

Algorithmic Impact Assessment Results

Version: 0.10.0

Project Details

1. Name of Respondent

Sgt Tom Guineau

2. Job Title

Technical Manager

3. Department

Royal Canadian Mounted Police

4. Branch

NHQ/Tech Ops/SSIS/NCECC/Tech Unit

5. Project Title

Griffeye

6. Project ID from IT Plan

Griffeye CS and Di Pro

7. Departmental Program (from Department Results Framework)

Griffeye CS and Di Pro

8. Project Phase

Implementation

[Points: 0]

9. Please provide a project description:

SOFTWARE already in Use - Tool designed to assist in the categorization/ classification of Child Exploitation images and Videos. It ingests media derived from Seized Exhibits, Submissions from partners (NCMEC, CyberTip and other Law Enforcement Agencies et al) and presents all of the media and associated embedded meta data (EXIF etc) into a workspace where Categories per Sec 163 of the Criminal Code can be applied, media can be grouped together using various filters, and pre-categorization performed against hash databases of known material. This tool does offer the ability to use an AI implementation to evaluate images and videos and determine whether Child Pornography is present by applying trained Classifiers to group media together for human review. This AI function is also able to Face Match (not Face Recognize) against other media contained within the workspace to attempt to group individuals with similar characteristics together. The inbuilt Face Matching feature can be used and does not reach out to other systems in order to resolve a face to a specific identifiable individual. Once work is complete, the Hashes and associated meta data can be exported to share out to other compatible tools using the PVIC Format, and shared as appropriate/ authorized.

About The System

10. Please check which of the following capabilities apply to your system.

Image and object recognition: Analyzing very large data sets to automate the recognition, classification, and context associated with an image or object.

Content generation: Analyzing large data sets to categorize, process, triage, personalize, and serve specific content for specific contexts

Risk assessment: Analyzing very large data sets to identify patterns and recommend courses of action and in some cases trigger specific actions

Section 1: Impact Level : 2

Current Score: 39

Raw Impact Score: 39

Mitigation Score: 16

Section 2: Requirements Specific to Impact Level 2

Peer review

Consult at least one of the following experts and publish the complete review or a plain language summary of the findings on a Government of Canada website:

- qualified expert from a federal, provincial, territorial or municipal government institution
- qualified members of faculty of a post-secondary institution
- qualified researchers from a relevant non-governmental organization
- contracted third-party vendor with a relevant specialization
- a data and automation advisory board specified by Treasury Board of Canada Secretariat.

OR

Publish specifications of the automated decision system in a peer-reviewed journal. Where access to the published review is restricted, ensure that a plain language summary of the findings is openly available.

Gender-based Analysis Plus

Ensure that the Gender-based Analysis Plus addresses the following issues:

- impacts of the automation project (including the system, data and decision) on gender and/or other identity factors;
- planned or existing measures to address risks identified through the Gender-based Analysis Plus.

Notice

Plain language notice posted through all service delivery channels in use (Internet, in person, mail or telephone).

Human-in-the-loop for decisions

Decisions may be rendered without direct human involvement.

Explanation

In addition to any applicable legal requirement, ensure that a meaningful explanation is provided to the client with any decision that results in the denial of a benefit or service, or involves a regulatory action. The explanation must inform the client in plain language of:

- the role of the system in the decision-making process;
- the training and client data, their source, and method of collection, as applicable;
- the criteria used to evaluate client data and the operations applied to process it;
- the output produced by the system and any relevant information needed to interpret it in the context of the administrative decision; and
- a justification of the administrative decision, including the principal factors that led to it.

Explanations must also inform clients of relevant recourse options, where appropriate.

A general description of these elements must also be made available through the Algorithmic Impact Assessment and discoverable via a departmental website.

Training

Documentation on the design and functionality of the system.

IT and business continuity management

None

Approval for the system to operate

None

Other requirements

The Directive on Automated Decision-Making also includes other requirements that must be met for all impact levels.

[Link to the Directive on Automated Decision-Making](#)

Contact your institution's ATIP office to discuss the requirement for a Privacy Impact Assessment as per the Directive on Privacy Impact Assessment.

Section 3: Questions and Answers

Section 3.1: Impact Questions and Answers

Reasons for Automation

1. What is motivating your team to introduce automation into this decision-making process?
(Check all that apply)

Improve overall quality of decisions

The system is performing tasks that humans could not accomplish in a reasonable period of time

Use innovative approaches

Existing backlog of work or cases
Lower transaction costs of an existing program
Other (please specify)

2. Please describe

With respect to Child Pornography and sexual abuse videos and images, human review and decision making is always required to quickly identify, locate and safeguard children being victimized. With the average single case containing in excess of 3 million media items to review, such tools are employed in order to identify which media has been reviewed or investigated previously, and then group 'new media' together for the human review and more quickly determine police response/rescue. By removing the need to review the exact same horrific media over and over again not only adds efficiency to safeguarding children, it lends itself to protecting the investigator's health and well-being through unnecessary exposure.

3. What client needs will the system address and how will this system meet them? If possible, describe how client needs have been identified.

There are several clients involved; primarily the victimized and abused children need to be identified, located and safeguarded from the abusive situation regardless of location in the world as soon as possible. If the scant law enforcement resources spend their time reviewing non-relevant or non-urgent materials it would delay discovery of the abuse for months or years, leaving the victim to be further abused despite the organization having timely information to save them in hand. Such systems aid in prioritizing media for human review in order to determine the likelihood of a child being currently at risk.

4. Please describe any public benefits the system is expected to have.

Further to above, the public has an expectation that law enforcement will act on information and safeguard people in a timely manner. If 911 call would take weeks to garner a response, public safety would be in jeopardy as would trust in the ability to provide security of the person. Such systems that aid in organizing data in a manner in which vulnerable persons are more quickly offered aid is a no-brainer.

5. How effective will the system likely be in meeting client needs?

Very effective

[Points: +0]

6. Please describe any improvements, benefits, or advantages you expect from using an automated system. This could include relevant program indicators and performance targets.

Griffeye Analyze is one of very few niche tools available that addresses categorizing media related to Child Pornography investigations. The benefits of using such tools are:

- more quickly identify, investigate and locate/safeguard children at risk
- ease of evidence and intelligence sharing with external and internal partners, as these tools utilize a standardized format in order to export/import data
- reduces psychological injury and grief to investigators by reducing the amount of trauma they are continuously exposed to
- ability to provide prosecutors with factual data as to the severity and scope of media collections by offenders
- be able to provide statistical data to government to convey the scope of abuse and the associate police response

7. Please describe how you will ensure that the system is confined to addressing the client needs identified above.

Griffeye Analyze is designed for media categorization specifically for this purpose (as are other competitive niche solutions), and as such is purpose built with these needs in mind.

8. Please describe any trade-offs between client interests and program objectives that you have considered during the design of the project.

None

9. Have alternative non-automated processes been considered?

Yes

[Points: +0]

10. If non-automated processes were considered, why was automation identified as the preferred option?

Historically, investigators would physically print off images of abuse originating from floppy disk drives and other digital storage, which resulted in few media items requiring human review and evaluation. With the advent of mass digital storage and the availability of high speed internet, the production, transmission and collection of videos and images are near instant and often result in a single investigation containing on average in excess of 2-3 million media items to be reviewed (numerous computers, cellphones, tablets and other common devices with large storage and connectivity can be found in everyone's residences). The need to human review and evaluate each video or image cannot be sustained by one person if they were to review each abuse image, each Windows logo or Amazon listing over and over again if they have already been seen in the past. Without digital processing and analysis of media that can identify previously seen content or the ability to group media together with various commonalities (same camera, same computer user, same website et al), an investigator would take a month just to evaluate and risk assess one case, let alone all of the reports that are submitted on a daily basis.

11. What would be the consequence of not deploying the system?

Service cannot be delivered in a timely or efficient manner

[Points: +2]

Service costs are too high

[Points: 0]

Service quality is not as high

[Points: 0]

Service delivery cannot achieve performance targets

[Points: 0]

Other (please specify)

[Points: 0]

12. Please describe

Children will continue to be filmed while being sexually assaulted or tortured due to delays in evaluating and prioritizing media under investigation.

Risk Profile

13. Is the project within an area of intense public scrutiny (e.g. because of privacy concerns) and/or frequent litigation?

No

[Points: +0]

14. Are clients in this line of business particularly vulnerable?

Yes

[Points: +3]

15. Are stakes of the decisions very high?

Yes

[Points: +4]

16. Will this project have major impacts on staff, either in terms of their numbers or their roles?

Yes

[Points: +3]

17. Will the use of the system create or exacerbate barriers for persons with disabilities?

No

[Points: +0]

Project Authority

18. Will you require new policy authority for this project?

No

[Points: +0]

About the Algorithm

19. The algorithm used will be a (trade) secret

No

[Points: +0]

20. The algorithmic process will be difficult to interpret or to explain

No

[Points: +0]

About the Decision

21. Please describe the decision(s) that will be automated.

The tool has the ability to group media together in a number of ways that are not AI driven (group by exif, file path, binary hash matching et al) but can also have the feature of leveraging additional modules of ML to group images/ videos together if they contain children, various sexual assault acts or levels of nudity. The decisions are to group media together and flag them to the investigator in order to prioritize their manual human review and evaluation - so effectively, the final decision always resides with the person operating the tool. The AI is to assist in more quickly locating media that contains elements of children and sexual abuse among masses of unrelated media such as internet web page media, operating system images and other more legitimate media.

22. Does the decision pertain to any of the categories below (check all that apply):

Other (please specify)

[Points: +1]

23. Please describe

The decisions more quickly identify Child Pornography media content within mass digital storage, in order to prioritize their human review.

Impact Assessment

24. Which of the following best describes the type of automation you are planning?

Partial automation (the system will contribute to administrative decision-making by supporting an officer through assessments, recommendations, intermediate decisions, or other outputs)

[Points: +2]

25. Please describe the role of the system in the decision-making process.

Please refer to previous information

26. Will the system be making decisions or assessments that require judgement or discretion?

No [Points: +0]

27. Please describe the criteria used to evaluate client data and the operations applied to process it.

The system (when using any AI feature) attempts to examine the content of images or videos to determine if certain visual or behavioral elements exist within (ie - is there a child present? Is there an adult? is there a sexual act occurring? is there genitalia present et al). Media containing selected elements is flagged and brought to the operator's attention as the prudent starting point to determine if Child Pornography is present within that set of media items.

28. Please describe the output produced by the system and any relevant information needed to interpret it in the context of the administrative decision.

The output results in media being flagged as a priority for human review, where an investigator can then watch the video or look at the image to ascertain and confirm whether the content/context therein is Child Pornography as defined by the Criminal Code, and thus decide if there is current risk to a person/persons requiring immediate intervention

29. Will the system perform an assessment or other operation that would not otherwise be completed by a human?

No [Points: +0]

30. Is the system used by a different part of the organization than the ones who developed it?

No [Points: +0]

31. Are the impacts resulting from the decision reversible?

Reversible [Points: +1]

32. How long will impacts from the decision last?

Impacts are most likely to be brief [Points: +1]

33. Please describe why the impacts resulting from the decision are as per selected option above.

The impact is that the AI is not the final decision as it rests with the human operator. The AI functions as an aid and not an authoritative role within the investigative process.

34. The impacts that the decision will have on the rights or freedoms of individuals will likely be:

Little to no impact [Points: +1]

35. Please describe why the impacts resulting from the decision are as per selected option above.

The AI functions as an aid and not an authoritative role within the investigative process.

36. The impacts that the decision will have on the equality, dignity, privacy, and autonomy of individuals will likely be:

Little to no impact [Points: +1]

37. Please describe why the impacts resulting from the decision are as per selected option

above.

The AI functions as an aid and not an authoritative role within the investigative process.

38. The impacts that the decision will have on the health and well-being of individuals will likely be:

Little to no impact

[Points: +1]

39. Please describe why the impacts resulting from the decision are as per selected option above.

The AI functions as an aid and not an authoritative role within the investigative process. However, the use of the tool will reduce the time in safeguarding children being sexually assaulted or tortured, which in turn is to the overall benefit to the victim if rescue were to occur as sooner rather than later. As well, a reduction in the repetitive traumatic exposure by investigators to such media will benefit their well being.

40. The impacts that the decision will have on the economic interests of individuals will likely be:

Little to no impact

[Points: +1]

41. Please describe why the impacts resulting from the decision are as per selected option above.

The AI functions as an aid and not an authoritative role within the investigative process.

42. The impacts that the decision will have on the ongoing sustainability of an environmental ecosystem, will likely be:

Little to no impact

[Points: +1]

43. Please describe why the impacts resulting from the decision are as per selected option above.

The AI functions as an aid and not an authoritative role within the investigative process.

About the Data - A. Data Source

44. Will the Automated Decision System use personal information as input data?

No

[Points: +0]

45. What is the highest security classification of the input data used by the system? (Select one)

Protected B / Protected C

[Points: +3]

46. Who controls the data?

Federal government

[Points: +1]

47. Will the system use data from multiple different sources?

Yes

[Points: +4]

48. Will the system require input data from an Internet- or telephony-connected device? (e.g. Internet of Things, sensor)

No

[Points: +0]

49. Will the system interface with other IT systems?

Yes

[Points: +4]

50. Who collected the data used for training the system?

A foreign government or non-government third party

[Points: +4]

51. Who collected the input data used by the system?

Your institution

[Points: +1]

52. Please describe the input data collected and used by the system, its source, and method of collection.

Input data (Images and Videos) consists of evidence submitted by Police, NGOs and other industry as a function of reporting a criminal offence. It can come by way of police seizure (Judicial Authorizations) or submitted via third party (CyberTip, NCMEC, Industry Reporting). Some is collected forensically, some electronically by the party that received it or as it traverses their IT systems and flagged for reporting to police. Police agencies receive such media either directly from victims/complainants or via assistance requests from other LE agencies

About the Data - B. Type of Data

53. Will the system require the analysis of unstructured data to render a recommendation or a decision?

No

[Points: 0]

Section 3.2: Mitigation Questions and Answers Consultations

1. Internal Stakeholders (federal institutions, including the federal public service)

Yes

[Points: +1]

2. Which Internal Stakeholders have you engaged?

Strategic Policy and Planning

Data Governance

Legal Services

Access to Information and Privacy Office

Program Policy

3. External Stakeholders (groups in other sectors or jurisdictions)

Yes

[Points: +1]

4. Which External Stakeholders have you engaged?

Academia

Industry

Governments in other Jurisdictions

International Organizations

De-Risking and Mitigation Measures - Data Quality

5. Do you have documented processes in place to test datasets against biases and other unexpected outcomes? This could include experience in applying frameworks, methods,

guidelines or other assessment tools.

No [Points: +0]

6. Is this information publicly available?

No [Points: +0]

7. Have you developed a process to document how data quality issues were resolved during the design process?

No [Points: +0]

8. Is this information publicly available?

No [Points: +0]

9. Have you undertaken a Gender Based Analysis Plus of the data?

No [Points: +0]

10. Is this information publicly available?

No [Points: +0]

11. Have you assigned accountability in your institution for the design, development, maintenance, and improvement of the system?

Yes [Points: +2]

12. Do you have a documented process to manage the risk that outdated or unreliable data is used to make an automated decision?

No [Points: +0]

13. Is this information publicly available?

No [Points: +0]

14. Is the data used for this system posted on the Open Government Portal?

No [Points: +0]

De-Risking and Mitigation Measures - Procedural Fairness

15. Does the audit trail identify the authority or delegated authority identified in legislation?

No [Points: +0]

16. Does the system provide an audit trail that records all the recommendations or decisions made by the system?

Yes [Points: +2]

17. Are all key decision points identifiable in the audit trail?

No [Points: +0]

18. Are all key decision points within the automated system's logic linked to the relevant legislation, policy or procedures?

No [Points: +0]

19. Do you maintain a current and up to date log detailing all of the changes made to the model and the system?

- No [Points: +0]
20. Does the system's audit trail indicate all of the decision points made by the system?
No [Points: +0]
21. Can the audit trail generated by the system be used to help generate a notification of the decision (including a statement of reasons or other notifications) where required?
No [Points: +0]
22. Does the audit trail identify precisely which version of the system was used for each decision it supports?
No [Points: +0]
23. Does the audit trail show who an authorized decision-maker is?
Yes [Points: +1]
24. Is the system able to produce reasons for its decisions or recommendations when required?
No [Points: +0]
25. Is there a process in place to grant, monitor, and revoke access permission to the system?
Yes [Points: +1]
26. Is there a mechanism to capture feedback by users of the system?
Yes [Points: +1]
27. Is there a recourse process established for clients that wish to challenge the decision?
No [Points: +0]
28. Does the system enable human override of system decisions?
Yes [Points: +2]
29. Is there a process in place to log the instances when overrides were performed?
Yes [Points: +1]
30. Does the system's audit trail include change control processes to record modifications to the system's operation or performance?
No [Points: +0]
31. Have you prepared a concept case to the Government of Canada Enterprise Architecture Review Board?
No [Points: +0]

De-Risking and Mitigation Measures - Privacy

32. If your system uses or creates personal information, have you undertaken a Privacy Impact Assessment, or updated an existing one?
Yes [Points: +1]
33. Please indicate the following in your answer: Title and scope of the Privacy Impact Assessment; How the automation project fits into the program; and Date of Privacy Impact Assessment completion or modification.
PIA is collateral to FVEYE Hash Sharing Agreement reviewed by RCMP Legal, ATIP, DSB, LSU and other internal stakeholders which defines personal

information as it pertains to this data type. Documentation regarding the use of hashes and derivative meta data have been produced for collateral issues and can be provided upon request.

34. Have you designed and built security and privacy into your systems from the concept stage of the project?

Yes

[Points: +1]

35. Is the information used within a closed system (i.e. no connections to the Internet, Intranet or any other system)?

Yes

[Points: +1]

36. If the sharing of personal information is involved, has an agreement or arrangement with appropriate safeguards been established?

Yes

[Points: +1]

37. Will you de-identify any personal information used or created by the system at any point in the lifecycle?

No

[Points: +0]