

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**

MONTRÉAL
AI ETHICS
INSTITUTE

 Centre for Research and Intervention
on Suicide and Euthanasia



The Canada Protocol

AI is a source of immense hopes and valid concerns. It is challenging to know how to remain ethical. That is why we started the **Canada Protocol**. It is an open access project for AI and Big Data developers, decision-makers, professionals, researchers and anyone thinking about using AI.

We have synthesized and analyzed over 40 reports on AI & Ethics, Professional Guidelines, key studies.

Our intention is to gather all the existing scientific and validated recommendations on how to address AI's ethical risks and challenges. We hope this project might help you!

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

If you have recommendations, guidelines or reports on the ethical use of AI in your field, please let us know by getting in touch with us: contact@canadaprotocol.com



How to use this tool

This version of the Canada Protocol is a checklist. It invites you to review 38 key ethical questions when AI is used in the context of Mental Healthcare or Suicide Prevention

Whether you are in the design or the deployment phase, you are asked to read each item and thus review your practices and how your Autonomous Intelligent System (IEEE, 2016) works.

We integrated some recommendations and ideas for action.

Abbreviation used:

AIS: Autonomous Intelligent System

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: <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

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

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 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)

Canada Protocol

Mental Health & Suicide Prevention Checklist

Mörch, Gupta, Mishara + canadaprotocol.com + 2018

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 & 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
Non-maleficience	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

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)

Follow the Canada Protocol

If you have recommendations, guidelines or reports on the ethical use of AI in your field, please let us know by getting in touch with us



contact@canadaprotocol.com



canadaprotocol.com



[@canadaprotocol](https://twitter.com/canadaprotocol)

