

Éthique et intelligence artificielle

Quels principes appliquer
pour le domaine de la santé mentale?



Webinaire Mars 2019
Carl Mörch, M.Psy.
Dominic Martin, Ph.D.

Regroupements stratégiques *Santé mentale, TIC
et Santé & Éthique* du Réseau de recherche en
santé des populations du Québec (RRSPQ)

Le webinaire en bref

- I. Pour commencer : Quel besoin de parler d'éthique de l'IA, et pourquoi en santé mentale? - Carl et Dominic
- II. Définir brièvement les concepts de Intelligence Artificielle et « Big Data » - Carl et Dominic
- III. Quels sont les enjeux éthiques importants en matière d'IA - Dominic
- IV. IA en Santé Mentale, applications et défis pour le domaine - Carl
- V. Comment renforcer la présence de l'éthique, dans le milieu des affaires, chez ceux qui développent l'IA, le ML en particulier - Dominic
- VI. Vers une éthique de l'IA en santé mentale, une proposition de protocole éthique - Carl et Dominic

Pourquoi parler d'éthique de l'IA en santé mentale

- ▶ **Évolutions récentes importantes** (en clinique, recherche et santé publique)
- ▶ **Décalage** entre la vitesse d'évolution technologique et l'évolution des formations en santé mentale
- ▶ **Besoins potentiels** d'une meilleure connaissance éthique et technique

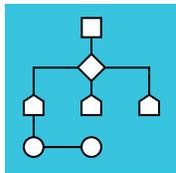
Définir l'Intelligence Artificielle

- ▶ Le terme désigne généralement le domaine qui s'intéresse à **l'étude et au design de machines intelligentes** (McCarthy, 1998).
- ▶ Ces machines sont parfois appelées **Autonomous Intelligent Systems (AIS)** (IEEE, 2017), peuvent être des robots.
- ▶ Son but est de « **construire des machines qui sont capables de réaliser des tâches que nous définirons comme requérant de l'intelligence** »
Exemple : le raisonnement, l'apprentissage, la planification la résolution de problème et la perception (Luxton, 2015, p. 2).
- ▶ Quelques techniques : réseaux neuronaux, deep learning, les algorithmes automatiques de classification

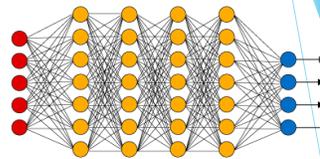
Définir l'IA

- ▶ Une discipline de l'intelligence artificielle est particulièrement utilisée : le **machine learning** ou apprentissage automatique
- ▶ Cette discipline permet à des systèmes informatiques **d'apprendre et d'exécuter des tâches de façon autonome**, en les entraînant d'abord
- ▶ Cette discipline **s'est développée grâce à la multiplication des données et l'augmentation de la puissance de calcul**

Two paradigms



'Classical' AI
Existing and explicit knowledge is coded or embedded in AI systems.



'New' AI
AI systems 'learn' by finding patterns in data with the help of structured feedback from humans.

Warning! We should be wary of any simple binary distinction. The notion of learning have been around for a long time (since the beginnings of AI in the 50's, see D. Cardon and A. Plasek work) and 'classical' AI also involves pattern recognition and adaption to user choices.

Le « Big Data » (mégadonnées) - 1

- ▶ Un concept qui existe depuis presque 30 ans et une expression de Mashey (1990)
 - ▶ Ce qui est trop grand pour être traité par les outils traditionnels
- ▶ Une définition large et peu précise scientifiquement
 - ▶ « Grosses données »
- ▶ Autres termes en français
 - ▶ Données massives
 - ▶ Mégadonnées

Le « Big Data » (mégadonnées) - 2

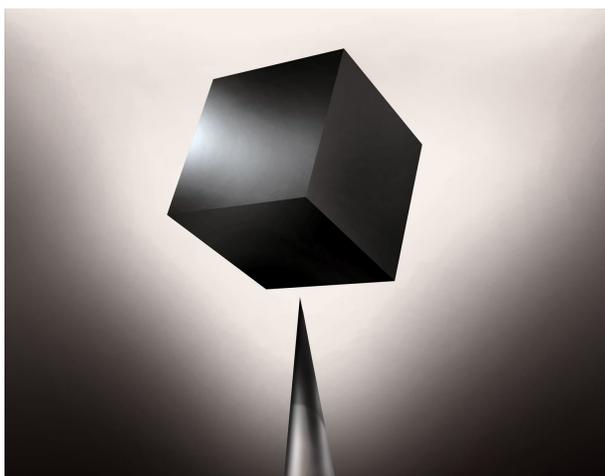
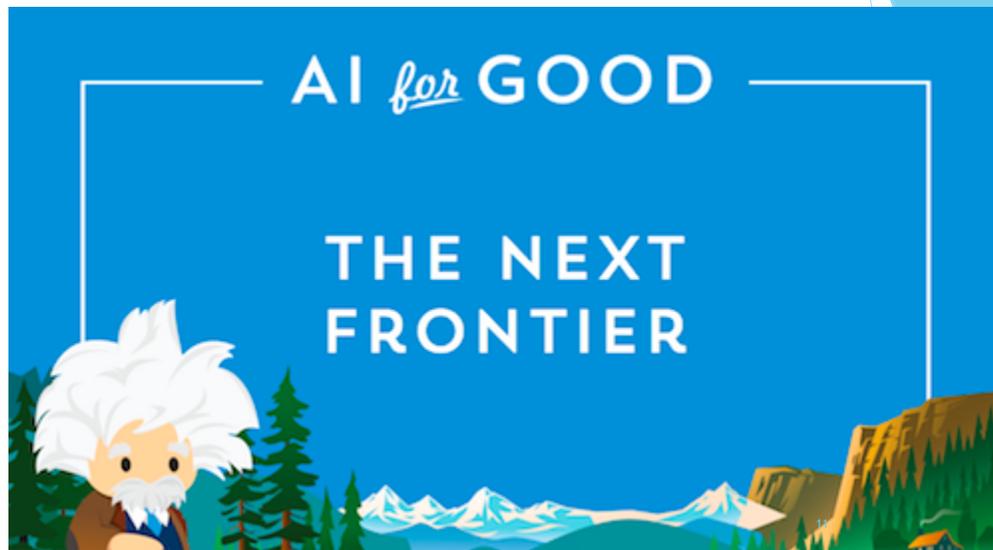
- ▶ Le terme désigne un phénomène vaste :
 - ▶ La multiplication actuelle des données et des moyens techniques de les analyser
- ▶ Consensus autour de *caractéristiques* (Kitchin, 2014)
 - ▶ Vitesse, volume, variété, résolution, vitesse...
- ▶ Le terme n'est pas tout à fait validé par la communauté technique mais repris par le grand public
 - ▶ Pas de critère strict de ce qui est du Big Data et ce qui n'en est pas.
Exemple : une base de données populationnelle vs. des Google Trends

Quels sont les enjeux éthiques importants en matière d'IA?

Three important changes

1. New AI Systems can accomplish tasks that we can't fully explain or define
(at least in a way that is both *formal* and *finite*).
E.g. Facial recognition
2. Performances *equal* to (or *better* than) humans
For narrow/specific tasks, especially in controlled environments.
3. No need for *human experts* to create ML systems
E.g. Detecting cancerous cells on X-rays without radiologists
E.g. Antidote software v. natural language processing with NN
E.g. Geoffrey Hinton and drug discovery (Harel and Radinsky 2018)

The Social Impacts of AI



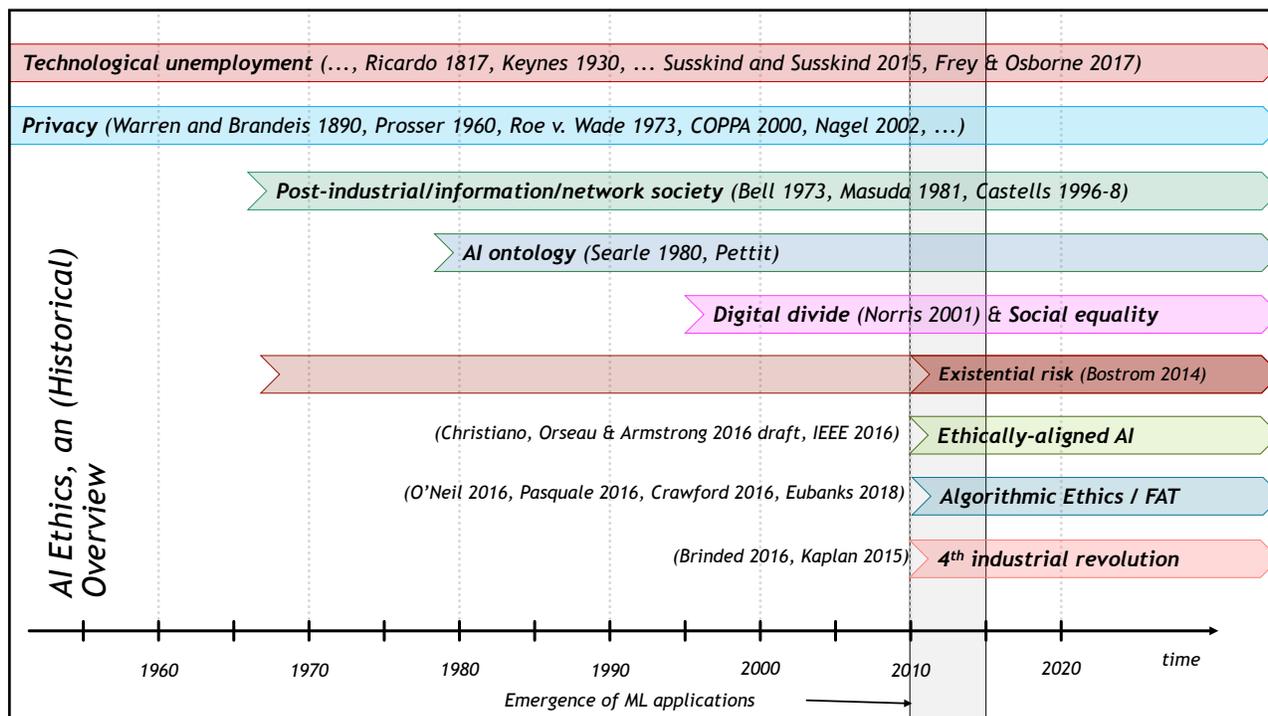
KEITH FRANKIN

Intelligent Machines

The Dark Secret at the Heart of AI

No one really knows how the most advanced algorithms do what they do. That could be a problem.

by Will Knight April 11, 2017



Automated surrogate decision making



AMA Journal of Ethics®

September 2018, Volume 20, Number 9: E902-910

VIEWPOINT: PEER-REVIEWED ARTICLE

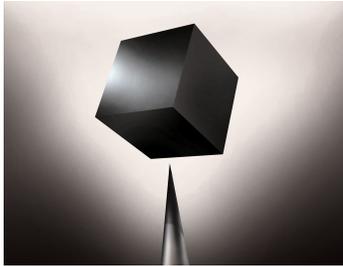
Should Artificial Intelligence Augment Medical Decision Making? The Case for an Autonomy Algorithm

Camillo Lamanna, MMathPhil, MBBS and Lauren Byrne, MBBS

Abstract: A significant proportion of elderly and psychiatric patients do not have the capacity to make health care decisions. We suggest that machine learning technologies could be harnessed to integrate data mined from electronic health records (EHRs) and social media in order to estimate the confidence of the prediction that a patient would consent to a given treatment. We call this process, which takes data about patients as input and derives a confidence estimate for a particular patient's predicted health care-related decision as an output, the autonomy algorithm. We suggest that the proposed algorithm would result in more accurate predictions than existing methods, which are resource intensive and consider only small patient cohorts. This algorithm could become a valuable tool in medical decision-making processes, augmenting the capacity of all people to make health care decisions in difficult situations.

Conclusions: In this essay, we have made the case that it should be possible to construct an autonomy algorithm to estimate confidence for predicted preferences of incapacitated patients by using machine learning technologies to analyze population-wide data sets, including EHRs and social media profiles. The proposed algorithm would result in more accurate predictions than existing methods, which are resource intensive and examine only small patient cohorts.

A Few Ethical Issues With AI/ML And Automation In The Healthcare Sector



- Human resources management
- Discrimination and biases (*There is no such thing as neutral data*)
- Transparency
- Explicability
- Accountability

L'IA en santé mentale - 1

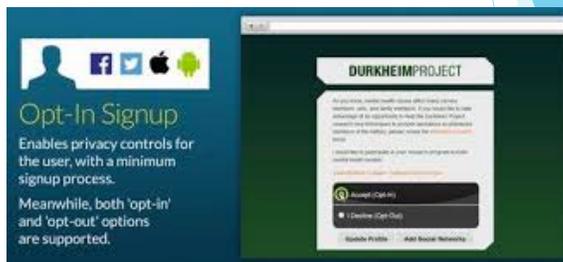
- ▶ Il existe plusieurs applications possibles en santé mentale (Luxton, 2016) :
 - ▶ L'aide à la décision ou au diagnostic
 - ▶ Le repérage
 - ▶ Les agents « affectifs » et conversationnels
 - ▶ Les outils behavioraux et thérapeutiques
 - ▶ La santé prédictive « big data »

L'IA en santé mentale - 2

- ▶ Voici des exemples du genre de recherches actuelles pour :
 - ▶ **Mieux comprendre des phénomènes cliniques**
Exemple : des études qui mêlent données biologiques et cliniques de patients schizophrènes
 - ▶ **Assister les cliniciens**
Exemple : signaler en temps réel des signes de risque suicidaire aux intervenants de ligne d'écoute (téléphone ou chat)
 - ▶ **Intervenir plus rapidement**
Exemple : les algorithmes de détection de messages à caractère suicidaire de Facebook
 - ▶ **Faire de « l'infodémiologie » (le suivi en temps réel de maladies à partir de données informatiques)**
Exemple : suivre la dépression sur Twitter

Exemples en prévention du suicide

- ▶ Le « *Durkheim Project* »
 - ▶ Projet du Veterans Affairs aux Etats-Unis
 - ▶ Repérage en temps presque réel du risque suicidaire chez les vétérans
 - ▶ Croisement de données cliniques et de réseaux sociaux



Exemples en prévention du suicide

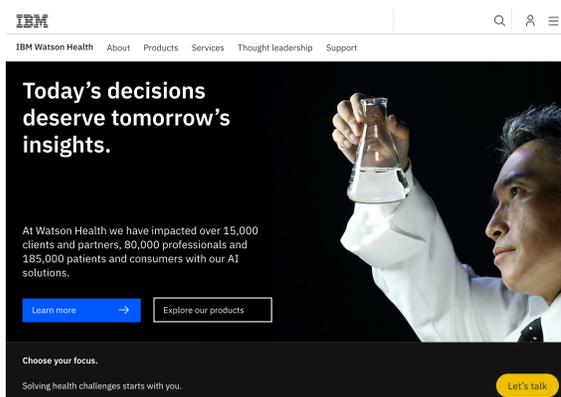
▶ Les chatbots



Enjeux éthiques de l'IA en santé mentale

- ▶ Le développement de l'Intelligence Artificielle a généré de nombreux problèmes scientifiques, juridiques et **éthiques** (Villani, 2017)
- ▶ **Il reste beaucoup de travail à faire** - au niveau des lois et des règles - pour répondre à ces enjeux (AI Now, 2017)

Promesses vs. craintes



IBM

IBM Watson Health About Products Services Thought leadership Support

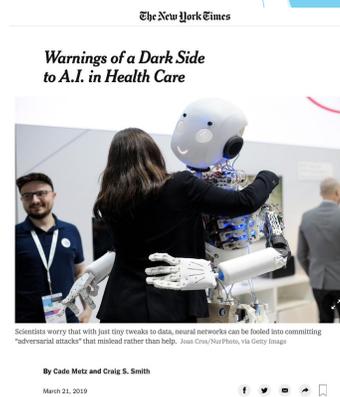
Today's decisions deserve tomorrow's insights.

At Watson Health we have impacted over 15,000 clients and partners, 80,000 professionals and 185,000 patients and consumers with our AI solutions.

[Learn more](#) → [Explore our products](#)

Choose your focus.
Solving health challenges starts with you. [Let's talk](#)

VS



The New York Times

Warnings of a Dark Side to A.I. in Health Care

Scientists worry that with just tiny tweaks to data, neural networks can be fooled into committing "adversarial attacks" that mislead rather than help. Joan Cres/NurPhoto, via Getty Image

By Cade Metz and Craig S. Smith
March 21, 2019

Quels principes éthiques appliquer?



- ▶ Plusieurs questions importantes ont émergé :
 - ▶ Quels sont les nouveaux enjeux éthiques soulevés par l'IA ?
 - ▶ Les nouveaux enjeux éthiques sont-ils couverts par les principes actuels en application ?
 - ▶ Doit-on adapter les principes éthiques « classiques » de la santé aux nouveaux défis soulevés par l'IA ?

Défis actuels - 1

- ▶ Il existe encore un manque de formation du côté de la santé sur les TIC. Il se crée de plus en plus des filières spécialisées. Mais les formations généralistes en santé mentale ont peu de contenu sur ces sujets.
 - ▶ Risque : face à la complexité technologique, le risque est d'adopter des technologies sans en comprendre les risques
- ▶ Enfin, l'IA ou le « Big Data » sont très discutés au niveau de leur impact social, mais encore peu voire pas par les acteurs de la santé mentale
 - ▶ Risque : défi d'utiliser des technologies assimilés de la surveillance
- ▶ Les récents scandales gouvernementaux (Snowden), corporatifs (Facebook) montrent qu'il est encore difficile d'avoir une éthique des données universelle et respectée.
 - ▶ Risque : Ceci pose de très grands défis pour un domaine dont le respect de la vie privée est au cœur des pratiques

Défis actuels - 2

- ▶ **Défi pour la recherche**
 - ▶ Les sciences humaines et la psychologie en particulier peuvent servir à développer des connaissances précieuses pour le profilage individuel
 - ▶ Risque : une utilisation à mauvais escient (ex. Cambridge Analytica)
 - ▶ En raison des coûts et des défis techniques, le milieu privé est souvent plus rapide et plus équipé que la recherche (publique en particulier)
- ▶ **Défi pour la pratique**
 - ▶ L'amélioration des stratégies de détection automatisée
 - ▶ Risque : les services de terrain risquent d'être dépassés (parfois inutilement). Ce qui risque de créer de l'épuisement
 - ▶ Risque 2 : les « fausses alarmes » sont aussi associées à de l'épuisement des ressources

Défis actuels - 3

▶ Défis sociaux

- ▶ On connaît l'existence du « digital divide », c'est-à-dire que les technologies de pointe ne sont pas utilisées par tous (en raison des inégalités sociales, d'accès aux TIC, etc.)
 - ▶ Risque : que les innovations technologiques en santé en profitent qu'aux personnes les plus fortunées.

▶ Défi de gouvernance et d'encadrement

- ▶ Les rapports nationaux (comme ceux de la Maison Blanche ou du gouvernement français) évoquent la difficulté de légiférer dans un contexte de compétition économique. Un encadrement légal trop strict risquerait de faire perdre la compétitivité.
 - ▶ Risque : le manque de cadre législatif proactif et exhaustif risque de renforcer les initiatives, applications ou utilisations non-éthiques de données liées à la santé mentale

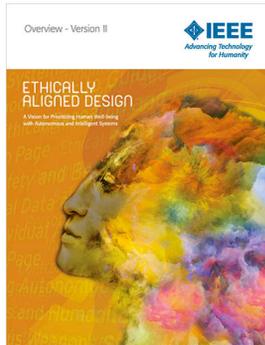
Comment renforcer la présence de l'éthique, dans le milieu des affaires, chez ceux qui développent l'IA, le ML en particulier?

- ▶ Reconnaître les enjeux!
- ▶ Protocoles, outils et déclarations
- ▶ Techniques de reddition de compte
 - ▶ Transparence du code (insuffisant)
 - ▶ Techniques exogènes
 - ▶ Techniques endogènes
- ▶ Trois questions
 1. Validité techniques (entraînement, type de réseau, etc.)?
 2. Adéquation avec les buts de notre organisation?
 3. D'autres problèmes éthiques?

Une proposition de protocole éthique

- ▶ Le Protocole Canadien (*Canada Protocol*) est une **initiative en accès libre**
- ▶ Son but est de contribuer à une meilleure **identification et prévention des enjeux éthiques en Intelligence Artificielle**
- ▶ Le projet vise à créer des **outils validés**, utilisant la littérature scientifique internationale
- ▶ Le premier outil diffusé est une **checklist éthique** à destination de tous ceux qui utiliseraient l'IA en santé mentale et prévention du suicide.

Une proposition de protocole éthique



- ▶ Des experts internationaux parlent désormais d'une « surcharge morale » chez les développeurs et promoteurs de l'IA (IEEE, 2017)
- ▶ C'est à dire : que de nombreux acteurs de l'IA se sentent dépassés par l'ampleur des enjeux, non formés pour y répondre, voire non concernés car c'est en dehors de leur champ de compétences.
- ▶ Ce constat existe au delà des sciences informatiques. Nous en avons fait le constat en santé mentale et prévention du suicide. L'IA y est de plus en plus utilisée, mais sans aucune mention d'enjeux (vie privée, sécurité...)

Une proposition de protocole éthique

- ▶ Comment rendre les acteurs plus sensibilisés aux enjeux éthiques liés à l'IA ?
En les aidant à les identifier les principaux défis reconnus dans les rapports internationaux en IA et éthique (Santé et TIC, etc.)
- ▶ Comment aider à prévenir les défis ou problèmes potentiels ?
En suggérant des pistes d'action, issues des recommandations d'expert
- ▶ Comment utiliser le potentiel de l'IA de façon prudente et mesurée ?
En prenant des décisions en pleine connaissance des risques encourus et des choix qu'il faut faire pour les limiter ou les éviter

contact@canadaprotocol.com

CANADA AI & ETHICS PROTOCOL

Homepage The Project Who Are We The Tools Our References FAQ

The Canada Protocol

Discover a scientific initiative to promote the ethical use of Artificial Intelligence. Whether you're in the design or the deployment phase of an intelligent system or just curious about AI & ethics, take a look at our project. We have different versions and more coming soon. Plus, it's all open-access!

LEARN MORE ABOUT THE PROJECT

DOWNLOAD THE FINAL VERSION OF THE AI & MENTAL HEALTH TOOL



But

- ▶ Le protocole a d'abord été créé pour répondre à un **manque identifié dans le champ spécifique de la santé mentale.**
- ▶ **Notre but** : faire penser aux risques et enjeux importants quand on veut utiliser l'IA pour faire de la prévention ou de l'intervention en Santé Mentale et pour prévenir le Suicide
- ▶ Nous avons donc créé une **checklist**, un outil fréquent en santé pour les diagnostics, les protocoles médicaux, de prévention ou de recherche (ex. PRISMA)



ETHICS + AI

CANADA PROTOCOL

Mental Health & Suicide Prevention version

2018

The Canada Protocol is an open access project designed to promote the ethical use of Artificial Intelligence (AI) + This version is a scientifically validated checklist focused on the challenges of using AI in the context of Mental Health Care or Suicide Prevention




Carl Mörch, M.Psy + Abhishek Gupta, B.A. + Brian L. Mishara, Ph.D.

1

Description

Delve into some of the main ethical challenges

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: https://ai-initiative.org/wp-content/uploads/2017/08/Making-the-AI-Revolution-Work-for-everyone-Report-to-OCED-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

Exemples

- ▶ **Funding** : 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
- ▶ **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")

2 Privacy & Transparency

Discover how respectful is your AIS

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	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

3 Security

Review some of the main technical risks and data-related issues

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

EXEMPLES

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

4 Health-Related Risks

Discover how respectful is your AIS

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 & contingency 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 behaviours and environments are considered as being at risk and explain the rationale in a simple way
Non-maleficence	Explain whether your Autonomous Intelligent System could harm, incommode, or embarrass 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

5 Biases

Prevent potential risks

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 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 issues	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)

EXEMPLES

- ▶ **Crisis and contingency 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
- ▶ **Ethics** : If you have requested an expertise on ethics during the design of your Autonomous Intelligent System, detail the parties involved and their contributions

Merci !

▶ Pour aller plus loin

- ▶ Luxton, D. D. (Ed.). (2015). *Artificial intelligence in behavioral and mental health care*. Academic Press.

▶ Pour nous contacter

- ▶ Dominic Martin martin.dominic@uqam.ca
- ▶ Carl Mörch cm.morch@gmail.com