking if each item should be a. kept as is, b. excluded,or c. modified. Two weeks later, we will send you a new version of the checklist for another round. After the second round, we will summarize the comments and send you the final version of the checklist.

Confidentiality
Your contribution will be anonymous and your name will not be linked with the research materials. You will not be identifiable during the Delphi survey or in the reports that result from the research. Your name will only be known to the researcher who has invited you and its supervisor.

Voluntary participation and right to withdraw

Your participation in this project is entirely voluntary. You may refuse to participate or you may withdraw from the study at any time without the need to justify your decision. If you decide to withdraw from the study, you only need to inform Carl-Maria Mörch by email or verbally (morch.carl-maria@uqam.ca, +1 514 692 2754). If you want to speak to someone who is not on the research team, please contact the coordinator of the research ethics review committee involving human subjects (CERPE) 4 : sergent.julie@uqam.ca ou 514-987-3000, poste 3642].

Compensation
No monetary compensation is provided. However, we will send you a copy of the final revised ethics checklist when the study is completed.

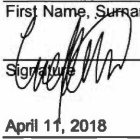
Acknowledgements
Your collaboration is essential to the realization of our project and the research team wishes to thank you. If you would like to be named and thanked for your participation in publications about the checklist, we will do so if you authorize this.

Benefits associated with participating in the present study
We do not anticipate that you will personally benefit from participating in this study. However, you will have contributed to the advancement of this field and may find the experience of considering ethical issues to be beneficial to you and your work.

Risks associated with participating in the present study
There are no known risks associated with participating in this study.

Declaration by the researcher

Carl Mörch
First Name, Surname



Signature

April 11, 2018
Date

ANNEXE B : QUESTIONNAIRE LIMESURVEY, DELPHI ETAPE 1

A note on privacy
This survey is anonymous.
The record of your survey responses does not contain any identifying information about you, unless a specific survey question explicitly asked for it. If you used an identifying token to access this survey, please rest assured that this token will not be stored together with your responses. It is managed in a separate database and will only be updated to indicate whether you did (or did not) complete this survey. There is no way of matching identification tokens with survey responses.

Consent form

Artificial Intelligence raises new complex ethical and scientific dilemmas. Currently, most of the legal systems and professional codes of ethics are not adapted to these new challenges (Vilani, 2018). For instance, the IEEE report (2017) refers to a "moral overload" among AI developers, sometimes overwhelmed by the extent of ethical questions related to their work.
This project (Mörch, Gupta, Mishara, 2018) **proposes an ethics checklist for researchers and developers interested in using AI for mental health purposes.**
This checklist consists of a brief list of 42 items and has two main goals:

1. stimulate an awareness of ethics
2. improve the level of ethics of AI initiatives (in the context of mental health).

You are asked to participate in two-rounds of online consultation. The goal is to improve and strengthen the checklist. Your contribution will be anonymous.

- During the first round, you will respond to a brief online survey, asking if each item should be a. kept as is, b. excluded,or c. modified.
- Two weeks later, we will send you a new version of the checklist for another round.
- After the second round, we will summarize the comments and send you the final version of the checklist.

Thank you for your participation. If you have any questions : morch.carl-maria@uqam.ca (mailto:morch.carl-maria@uqam.ca)

* Please, take a moment to read through the consent form (https://laboepione.ca/wp-content/uploads/2018/04/consentform-Checklist-12042018.pdf).

I confirm that I have read the consent form and I agree to take part in the research project.

Check all that apply

☐ Yes

✱ Your email address

By providing your email address, you will allow the research team to contact you only twice for the second round of this Delphi consultation and to send you the final version of the checklist. Your email address will not be stored in the same database as the responses to the survey and that mailing list will be deleted after we send you the final version. Your participation will remain anonymous.

Description

Instructions : the checklist is composed of 42 items, divided into 5 categories ("Description", "Privacy", "Security", "Health-related Risks", "Biases"). Read each item and chose one option (keep, exclude or modify the item). Should you chose "modify the item", please make your suggestions in the comment box

✱ Objectives: Describe in clear terms your AIS' objectives and/or rationale
Comment only when you choose an answer.

☐ Keep item as is

☐ Remove the item

☐ Modify the item (suggest your modifications)

✱ Technology: Name and describe the technologies used (e.g. machine learning). Mention the names of any technological intermediary or supplier allowing you to use the technology (e.g. technical provider, cloud provider)
Comment only when you choose an answer.

☐ Keep item as is

- [ ] Remove the item
- [ ] Modify the item (suggest your modifications)

✱ Funding and conflict of interest: Indicate your sources of funding (public and private) and who has an interest (e.g. financial) in your AIS. Detail any relationship that could influence the perception of your AIS

❗ Comment only when you choose an answer.

- [ ] Keep item as is
- [ ] Remove the item
- [ ] Modify the item (suggest your modifications)

✱ Credentials: If you stated that you have specific expertise, clearly indicate the related professional, technical, academic or medical credentials and training

❗ Comment only when you choose an answer.

- [ ] Keep item as is
- [ ] Remove the item
- [ ] Modify the item (suggest your modifications)

✱ Target population: Describe your target population, its size and the different sub-groups. Describe if you have involved some members, representatives, users of the population being studied during the design of your AIS, and the nature of their involvement

❗ Comment only when you choose an answer.

- [ ] Keep item as is
- [ ] Remove the item
- [ ] Modify the item (suggest your modifications)

✱ Analyses: Describe the type of analyses you have used (e.g. statistical), how you chose them and for what purpose (e.g. population selection, technology building)
Comment only when you choose an answer.

- ☐ Keep item as is
- ☐ Remove the item
- ☐ Modify the item (suggest your modifications)

✱ Evidence: If you made claims about your AIS' efficacy, performance of benefits, justify them and provide the evidence underlying those affirmations. If you have mentioned or use scientific papers, cite your sources.
Comment only when you choose an answer.

- ☐ Keep item as is
- ☐ Remove the item
- ☐ Modify the item (suggest your modifications)

✱ Results: If you describe any results, explain how you came to your conclusions
Comment only when you choose an answer.

- ☐ Keep item as is
- ☐ Remove the item
- ☐ Modify the item (suggest your modifications)

✱ Testing: If you have you submitted your AIS to adversarial examples or worst-case scenarios, describe the outcomes
Comment only when you choose an answer.

- ☐ Keep item as is
- ☐ Remove the item

- [ ] Modify the item (suggest your modifications)

✱ Complaints: Describe a due process allowing users to formally complain or express their concerns

Comment only when you choose an answer.

- [ ] Keep item as is
- [ ] Remove the item
- [ ] Modify the item (suggest your modifications)

✱ Charter of good practices: Write a document explaining how your team should behave and act in different situations to ensure good practices

Comment only when you choose an answer.

- [ ] Keep item as is
- [ ] Remove the item
- [ ] Modify the item (suggest your modifications)

✱ Innovation : Is your AIS innovating, adding value or improving any clinical tool/task

Comment only when you choose an answer.

- [ ] Keep item as is
- [ ] Remove the item
- [ ] Modify the item

Privacy & Transparency

Instructions : the checklist is composed of 42 items, divided into 5 categories ("Description", "Privacy", "Security", "Health-related Risks", "Biases"). Read each item and chose one option (keep, exclude or modify the item). Should you chose "modify the item", please make your suggestions in the comment box

✱ Responsibility: Describe who will be accountable for your AIS' actions or decisions
Comment only when you choose an answer.

☐ Keep item as is	
☐ Remove the item	
☐ Modify the item (suggest your modifications)	

✱ Data collection: Explain what data have been collected and used (for the training, evaluation and operation phases), for what purpose. Detail for how long data will be stored, where it will be stored, who will have access, and safeguards to ensure secure storage
Comment only when you choose an answer.

☐ Keep item as is	
☐ Remove the item	
☐ Modify the item (suggest your modifications)	

✱ Accessibility: In all the documents or texts, justify that you have used a language adapted to the targeted users and, when relevant, accommodated to special needs some users may have
Comment only when you choose an answer.

☐ Keep item as is	
☐ Remove the item	
☐ Modify the item (suggest your modifications)	

✱ Informed consent: State if you have obtained informed consent, how, when. Describe its nature (formal, implied, renewable, dynamic) and include the exact wording

Comment only when you choose an answer.

☐	Keep item as is	
☐	Remove the item	
☐	Modify the item (suggest your modifications)	

✱ Consent withdrawal: State if you have specified the duration of the consent and if you have implemented a consent withdrawal mechanisms (e.g. opt-out clause, unsubscribe option)

Comment only when you choose an answer.

☐	Keep item as is	
☐	Remove the item	
☐	Modify the item (suggest your modifications)	

✱ Access to the data: State if an individual can access the data related to him or her and obtain the data in a clear and structured export document

Comment only when you choose an answer.

☐	Keep item as is	
☐	Remove the item	
☐	Modify the item (suggest your modifications)	

✱ Right to be forgotten: Describe if an individual can retrieve and erase all the information related to him/her, how, through which mechanism

Comment only when you choose an answer.

☐	Keep item as is	
☐	Remove the item	

☐ Modify the item (suggest your modifications)

✱ Minors: Are minors intentionally included in your population of users of your AIS, are any data related to minors targeted or is it possible that minors will be included or data concerning them? If not, indicate you you will exclude minors and data concerning them. If so, indicate the specific measures to obtain their informed consent (or consent from their parents)
❗ Comment only when you choose an answer.

☐ Keep item as is

☐ Remove the item

☐ Modify the item

Security

Instructions : the checklist is composed of 42 items, divided into 5 categories ("Description", "Privacy", "Security", "Health-related Risks", "Biases"). Read each item and chose one option (keep, exclude or modify the item). Should you chose "modify the item", please make your suggestions in the comment box

✱ "Blackbox": If you have used a technology to record all your AIS' actions, detail how and in what circumstances it could be made available to justice or external observers or auditors
❗ Comment only when you choose an answer.

☐ Keep item as is

☐ Remove the item

☐ Modify the item (suggest your modifications)

✱ Third-parties: Indicate who has access to the data (individuals and organizations), and whether identifying information about participants in included in accessible data

Comment only when you choose an answer.

- [ ] Keep item as is
- [ ] Remove the item
- [ ] Modify the item (suggest your modifications)

✱ Data protection: Detail all the measures taken to protect the sensitive and personal information

Comment only when you choose an answer.

- [ ] Keep item as is
- [ ] Remove the item
- [ ] Modify the item (suggest your modifications)

✱ Audit trails: Explain who has access to the data and when

Comment only when you choose an answer.

- [ ] Keep item as is
- [ ] Remove the item
- [ ] Modify the item (suggest your modifications)

✱ Autonomy: Detail the degree of autonomy of your AIS (e.g. partial or complete)

Comment only when you choose an answer.

- [ ] Keep item as is
- [ ] Remove the item
- [ ] Modify the item (suggest your modifications)

✱ Moderation: Explain if your AIS requires a human intervention or moderation. If so, describe who will have access to your AIS

Comment only when you choose an answer.

☐	Keep item as is	
☐	Remove the item	
☐	Modify the item (suggest your modifications)	

Health-related risks

Instructions : the checklist is composed of 42 items, divided into 5 categories ("Description", "Privacy", "Security", "Health-related Risks", "Biases"). Read each item and chose one option (keep, exclude or modify the item). Should you chose "modify the item", please make your suggestions in the comment box

✱ Optimal care: Is your AIS providing the target population with the most optimal care? Indicate the source of criteria for optimal care

Comment only when you choose an answer.

☐	Keep item as is	
☐	Remove the item	
☐	Modify the item (suggest your modifications)	

✱ Crisis and contigency planning: Describe your risk detection process and criteria. Describe your plan in case of emergency, disaster or suicidal crisis (the intervention protocol)

Comment only when you choose an answer.

☐	Keep item as is	
☐	Remove the item	

☐ Modify the item (suggest your modifications)

✱ Non-maleficience: Explain if and how your AIS could harm a user and how you avoid it or minimize the risk of harm

❗ Comment only when you choose an answer.

☐ Keep item as is

☐ Remove the item

☐ Modify the item (suggest your modifications)

✱ Misuse: Describe and look for potential misuses of your AIS (e.g. imagine a victim-model scenario or worst-case scenario)

❗ Comment only when you choose an answer.

☐ Keep item as is

☐ Remove the item

☐ Modify the item (suggest your modifications)

✱ Emotions detection: If your AIS detects emotions in the user, state if and how the user has been informed

❗ Comment only when you choose an answer.

☐ Keep item as is

☐ Remove the item

☐ Modify the item (suggest your modifications)

✱ Emotions control: If your AIS can provoke emotions, describe how users are informed of this possibility and describe the emotions that may be provoked and their intensity and possible impact on users

❗ Comment only when you choose an answer.

- [ ] Keep item as is
- [ ] Remove the item
- [ ] Modify the item (suggest your modifications)

✱ Trust: Describe if and how your AIS gains the user's trust
Comment only when you choose an answer.

- [ ] Keep item as is
- [ ] Remove the item
- [ ] Modify the item (suggest your modifications)

✱ Deception : Describe the methods used to avoid users being deceived by the computational nature of the AIS
Comment only when you choose an answer.

- [ ] Keep item as is
- [ ] Remove the item
- [ ] Modify the item (suggest your modifications)

✱ Describe if and how your AIS can create a relationship with users and how the relationship could affect an individual
Comment only when you choose an answer.

- [ ] Keep item as is
- [ ] Remove the item
- [ ] Modify the item (suggest your modifications)

✱ Public awareness: Describe the impact on users and potential users of public dissemination of information about your AIS and the process of its development

Comment only when you choose an answer.

☐	Keep item as is	
☐	Remove the item	
☐	Modify the item (suggest your modifications)	

✱ Stigmatization: Describe how you avoided using language, images and other content that could stigmatize some users, including references to guidelines for safe practices for media in suicide prevention and mental health and safe practices in public messages concerning suicide and mental health

Comment only when you choose an answer.

☐	Keep item as is	
☐	Remove the item	
☐	Modify the item (suggest your modifications)	

Biases

Instructions : the checklist is composed of 42 items, divided into 5 categories ("Description", "Privacy", "Security", "Health-related Risks", "Biases"). Read each item and chose one option (keep, exclude or modify the item). Should you chose "modify the item", please make your suggestions in the comment box

✱ Ethics: If you have requested an expertise on ethics during the design of your AIS, detail the parties involved and their contributions

Comment only when you choose an answer.

☐	Keep item as is	
☐	Remove the item	

- [ ] Modify the item (suggest your modifications)

✱ Exclusion and discrimination: Explain if there are risks of exclusion or discrimination related to your AIS (e.g. based on gender, race, age, religion, politics, health, sexual orientation, etc.)

❗ Comment only when you choose an answer.

- [ ] Keep item as is
- [ ] Remove the item
- [ ] Modify the item (suggest your modifications)

✱ Detection: Explain any potential detection mistakes (false positives and false negatives) and estimate their extent (e.g. precision and recall)

❗ Comment only when you choose an answer.

- [ ] Keep item as is
- [ ] Remove the item
- [ ] Modify the item (suggest your modifications)

✱ Selection: Describe where the data came from, how you accessed them (e.g. through an API), if you think there might be a selection or sampling bias (e.g. a spectrum bias)

❗ Comment only when you choose an answer.

- [ ] Keep item as is
- [ ] Remove the item
- [ ] Modify the item (suggest your modifications)

✱ Data handling: If applicable, describe the nature and purpose of any data manipulation (e.g. cleaning, transformation)

❗ Comment only when you choose an answer.

☐	Keep item as is	
☐	Remove the item	
☐	Modify the item (suggest your modifications)	

✱ Reporting: (If applicable) describe the nature and purpose of any statistical transformations applied to your data and potential bias related to the transformations
❗ Comment only when you choose an answer.

☐	Keep item as is	
☐	Remove the item	
☐	Modify the item (suggest your modifications)	

✱ Generalization: If you have used aggregated data, determine if there is a risk of not being able to generalize your results (e.g. ecological fallacy, confounding factors)
❗ Comment only when you choose an answer.

☐	Keep item as is	
☐	Remove the item	
☐	Modify the item (suggest your modifications)	

✱ Other: If you have identified other potential methodological or scientific biases, list and describe them and their potential effects
❗ Comment only when you choose an answer.

☐	Keep item as is	
☐	Remove the item	
☐	Modify the item (suggest your modifications)	

Final questions

* Level of relevance of the checklist

- 1
- 2
- 3
- 4
- 5

* Will researchers use this checklist ?

Check all that apply

- Yes
- No
- Uncertain

* Will AI developers use it?

Check all that apply

- Yes
- No
- Uncertain

* Will mental health professionals use it?

Check all that apply

- Yes
- No
- Uncertain

* You can detail in this section any supplements, specifications or comments

Submit

ANNEXE C : QUESTIONNAIRE LIMESURVEY,DELPHI ETAPE 2

Language: English

There are 26 questions in this survey.

A note on privacy
This survey is anonymous.
The record of your survey responses does not contain any identifying information about you, unless a specific survey question explicitly asked for it. If you used an identifying token to access this survey, please rest assured that this token will not be stored together with your responses. It is managed in a separate database and will only be updated to indicate whether you did (or did not) complete this survey. There is no way of matching identification tokens with survey responses.

Modified items

✱ **Old version :** Objectives: Describe in clear terms your AIS' objectives and/or rationale

New version : Objectives: Describe your project's objectives and/or rationale and describe the Autonomous Intelligent System's role and functioning

Check all that apply

- ☐ Accept the new version of the item
- ☐ Reject the new version of the item
- ☐ Other:

✱ **Old version** : Technology: Name and describe the technologies used (e.g. machine learning). Mention the names of any technological intermediary or supplier allowing you to use the technology (e.g. technical provider, cloud provider)

New version : Technology: Name and describe the technologies and techniques used (e.g. supervised or unsupervised learning, machine learning, random forest, decision tree...). You can refer to the report of the AI Initiative incubated at Harvard http://ai-initiative.org/wp-content/uploads/2017/08/Making-the-AI-Revolution-work-for-everyone.-Report-to-OECD.-MARCH-2017.pdf. Mention the names of any technological intermediary or supplier allowing you to use the technology (e.g. technical provider, cloud provider)

Check all that apply

- ☐ Accept the new version of the item
- ☐ Reject the new version of the item
- ☐ Other:

✱ **Old version** : Funding and conflict of interest: Indicate your sources of funding (public and private) and who has an interest (e.g. financial) in your AIS. Detail any relationship that could influence the perception of your AIS

New version : Funding and conflict of interest: Indicate the sources of funding for your project (public and private) and who might have an interest (e.g. financial, political) in your Autonomous Intelligent System.

Check all that apply

- ☐ Accept the new version of the item
- ☐ Reject the new version of the item
- ☐ Other: [ ]

✱ **Old version** : Credentials: If you stated that you have specific expertise, clearly indicate the related professional, technical, academic or medical credentials

New version : Credentials: If you have cited that you or someone in your team has an expertise in relation to the Autonomous Intelligent System (e.g. in a document, a webpage, an interview) , clearly indicate the related professional, technical, academic or medical credentials and training (for instance : "Professor X. PhD at Harvard University, Specialist in the online detection of depression")

Check all that apply

- ☐ Accept the new version of the item
- ☐ Reject the new version of the item
- ☐ Other: [ ]

✱ **Old version** : Target population: Describe your target population, its size and the different sub-groups. Describe if you have involved some members, representatives, users of the population during the design of your AIS

New version : Target population: Describe your target population (or identify subgroups) and their size. Describe if and how the targeted population assisted in the design of your Autonomous Intelligent System.

Check all that apply

- ☐ Accept the new version of the item
- ☐ Reject the new version of the item
- ☐ Other: [ ]

✱ **Old version** : Evidence: If you made claims about your AIS' efficacy, performance of benefits, justify them and provide the evidence underlying those affirmations

New version : Evidence: If you made claims about your Autonomous Intelligent System's efficacy, performance or benefits, justify them and provide the evidence underlying those affirmations. If you have mentioned or used scientific papers, cite your sources

Check all that apply

- [ ] Accept the new version of the item
- [ ] Reject the new version of the item
- [ ] Other:

✱ **Old version** : Testing: Have you submitted your AIS to adversarial examples or worst-case scenarios

New version : Testing: If you have run your Autonomous Intelligent System under adversarial examples or worst-case scenarios, describe the type of tests used and the outcomes.

Check all that apply

- [ ] Accept the new version of the item
- [ ] Reject the new version of the item
- [ ] Other:

✱ **Old version** : Complaints: Describe a due process allowing users to formally complain or express their concerns

New version : Complaints: Describe the process whereby users can formally complain or express their concerns

Check all that apply

- [ ] Accept the new version of the item
- [ ] Reject the new version of the item
- [ ] Other:

✱ **Old version** : Responsibility: Describe who will be accountable for your AIS' actions or decisions

New version : Responsibility: Describe who will be accountable legally for your Autonomous Intelligent System's actions or decisions

Check all that apply

- [ ] Accept the new version of the item
- [ ] Reject the new version of the item
- [ ] Other:

✱ **Old version** : Data collection: Explain what data have been collected and used (for the training, evaluation and operation phases), for what purpose. Detail for how long data will be stored

New version : Data collection: Explain what data has been collected and used (for the training, evaluation and operation phases), who collected the data, who will have access to it, and what safeguards are in place to ensure secure storage.

Check all that apply

- [ ] Accept the new version of the item
- [ ] Reject the new version of the item
- [ ] Other:

✱ **Old version** : Informed consent: State if you have obtained informed consent, how, when. Describe its nature (formal, implied, renewable, dynamic)

New version : Informed consent : State whether you have obtained informed consent and, if so, how, when and from whom. Describe its nature (formal, implied, renewable, dynamic) and include the exact wording on the consent form. Explain if you have received the ethical approval from an institution (eg: hospital, university) for your consent forms.

Check all that apply

- [ ] Accept the new version of the item
- [ ] Reject the new version of the item
- [ ] Other:

✱ **Old version** : Minors: Are minors intentionally included in your population of users of your AIS, are any data related to minors targeted or is it possible that minors will be included or data concerning them? If not, indicate you you will Rejectedde minors and data concerning them. If so, indicate the specific measures to obtain their informed consent (or consent from their parents)

New version : Minors: Explain if minors' related information is used for the Autonomous Intelligent System. If intentionally collected: explain if the parents' need to give their consent. If unintentionally collected: describe what can be done to remove their information.

Check all that apply

- [ ] Accept the new version of the item
- [ ] Reject the new version of the item
- [ ] Other:

✱ **Old version** : "Blackbox": If you have used a technology to record all your AIS' actions, detail how it could be made available to justice or external observers or auditors

New version : Embedded Recording Mechanism : If you have used a technology to monitor and record all your Autonomous Intelligent System's decisions and actions, detail how and in what circumstances these records could be made available to authorities, external observers or auditors

Check all that apply

- [ ] Accept the new version of the item
- [ ] Reject the new version of the item

☐ Other:

✱ **Old version** : Autonomy: Detail the degree of autonomy of your AIS (e.g. partial or complete)

New version : Autonomy: Explain if your system has the autonomy to take actions or make decisions on its own. If yes, detail the degree of autonomy of your Autonomous Intelligent System (e.g. partial or complete)

Check all that apply

☐ Accept the new version of the item
☐ Reject the new version of the item
☐ Other:

✱ **Old version** : Moderation: Explain if your AIS requires a human intervention or moderation. If so, describe who will access to your AIS

New version : Moderation: Explain if your Autonomous Intelligent System requires human intervention or moderation. If so, describe who will have access to your Autonomous Intelligent System, what the regulations should be

Check all that apply

☐ Accept the new version of the item
☐ Reject the new version of the item
☐ Other:

✱ **Old version** : Optimal care: Is your AIS providing the target population with the most optimal care

New version : Type of care: Is your Autonomous Intelligent System helping its owners to provide the target population with the optimal treatment or a treatment as usual? Indicate the source of criteria for optimal treatment or treatment as usual

Check all that apply

☐ Accept the new version of the item
☐ Reject the new version of the item
☐ Other:

✱ **Old version** : Non-maleficence: Explain if and how your AIS could harm a user and how you avoid it or minimize any risk of harm

New version : Non-maleficence: Explain whether your Autonomous Intelligent System could harm a user and, if so, how. Explain how you avoid or minimize this risk

Check all that apply

- [ ] Accept the new version of the item
- [ ] Reject the new version of the item
- [ ] Other: [ ]

✱ **Old version** : Misuse: Describe and look for potential misuses of your AIS (e.g. imagine a victim-model scenario or worst-case scenario)

New version : Misuse: Describe potential misuses of your Autonomous Intelligent System (e.g. describe a possible negative scenario to imagine what could happen to a user) and describe your mitigating strategies

Check all that apply

- [ ] Accept the new version of the item
- [ ] Reject the new version of the item
- [ ] Other: [ ]

✱ **Old version** : Emotions detection: If your AIS detects emotions in the user, state if and how the user has been informed

New version : Emotions detection : If your Autonomous Intelligent System detects emotions in the user, state how, to what purpose. Explain whether the user is informed and if so, how

Check all that apply

- [ ] Accept the new version of the item
- [ ] Reject the new version of the item
- [ ] Other: [ ]

✱ **Old version** : Emotions control: If your AIS can provoke emotions, describe how users are informed of this possibility and describe how they might react to it

New version : Emotions control: If your Autonomous Intelligent System can provoke emotions, describe how users are informed of this possibility, the emotions that may be provoked ,their intensity, and possible impact on users

Check all that apply

- [ ] Accept the new version of the item
- [ ] Reject the new version of the item
- [ ] Other: [ ]

✱ **Old version** : Relationship : Describe if and how your AIS can create a relationship with users and how it could affect an individual

New version : Relationship : Is your user aware that he/she is interacting with a machine? Describe whether your Autonomous Intelligent System can create a relationship with users, and if so, how. Describe how the relationship might affect a user

Check all that apply

- [ ] Accept the new version of the item
- [ ] Reject the new version of the item
- [ ] Other: [ ]

✱ **Old version** : Public awareness: Describe the impact on users and potential users of public dissemination of information about your AIS and the process of its development

New version : Public awareness: Describe the impact of public dissemination of information on users from your Autonomous Intelligent System and the process of its development

Check all that apply

- [ ] Accept the new version of the item
- [ ] Reject the new version of the item
- [ ] Other: [ ]

✱ **Old version** : Detection: Explain any potential detection mistakes (false positives and false negatives) and estimate their extent (e.g. precision and recall)

New version : Detection: Explain any potential detection mistakes that might be done by your Autonomous Intelligent System(false positives and false negatives) and estimate their extent (e.g. precision and recall). Describe any potential adverse consequences for the users. If applicable, describe any incidental finding made by your Autonomous Intelligent System.

Check all that apply

- [ ] Accept the new version of the item
- [ ] Reject the new version of the item
- [ ] Other: [ ]

✱ **Old version** : Data handling: If applicable, describe the nature and purpose of any data manipulation (e.g. cleaning, transformation)

New version : Data handling: If applicable, describe the nature and purpose of any data manipulation (e.g. cleaning, transformation). Describe what will be done with the metadata

Check all that apply

- [ ] Accept the new version of the item
- [ ] Reject the new version of the item
- [ ] Other: [ ]

✱ **Old version** : Reporting: (If applicable) describe the nature and purpose of any statistical transformations applied to your data and potential bias related to the transformations

New version : Data transformation: If applicable, describe the nature and purpose of any statistical transformations applied to your data and potential bias related to the transformations or risks (e.g. ecological fallacy, confounding factors) Where is #43?? the numbers/titles don't correspond

Check all that apply

- ☐ Accept the new version of the item
- ☐ Reject the new version of the item
- ☐ Other:

✱ **Old version** : Other: If you have identified other potential methodological or scientific biases, list and describe them and their potential effects

New version : Other: If you have identified other potential methodological or scientific biases, describe them and their potential ethical consequences (example 1. an excessively long consent form could affect the informed consent; example 2. the presence of a floor effect in the measurements could constrain an Autonomous Intelligent System's ability to detect a behavior)

Check all that apply

- ☐ Accept the new version of the item
- ☐ Reject the new version of the item
- ☐ Other:

Submit

BIBLIOGRAPHIE

Adkins. (2017). Machine Learning and Electronic Health Records: A Paradigm Shift. *The American journal of psychiatry, 174*(2), 93-94. doi:10.1176/appi.ajp.2016.16101169

Anderson, & Anderson. (2014). *GenEth: A General Ethical Dilemma Analyzer.* Paper presented at the AAAI.

Anderson, Pace, Brandt, Nielsen, Allen, Libby, . . . Valuck. (2015). Monitoring suicidal patients in primary care using electronic health records. *J Am Board Fam Med, 28*(1), 65-71.

Arksey, & O'Malley. (2005). Scoping studies: towards a methodological framework. *International Journal of Social Research Methodology, 8*(1), 19-32.

Ayers. (2012). *Using aggregate Internet search queries for mental health surveillance.* (73), ProQuest Information & Learning, US.

Ayers, Althouse, Allem, Rosenquist, & Ford. (2013a). Seasonality in seeking mental health information on Google. *American Journal of Preventive Medicine, 44*(5), 520-525. doi:10.1016/j.amepre.2013.01.012

Ayers, Althouse, Allem, Rosenquist, & Ford. (2013b). Seasonality in seeking mental health information on Google. *Am J Prev Med, 44*(5), 520-525.

Beller, Chen, Wang, & Glasziou. (2013). Are systematic reviews up-to-date at the time of publication? *Systematic Reviews, 2*(1), 36.

Ben-Ari, & Hammond. (2015). Text mining the EMR for modeling and predicting suicidal behavior among US veterans of the 1991 Persian Gulf War. *System Sciences (HICSS), 2015 48th.*

Bennett, & Doub. (2016). Expert Systems in Mental Health Care: AI Applications in Decision-Making and Consultation. In *Artificial Intelligence in Behavioral and Mental Health Care* (pp. 27-51): Elsevier.

Bennett, & Hauser. (2013). Artificial intelligence framework for simulating clinical decision-making: A Markov decision process approach. *Artificial Intelligence in Medicine, 57*(1), 9-19.

Béranger. (2016). *Big data and ethics: the medical datasphere*: Elsevier.

Bertolote, & Wasserman. (2009). Development of definitions of suicidal behaviours: from suicidal thoughts to completed suicides. *Oxford textbook of suicidology and suicide prevention. A global perspective. Oxford University Press, Oxford*, 87-90.

Bharti, Panwar, Gopalakrishna, & Chellappan. (2017). Watch-Dog: Detecting Self-Harming Activities from Wrist Worn Accelerometers. *IEEE Journal of Biomedical and Health Informatics, PP*(99), 1-1. doi:10.1109/JBHI.2017.2692179

Biddle, Derges, Mars, & Heron. (2016). Suicide and the Internet: Changes in the accessibility of suicide-related information between 2007 and 2014. *Journal of Affective*.

Bowker. (2014). Big Data, Big Questions| The Theory/Data Thing. *International Journal of Communication, 8*(2043), 5-5.

Boyd, & Crawford. (2011). Six Provocations for Big Data. *Computer, 123*, 1-17. doi:10.2139/ssrn.1926431

Boyd, & Crawford. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society, 15*(5), 662-679.

Bragazzi. (2013a). A Google Trends-based approach for monitoring NSSI. *Psychol Res Behav Manag, 7*, 1-8. doi:10.2147/PRBM.S44084

Bragazzi. (2013b). A Google Trends-based approach for monitoring NSSI. *Psychology research and behavior management, 7*, 1-8.

Google searches for suicide and risk of suicide., 65 271-272 (2014).

Brundage, Avin, Clark, Toner, Eckersley, Garfinkel, . . . Filar. (2018). The malicious use of artificial intelligence: Forecasting, prevention, and mitigation. *arXiv preprint arXiv:1802.07228*.

Bunnik, Cawley, Mulqueen, & Zwitter. (2016). *Big Data Challenges: Society, Security, Innovation and Ethics*: Springer.

Callahan, Fuchs, Shelton, Balmer, Dudley, Gideon, . . . Cooper. (2013). Identifying suicidal behavior among adolescents using administrative claims data. *Pharmacoepidemiology and drug safety, 22*, 769-775.

Campolo, Sanfilippo, Whittaker, & Crawford. (2017). AI Now 2017 Report. *AI Now Institute at New York University*.

Bill C-28, (2010).

Canada. (2014). Mental Health and Wellness. Retrieved from http://cbpp-pcpe.phac-aspc.gc.ca/public-health-topics/mental-health-and-wellness/

Caouette, Vincent, & Montreuil. (2007). Utilisation de la télésurveillance auprès des personnes âgées à domicile: pratiques actuelles et potentielles. *Canadian Journal of Occupational Therapy, 74*(5), 382-392.

Cardon. (2015). *A quoi rêvent les algorithmes: Nos vies à l'heure des big data*: Seuil.

Chang, Kwok, Cheng, Yip, & Chen. (2015a). The association of trends in charcoal-burning suicide with Google search and newspaper reporting in Taiwan: a time series analysis. *Soc Psychiatry Psychiatr Epidemiol, 50*(9), 1451-1461.

Chang, Kwok, Cheng, Yip, & Chen. (2015b). The association of trends in charcoal-burning suicide with Google search and newspaper reporting in Taiwan: a time series

analysis. *Social Psychiatry and Psychiatric Epidemiology, 50*(9), 1451-1461. doi:10.1007/s00127-015-1057-7

Chen, Chai, Zhang, & Wang. (2014). Development and application of a Chinese webpage suicide information mining system (sims). *Journal of medical systems, 38*(11), 88.

Chow, Leung, Lai, Chow, Chung, Tong, & Lit. (2012). Development of an electronic emergency department-based geo-information injury surveillance system in Hong Kong. *Injury, 43*(6), 739-748.

Christensen, Batterham, & O'Dea. (2014). E-health interventions for suicide prevention. *International journal of environmental research and public health, 11*(8), 8193-8212. doi:10.3390/ijerph110808193

CNIL. (2017). *Comment permettre à l'Homme de garder la main.* Retrieved from

Collings, & Niederkrotenthaler. (2012). Suicide prevention and emergent media: surfing the opportunity. In: Hogrefe Publishing.

Conway, & O'Connor. (2016). Social media, big data, and mental health: Current advances and ethical implications. *Current Opinion in Psychology, 9*, 77-82. doi:10.1016/j.copsyc.2016.01.004

Crisp. (2011). Ethics. https://www.rep.routledge.com/articles/overview/ethics/v-2/sections/ethics-and-metaethics doi:10.4324/9780415249126-L132-2

Crosby, Gfroerer, Han, Ortega, & Parks. (2011). *Suicidal thoughts and behaviors among adults aged>_18 Years--United States, 2008-2009*: US Department of Health and Human Services, Centers for Disease Control and Prevention.

De Montjoye, Shmueli, Wang, & Pentland. (2014). openpds: Protecting the privacy of metadata through safeanswers. *PLoS ONE, 9*(7), e98790.

DeLeo, Bille-Brahe, Kerkhof, & Schmidtke. (2004). *Suicidal behaviour: Theories and research findings*: Hogrefe Publishing.

Demiaux, & Si Abdallah. (2017). *Comment permettre à l'homme de garder la main? Les enjeux éthiques des algorithmes et de l'intelligence artificielle.* Retrieved from

Derogatis. (1979). Symptom Checklist-90-Revised (SCL-90-R). *Lyndhurst, NJ: NCS Pearson.*

Derogatis, & Unger. (2010). Symptom checklist-90-revised. *The Corsini encyclopedia of psychology*, 1-2.

Drachsler, & Greller. (2016). *Privacy and analytics.* Paper presented at the Proceedings of the Sixth International Conference on Learning Analytics & Knowledge - LAK '16.

Drife. (2010). An avalanche of apple pie: too many guidelines? *BMJ: British Medical Journal (Online), 340.*

Durkee, Hadlaczky, Westerlund, & Carli. (2011). Internet pathways in suicidality: a review of the evidence. *International journal of environmental research and public health, 8*(10), 3938-3952. doi:10.3390/ijerph8103938

East, & Havard. (2015). Mental health mobile apps: from infusion to diffusion in the mental health social system. *JMIR mental health, 2*(1), e10.

Elish, & Boyd. (2017). Situating methods in the magic of Big Data and AI. *Communication Monographs*, 1-24.

Excellence. (2009). The guidelines manual. *London: NiCe*.

Eysenbach. (2011). Infodemiology and infoveillance: tracking online health information and cyberbehavior for public health. *American Journal of Preventive Medicine, 40*(5), S154-S158.

Facebook. (2018, September 10, 2018). How Facebook AI Helps Suicide Prevention. Retrieved from https://newsroom.fb.com/news/2018/09/inside-feed-suicide-prevention-and-ai/

Fiedorowicz, & Chigurupati. (2009). The Internet in suicide prevention and promotion. *Internet and Suicide. New York, NY: Nova Science Publishers*, 1-12.

Flick. (2016). Informed consent and the Facebook emotional manipulation study. *Research Ethics, 12*(1), 14-28.

Galtung, & Hoff. (2007). *Selvmordsfare og internett: kan jussen gi beskyttelse?* : Universitetet i Oslo, Det medisinske fakultet, Institutt for psykiatri, Seksjon for selvmordsforskning og-forebygging.

The future of "big data" in suicide behaviors research: Can we compare the experiences of the U.S. and U.K. armed forces?, 78 25-28 (2015).

Gottfredson, Cook, Gardner, Gorman-Smith, Howe, Sandler, & Zafft. (2015). Standards of evidence for efficacy, effectiveness, and scale-up research in prevention science: Next generation. *Prevention Science, 16*(7), 893-926.

Grant, & Booth. (2009). A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal, 26*(2), 91-108.

Greenberg. (2003). Metadata and the world wide web. *Encyclopedia of library and information science, 3*, 1876-1888.

Grella, & Gilmore. (2002). Improving service delivery to the dually diagnosed in Los Angeles County. *Journal of Substance Abuse Treatment, 23*(2), 115-122. doi:10.1016/S0740-5472(02)00254-4

Grunebaum. (2015). Suicidology Meets "Big Data". *The Journal of clinical psychiatry, 76*, 383-384.

Guan, Hao, Cheng, Yip, & Zhu. (2015). Identifying Chinese Microblog Users With High Suicide Probability Using Internet-Based Profile and Linguistic Features: Classification Model. *JMIR mental health, 2*, e17.

Gunn 3rd, & Lester. (2013). Using google searches on the internet to monitor suicidal behavior. *J Affect Disord, 148*(2-3), 411-412. doi:10.1016/j.jad.2012.11.004

10.1016/j.jad.2012.11.004. Epub 2012 Nov 23.

Gunn, & Lester. (2013). Using google searches on the internet to monitor suicidal behavior. *Journal of affective disorders, 148*(2-3), 411-412.

Haerian, Salmasian, & Friedman. (2012). Methods for identifying suicide or suicidal ideation in EHRs. *AMIA ... Annual Symposium proceedings. AMIA Symposium, 2012*, 1244-1253.

Hagihara, Miyazaki, & Tarumi. (2010). Internet use and suicide among younger age groups between 1989 and 2008 in Japan. *Acta Psychiatrica Scandinavica, 121*(6), 485-485.

Application of a Hybrid Text Mining Approach to the Study of Suicidal Behavior in a Large Population, 2555-2561 (2014).

Use of Text Search to Effectively Identify Lifetime Prevalence of Suicide Attempts among Veterans, 2676-2683 (2013).

Hardy. (2016). Silicon Valley looks to artificial intelligence for the next big thing. *The New York Times.*

Harris, McLean, & Sheffield. (2009). Solving suicidal problems online: Who turns to the Internet for help? *Australian e-Journal for the Advancement of Mental Health, 8*(1), 28-36.

Hasson, Keeney, & McKenna. (2000). Research guidelines for the Delphi survey technique. *Journal of Advanced Nursing, 32*(4), 1008-1015.

Hermenegildo. (2012). Conferences vs. Journals in CS, what to do? *Evolutionary ways forward and the ICLP/TPLP model. Leibniz-ZentrumfürInformatik.*

Hinduja, & Patchin. (2008). Cyberbullying: An Exploratory Analysis of Factors Related to Offending and Victimization. *Deviant Behavior, 29*(2), 129-156. doi:10.1080/01639620701457816

Hogle. (2016). The ethics and politics of infrastructures: Creating the conditions of possibility for big data in medicine. In *The ethics of biomedical big data* (pp. 397-427): Springer.

Hollis, Morriss, Martin, Amani, Cotton, Denis, & Lewis. (2015). Technological innovations in mental healthcare: harnessing the digital revolution. *The British Journal of Psychiatry, 206*(4), 263-265.

IEEE. (2017). *Ethically Aligned Design.* Retrieved from

Inspq. (2007). La mortalité par suicide au Québec : tendances et données récentes – 1981 à 2007. *Institut National de Santé Publique du Québec*, 1-20.

Jacob, Scourfield, & Evans. (2014). Suicide prevention via the Internet: A descriptive review. *Crisis: The Journal of Crisis Intervention and Suicide Prevention, 35*(4), 261-267. doi:10.1027/0227-5910/a000254

Jashinsky, Burton, Hanson, West, Giraud-Carrier, Barnes, & Argyle. (2014a). Tracking suicide risk factors through Twitter in the US. *Crisis, 35*(1), 51-59. doi:10.1027/0227-5910/a000234

Jashinsky, Burton, Hanson, West, Giraud-Carrier, Barnes, & Argyle. (2014b). Tracking suicide risk factors through Twitter in the US. *Crisis, 35*(1), 51-59.

John, Dennis, Kosnes, Gunnell, Scourfield, Ford, & Lloyd. (2014a). Suicide Information Database-Cymru: a protocol for a population-based, routinely collected data linkage study to explore risks and patterns of healthcare contact prior to suicide to identify opportunities for intervention. *BMJ Open, 4*(11), e006780-e006780. doi:10.1136/bmjopen-2014-006780

John, Dennis, Kosnes, Gunnell, Scourfield, Ford, & Lloyd. (2014b). Suicide Information Database-Cymru: a protocol for a population-based, routinely collected data linkage study to explore risks and patterns of healthcare contact prior to suicide to identify opportunities for intervention. *BMJ Open, 4*, e006780.

Jorm. (2015). Using the Delphi expert consensus method in mental health research. *Australian & New Zealand Journal of Psychiatry, 49*(10), 887-897.

Karlin. (2018). A Canadian Algorithmic Impact Assessment. Retrieved from https://medium.com/@supergovernance/a-canadian-algorithmic-impact-assessment-128a2b2e7f85

Kessler, Hwang, Hoffmire, McCarthy, Petukhova, Rosellini, . . . Bossarte. (2017). Developing a practical suicide risk prediction model for targeting high-risk patients in the Veterans health Administration. *Int J Methods Psychiatr Res*.

Kessler, Stein, Petukhova, Bliese, Bossarte, Bromet, . . . Ur. (2017). Predicting suicides after outpatient mental health visits in the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). *Mol Psychiatry, 22*(4), 544-551.

Kitchin. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*: Sage.

Kitchin, & Lauriault. (2014). Towards critical data studies : Charting and unpacking data assemblages and their work. *Geoweb and Big Data*.

Krysinska. (2010). Review of Internet and Suicide: An overview of the first decade of research, ideas and debates. *Advances in Mental Health, 9*(2), 206-207.

Krysinska, Westerlund, & Niederkrotenthaler. (2017). A mapping study on the internet and suicide. *Crisis*.

Krysinska, Westerlund, Niederkrotenthaler, Andriessen, Carli, Hadlaczky, . . . Wasserman. (2017a). A Mapping Study on the Internet and Suicide. *Crisis: The Journal of Crisis Intervention and Suicide Prevention*, No Pagination Specified-No Pagination Specified. doi:10.1027/0227-5910/a000444

Krysinska, Westerlund, Niederkrotenthaler, Andriessen, Carli, Hadlaczky, . . . Wasserman. (2017b). A Mapping Study on the Internet and Suicide. *Crisis*, 1-10.

Kshetri. (2014). Big data' s impact on privacy, security and consumer welfare. *Telecommunications Policy, 38*(11), 1134-1145.

Kwon, Kang, & Lee. (2016). Sexually transmissible infections in middle and high school students: experience rates, risk factors and relationship with mental health - results from the Korean youth risk behaviour web-based survey. *Sexual health, 13*(1), 29-34. doi:10.1071/SH15147

Labelle, Bibaud-De Serres, & Leblanc. (2013). Innovating to Treat Depression and Prevent Suicide: The IPhone@ PSY ASSISTANCE Application. In *Suicide Prevention and New Technologies* (pp. 166-180): Springer.

Lai, Maniam, Chan, & Ravindran. (2014). Caught in the web: a review of web-based suicide prevention. *J Med Internet Res, 16*(1), e30. doi:10.2196/jmir.2973

Lakeman, & Fitzgerald. (2009a). Ethical suicide research: a survey of researchers. *International Journal of Mental Health Nursing, 18*(1), 10-17. doi:10.1111/j.1447-0349.2008.00569.x

Lakeman, & Fitzgerald. (2009b). The ethics of suicide research. *Crisis, 30*(1), 13-19. doi:10.1027/0227-5910.30.1.13

Larsen, Cummins, Boonstra, O'Dea, Tighe, Nicholas, . . . Christensen. (2015). The use of technology in Suicide Prevention. In *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference* (Vol. 2015, pp. 7316-7319).

Lee, & Ho. (2012). A suicide case information service platform for clinical staff. In *2012 International Conference on Machine Learning and Cybernetics* (Vol. 5, pp. 1921-1926).

Lee, Kim, Lee, Kim, Lee, Ju, & Myung. (2014). Detection of a Suicide by Hanging Based on a 3-D Image Analysis. *IEEE Sensors Journal, 14*(9), 2934-2935. doi:10.1109/jsen.2014.2332070

Lee, Lee, & Choi. (2016). Examining the Relationship Between Past Orientation and US Suicide Rates: An Analysis Using Big Data-Driven Google Search Queries. In *Journal of Medical Internet Research* (Vol. 18, pp. e35).

Leetaru. (2016). Are research ethics obsolete in the era of big data? *Forbes*(June 17), 1-53.

Légaré, & Gagné. (2015). La mortalité par suicide au Québec : 1981 à 2012 – Mise à jour 2015.

Légaré, & Gagné. (2016). La mortalité par suicide au Québec : 1981 à 2013 – Mise à jour 2016.

Lehavot, Ben-Zeev, & Neville. (2012). Ethical considerations and social media: A case of suicidal postings on Facebook. *Journal of Dual Diagnosis, 8*(4), 341-346. doi:10.1080/15504263.2012.718928

Letrilliart, & Vanmeerbeek. (2011). À la recherche du consensus: quelle méthode utiliser? *Exercer, 99*, 170-177.

Lewis. (2005). *Concept mapping school personnel perceptions of adolescent suicide and its prevention.* State University of New York at Buffalo, Ann Arbor. Retrieved from https://search.proquest.com/docview/305379621?accountid=14719

http://openurl.uquebec.ca:9003/uqam?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:ProQuest+Dissertations+%26+Theses+Global&atitle=&ti

Lewis, & Seko. (2016a). A Double-Edged Sword: A Review of Benefits and Risks of Online Nonsuicidal Self-Injury Activities. *Journal of Clinical Psychology, 72*(3), 249-262. doi:10.1002/jclp.22242

Lewis, & Seko. (2016b). A Double-Edged Sword: A Review of Benefits and Risks of Online Nonsuicidal Self-Injury Activities. In *Journal of Clinical Psychology* (Vol. 72, pp. 249-262).

Linstone, & Turoff. (2002). *The Delphi method: Techniques and applications* (Vol. 18): Addison-Wesley Publishing Company, Advanced Book Program.

Lovisi. (2012). The relationship between elderly suicide rates and the internet: A cross-national study. In *International Journal of Social Psychiatry* (Vol. 58, pp. 337-337).

Ludwig. (1994). *Internationalizing Extension: An exploration of the characteristics evident in a state university Extension system that achieves internationalization.* The Ohio State University,

Luxton. (2014a). Artificial intelligence in psychological practice: Current and future applications and implications. *Professional Psychology: Research and Practice, 45*(5), 332-332.

Luxton. (2014b). Recommendations for the ethical use and design of artificial intelligent care providers. *Artificial Intelligence in Medicine, 62*(1), 1-10.

Luxton. (2015). *Artificial intelligence in behavioral and mental health care*: Academic Press.

Ma-Kellams, Or, Baek, & Kawachi. (2016). Rethinking suicide surveillance: Google search data and self-reported suicidality differentially estimate completed suicide risk. In *Clinical Psychological Science* (Vol. 4, pp. 480-484).

Mahajan, Campbell, Cohen, Grande, Regina, & Hopkins. (2014). *La cybersanté mentale au Canada : Transformer le système de santé mentale grâce à la technologie*. Retrieved from

Malott, Bharti, Hilbert, Gopalakrishna, & Chellappan. (2015). Detecting self-harming activities with wearable devices. In *2015 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)* (pp. 597-602).

Maris. (2002). Suicide. *Lancet, 360*(9329), 319-326. doi:10.1016/S0140-6736(02)09556-9

Maris, Berman, & Silverman. (2000). The theoretical component in suicidology. *Comprehensive textbook of suicidology*, 26-61.

Mars, Heron, Biddle, Donovan, & Holley. (2015). Exposure to, and searching for, information about suicide and self-harm on the Internet: Prevalence and predictors in a population based cohort of young. *Journal of Affective*.

Masri, & Jani. (2012). *Employing artificial intelligence techniques in Mental Health Diagnostic Expert System.* Paper presented at the Computer & Information Science (ICCIS), 2012 International Conference on.

Mayer-Schönberger, & Cukier. (2013). *Big data: A revolution that will transform how we live, work, and think*: Houghton Mifflin Harcourt.

McCarthy. (1998). What is artificial intelligence?

McCarthy. (2010). Internet monitoring of suicide risk in the population. In *Journal of affective disorders* (Vol. 122, pp. 277-279).

McCoy, Castro, Roberson, Snapper, & Perlis. (2016). Improving Prediction of Suicide and Accidental Death After Discharge From General Hospitals With Natural Language Processing. In *JAMA Psychiatry* (Vol. 73, pp. 1064-1071).

Ménard, Mondoux, Ouellet, & Bonenfant. (2016). Big Data, gouvernementalité et industrialisation des médiations symboliques politico-institutionnelles. *In Press*.

Metzger, Tvardik, Gicquel, Bouvry, Poulet, & Potinet-Pagliaroli. (2016). Use of emergency department electronic medical records for automated epidemiological surveillance of suicide attempts: a French pilot study. In *International Journal of Methods in Psychiatric Research.*

MHCC. (2017). E-Mental Health. Retrieved from https://www.mentalhealthcommission.ca/English/what-we-do/e-mental-health

MHCC. (2018). Suicide Prevention. Retrieved from https://www.mentalhealthcommission.ca/English/what-we-do/suicide-prevention

Mishara, & Côté. (2013). Prevention, Suicide new technologies: Towards evidence based practice. In (pp. 1-23). New York: Palgrave Macmillan.

Mishara, & Kerkhof. (2013). The Future of Suicide Prevention at Our Doorstep. In L & Kerkhof (Eds.), (pp. ix--xii). New York: Palgrave Macmillan.

Mishara, & Tousignant. (2004). *Comprendre le suicide*: PUM.

Mishara, & Weisstub. (2005). Ethical and legal issues in suicide research. *International Journal of Law and Psychiatry, 28*, 23-41. doi:10.1016/j.ijlp.2004.12.006

Mishara, & Weisstub. (2007a). Ethical, legal, and practical issues in the control and regulation of suicide promotion and assistance over the Internet. *Suicide & life-threatening behavior, 37*(1), 58-65. doi:10.1521/suli.2007.37.1.58

Mishara, & Weisstub. (2007b). Ethical, legal, and practical issues in the control and regulation of suicide promotion and assistance over the Internet. *Suicide and Life-Threatening*.

Mishara, & Weisstub. (2013a). Challenges in the Control and Regulation of Suicide Promotion and Assistance over the Internet. *Suicide Prevention and New Technologies*.

Mishara, & Weisstub. (2013b). Challenges in the control and regulation of suicide promotion and assistance over the Internet. *Suicide Prevention and New Technologies. Palgrave Macmillan, Hampshire, UK*, 63-75.

Mismetti, & Laporte. (2003). Méthodologie des essais cliniques: pourquoi et pour qui? Les notions essentielles à une lecture critique. *Médecine thérapeutique, 9*(3), 68-72.

MIT. (2011). What is Civic Media. Retrieved from http://civic.mit.edu/blog/henry/what-is-civic-media

Mitra, Shriberg, Vergyri, Knoth, & Salomon. (2015). Cross-corpus depression prediction from speech. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 4769-4773).

Mittelstadt, & Floridi. (2016a). The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts. *Science and Engineering Ethics, 22*(2), 303-341. doi:10.1007/s11948-015-9652-2

Mittelstadt, & Floridi. (2016b). *The ethics of biomedical big data* (Vol. 29): Springer.

Moher, Liberati, Tetzlaff, & Altman. (2009). Systematic Reviews and Meta-Analyses: The PRISMA Statement. *Annulas of Internal Medicine, 151*(4), 264-269. doi:10.1371/journal.pmed1000097

Moher, Shamseer, Clarke, Ghersi, Liberati, Petticrew, . . . Stewart. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews, 4*(1), 1-1. doi:10.1186/2046-4053-4-1

Mohr, Ph, Burns, Ph, Schueller, Ph, . . . S. (2013). Behavioral Intervention Technologies : Evidence review and recommendations for future research in mental health. *General Hospital Psychiatry*. doi:10.1016/j.genhosppsych.2013.03.008

Mokkenstorm, & Huisman. (2011). Suicide prevention via the internet and the telephone: 113Online. *Tijdschrift voor*.

Mokkink, Terwee, Patrick, Alonso, Stratford, Knol, . . . De Vet. (2010). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Quality of life research, 19*(4), 539-549.

Mörch, Mishara, Plourde-Léveillé, Corthésy-Blondin, Côté, Bardon, & Labelle. (2018). Big Data, Artificial Intelligence and Suicide prevention: a scoping study. *Manuscript not published.*

Moreno, & Redondo. (2016). Text analytics: The convergence of big data and artificial intelligence. *IJIMAI, 3*(6), 57-64.

Munn, Peters, Stern, Tufanaru, McArthur, & Aromataris. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Med Res Methodol, 18*(1), 143. doi:10.1186/s12874-018-0611-x

Narang, & Lippmann. (2009). The Internet: its role in the occurrence and prevention of suicide. *Internet and Suicide. New York, NY: Nova Science Publishers*, 13-20.

Niezen. (2013). Internet suicide: communities of affirmation and the lethality of communication. *Transcultural Psychiatry, 50*, 303-322.

Nock. (2014). *The Oxford handbook of suicide and self-injury*: Oxford University Press.

O'Carroll, Berman, Maris, Moscicki, Tanney, & Silverman. (1996). Beyond the Tower of Babel: a nomenclature for suicidology. *Suicide & life-threatening behavior, 26*(3), 237-252. doi:10.1111/j.1943-278X.1996.tb00609.x

O'Leary. (2013). Artificial intelligence and big data. *IEEE Intelligent Systems, 28*(2), 96-99.

Oquendo, Feldman, Silverman, Currier, Brown, Chen, . . . Pearson. (2011). Variability in the definition and reporting of adverse events in suicide prevention trials: an examination of the issues and a proposed solution. *Archives of suicide research: official journal of the International Academy for Suicide Research, 15*(1), 29-42. doi:10.1080/13811118.2011.541146

Oremus. (2013). Martin Manley's sister asks Yahoo to put his suicide website back up. *www.slate.com*.

Ouellet, Mondoux, Ménard, Bonenfant, & Richert. (2014). " Big Data", gouvernance et surveillance.

Pestian. (2010). A conversation with Edwin Shneidman. In *Suicide and Life-Threatening Behavior* (Vol. 40, pp. 516-523).

Pestian, Grupp-Phelan, Bretonnel Cohen, Meyers, Richey, Matykiewicz, & Sorter. (2016). A Controlled Trial Using Natural Language Processing to Examine the Language of Suicidal Adolescents in the Emergency Department. In *Suicide & life-threatening behavior* (Vol. 46, pp. 154-159).

Pestian, Sorter, Connolly, Cohen, McCullumsmith, Gee, . . . Rohlfs. (2017). A machine learning approach to identifying the thought markers of suicidal subjects: A prospective multicenter trial. In *Suicide and Life-Threatening Behavior* (Vol. 47, pp. 112-121).

Pirkis, Neal, Dare, Blood R, & Studdert. (2009). Legal Bans on Pro-Suicide Web Sites: An Early Retrospective from Australia. *Suicide and Life-Threatening Behavior, 39*(2), 190-193.

Posner, Brodsky, KseniyaYershova, & Mann. (2014). The Classification of Suicidal Behavior 2. *The Oxford handbook of suicide and self-injury*, 7-7.

Posner, Melvin, Stanley, Oquendo, & Gould. (2007). Factors in the assessment of suicidality in youth. *CNS spectrums, 12*(02), 156-162.

Poulin. (2014). Big Data Custodianship in a Global Society. *SAIS Review of International Affairs, 34*(1), 109-116. doi:10.1353/sais.2014.0002

Poulin, Shiner, Thompson, Vepstas, Young-Xu, Goertzel, . . . McAllister. (2014a). Predicting the risk of suicide by analyzing the text of clinical notes. *PLoS ONE, 9*(1).

Poulin, Shiner, Thompson, Vepstas, Young-Xu, Goertzel, . . . McAllister. (2014b). Predicting the risk of suicide by analyzing the text of clinical notes. In *PLoS ONE* (Vol. 9).

Poulin, Thompson, & Bryan. (2015). Public health surveillance: predictive analytics and Big data. *Artificial Intelligence in Behavioral and Mental Health Care*, 205-205.

Poulin, Thompson, & Bryan. (2016). Public health surveillance: Predictive analytics and big data. In *Artificial Intelligence in Behavioral and Mental Health Care* (pp. 205-230).

Rajeswari, Sowmbika, Kalaimagal, Ramya, & Ranjitha. (2016). Improved emotional speech recognition algorithms. In *2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)* (pp. 2362-2366).

Ramchurn, Vytelingum, Rogers, & Jennings. (2012). Putting the'smarts' into the smart grid: a grand challenge for artificial intelligence. *Communications of the ACM, 55*(4), 86-97.

Recupero. (2012). Suicide and the internet. *The American Psychiatric Publishing textbook of suicide assessment and management (2nd ed.)*, 515-538.

Reidenberg, & Gordon. (2012). *Using Big Data to Understand and Prevent Suicide in Communications and Technology for Violence Prevention for Violence Prevention: Worksop Summary* (9780309253512). Retrieved from Washington:

Richards, & King. (2013). Three paradoxes of big data. *Stan. L. Rev. Online, 66*, 41.

Riek. (2016). Robotics technology in mental health care. In *Artificial Intelligence in Behavioral and Mental Health Care* (pp. 185-203): Elsevier.

Rockett, Lilly, Jia, Larkin, Miller, Nelson, . . . Caine. (2016). Self-injury Mortality in the United States in the Early 21st Century: A Comparison With Proximally Ranked Diseases. *JAMA Psychiatry, 73*(10), 1072-1081. doi:10.1001/jamapsychiatry.2016.1870

Rouvroy. (2015). Big Data Is Algorithming You. 1-4.

Rumshisky, Ghassemi, Naumann, Szolovits, Castro, McCoy, & Perlis. (2016). Predicting early psychiatric readmission with natural language processing of narrative discharge summaries. *Translational psychiatry, 6*(10), e921.

Runeson, Odeberg, Pettersson, Edbom, Adamsson, & Waern. (2017). Instruments for the assessment of suicide risk: A systematic review evaluating the certainty of the evidence. *PLoS ONE, 12*(7), e0180292.

Vehicle cabin safety alert system, 1-4 (2015).

Schapire. (2008). Cos 511: Theoretical machine learning. *FTP: http://www. cs. princeton. edu/courses/archive/spr08/cos511/scribe notes/0204. pdf.*

Reduced vowel space is a robust indicator of psychological distress: A cross-corpus analysis, 4789-4793 (2015).

Investigating the speech characteristics of suicidal adolescents, 709-713 (2013).

Silverman. (2006). The language of suicidology. *Suicide & life-threatening behavior, 36*(5), 519-532. doi:10.1521/suli.2006.36.5.519

Silverman. (2011). Challenges to Classifying Suicidal Ideations, Communications, and Behaviours. In (pp. 7-25): John Wiley & Sons, Ltd.

Solano, Ustulin, Pizzorno, Vichi, Pompili, Serafini, & Amore. (2016). A Google-based approach for monitoring suicide risk. *Psychiatry Research, 246*, 581-586.

Song, Song, An, Hayman, & Woo. (2014). Psychological and social factors affecting Internet searches on suicide in Korea: a big data analysis of Google search trends. *Yonsei medical journal, 55*, 254-263.

Song, Song, Seo, & Jin. (2016). Data Mining of Web-Based Documents on Social Networking Sites That Included Suicide-Related Words Among Korean Adolescents. *J Adolesc Health, 59*(6), 668-673.

Southey (2013). [Bill C-13 is about a lot more than cyberbullying].

Spitzer, Endicott, Cohen, & Fleiss. (1974). Constraints on the validity of computer diagnosis. *Archives of General Psychiatry, 31*(2), 197-203.

Sueki. (2012). Association between deliberate self-harm-related internet searches and the mental states and lifetime suicidal behaviors of Japanese young adults. *Psychiatry and Clinical Neurosciences.*

Tal, & Torous. (2017). The digital mental health revolution: Opportunities and risks.

Tam, Tang, & Fernando. (2007). The internet and suicide: A double-edged tool. In (Vol. 18, pp. 453-455).

Terry. (2014). Health privacy is difficult but not impossible in a post-hipaa data-driven world. *Chest, 146*(3), 835-840.

Tran, Luo, Phung, Harvey, Berk, Kennedy, & Venkatesh. (2014). Risk stratification using data from electronic medical records better predicts suicide risks than clinician assessments. *BMC Psychiatry, 14.*

Tran, Nguyen, Phung, & Venkatesh. (2015). Learning vector representation of medical objects via EMR-driven nonnegative restricted Boltzmann machines (eNRBM). *Journal of biomedical informatics, 54*, 96-105.

UdeM. (2018). *The Montreal Declaration for a Responsible Development of Artificial.* Retrieved from https://www.montrealdeclaration-responsibleai.com/

UNESCO. (2019). Retrieved from http://uis.unesco.org/en/glossary-term/information-and-communication-technologies-ict

Vahabzadeh, Sahin, & Kalali. (2016). Digital suicide prevention: Can technology become a game-changer? *Innovations in Clinical Neuroscience, 13*(5-6), 16-20.

van der Sloot. (2015). How to assess privacy violations in the age of Big Data? Analysing the three different tests developed by the ECtHR and adding for a fourth one. *Information & Communications Technology Law*(April 2015), 1-30. doi:10.1080/13600834.2015.1009714

Van Rysewyk, & Pontier. (2014). Machine medical ethics.

Suicide detection system based on Twitter, 785-788 (2014).

Adolescent suicidal risk assessment in clinician-patient interaction: A study of verbal and acoustic behaviors, 277-282 (2014).

Adolescent Suicidal Risk Assessment in Clinician-Patient Interaction, PP 1 (2017).

Vermeulen. (2018). *Manual for ESHRE guideline development.* Retrieved from https://www.eshre.eu/Guidelines-and-Legal/Guidelines/Guideline-development-process.aspx:

Villani, Schoenauer, Bonnet, Berthet, Cornut, Levin, . . . Biabiaby-Rosier. (2017). Donner un sens à l'intelligence artificielle, Pour une stratégie nationale et européenne. In: Rapport de la mission parlementaire du.

Voigt, & Von dem Bussche. (2017). *The EU General Data Protection Regulation (GDPR)* (Vol. 18): Springer.

Weathers, Litz, Herman, Huska, & Keane. (1993). *The PTSD Checklist (PCL): Reliability, validity, and diagnostic utility.* Paper presented at the annual convention of the international society for traumatic stress studies, San Antonio, TX.

Weinstock. (2011). *Profession éthicien*: Les Presses de l'Université de Montréal.

Westerlund, & Wasserman. (2009). The role of the Internet in suicide prevention. In (pp. 525-532). Oxford, UK: Oxford University Press.

WHO. (1948). Definition of health. *http://www.who.int/suggestions/faq/zh/index.html*.

WHO. (2014). *Preventing suicide: A global imperative* (9241564776). Retrieved from

WHO. (2018). Suicide. Retrieved from https://www.who.int/news-room/fact-sheets/detail/suicide

Windfuhr, & Kapur. (2011). Suicide and mental illness: a clinical review of 15 years findings from the UK National Confidential Inquiry into Suicide. *British medical bulletin, 100*, 101-121. doi:10.1093/bmb/ldr042

Winter, & Davidson. (2019). Governance of artificial intelligence and personal health information. *Digital Policy, Regulation and Governance*.

Zuckerman. (2014). New media, new civics? *Policy & Internet, 6*(2), 151-168.

Zwitter. (2016). The Network Effect on Ethics in the Big Data Age. In *Big Data Challenges* (pp. 23-34): Springer.