Information Accountability Foundation

Assessments in an AI World

Risk/Data Protection Assessment as Required by U.S. State Privacy Laws

Lynn Goldstein and Peter Cullen

August 2024

IAF Data Protection Assessment to Meet State Laws

Report

Background

Data protection assessments required by US state call for balancing of risks and benefits for all stakeholders' but don't tell how to balance them.

The Information Accountability Foundation (IAF) for many years has advocated for a multi-dimensional stakeholder interests, risks, and benefits balancing assessment. This expanded balancing assessment makes sure all interests are reflected in today's complex data use scenarios. The development of these assessments began with the big data revolution. The Mayer-Schönberger and Cukier 2013 book Big Data raised the issue of big data governance, and the IAF believed to establish big data governance, and in particular the impacting use of this data, a common ethical frame based on key values and an assessment of interests and risks framework was needed. Working with academics, civil society, regulators, and global organizations using big data, the IAF in 2015 developed the Big Data Assessment Framework. In 2016 the Office of the Privacy Commissioner of Canada gave the IAF a grant to develop a Canadian Version of Big Data Assessment, in 2018 the Office of the Privacy Commissioner for Personal Data funded the IAF's drafting of an Ethical Accountability Framework for Hong Kong China, and in 2021 the IAF published a Model Algorithmic Impact Assessment (Model AIA). Each of these research projects advanced and matured multi-dimensional assessments.

A common theme throughout all these iterations of the IAF's advanced data analytics assessments is multi-dimensional balancing – the assessing of the interests of multiple stakeholders and a broad set of benefits and risks. As a result, the IAF published in 2022 a briefing paper on multi-dimensional proportionality, A Principled Approach to Rights and Interest Balancing (Balancing Briefing Paper).

The IAF learned, as a result of the Balancing Briefing Paper and the related Model Artificial Intelligence Assessment (AIA), that a methodology is needed to, in an orderly and repeatable fashion, identify and demonstrate three components of multi-dimensional balancing:

- **The individual rights and benefits** (in a fundamental rights-based system) or established legitimate interests (in legal systems when fundamental rights are not established),
- **The full range of stakeholders whose rights or interests are involved**, and
- **The adverse processing impacts to all stakeholders that may be involved, their likelihood and level of consequence, recognizing that adverse processing impacts sometimes only may be reduced and not eliminated.**

To create that robust process, the IAF developed a directory of rights, interests, stakeholders, and consequences, and used colors and symbols and a mathematic model to represent those factors. All these factors are then balanced – multi-dimensional balancing – and the results can be demonstrated in a narrative form and can be supported pictorially, mathematically, or both. This multi-dimensional balancing provides form and substance in a world that demands both protection and advanced probabilistic programs.

With the requirement of a data protection assessment under the many new U.S. state privacy laws, which explicitly require this type of multi-dimensional balancing, the next obvious iteration of the IAF's assessments is the Demonstrable Assessments to Meet U.S. State Privacy Laws (U.S. Assessment). The requirement in the two of the U.S. state privacy laws to produce assessments to regulators on demand or provide them at least annually will trigger requests to show or demonstrate new accountability requirements. To meet these laws' explicit and implicit requirements, the IAF believes organizations will have to adopt new governance processes, which includes substantive guidelines, policies, and procedures. The U.S. Assessment expands upon the requirements of the laws and regulations of the U.S. state privacy laws. To develop the U.S. Assessment, the IAF created the Demonstrable Accountability Project (the Project).

The Project

Since the Global Accountability Dialogue was first established in 2009, the components of what it means for an organization to be accountable as part of data protection requirements, privacy and data protection regulators have called for "demonstrable accountability."[1] The growing use of data and data driven technologies, such as Artificial Intelligence (AI) and large language models (LLMs), have increased the expectations of policymakers and regulators for more demonstrable accountability processes.

The expectation for demonstrable accountability is found in the assessment requirements set forth in many of the new state

[1] In an October 2019 Blog, the Global Privacy Assembly (GPA) said that under the General Data Protection Regulation (GDPR), an accountable organization must be able to *demonstrate* its compliance.

privacy laws in the United States[2] to weigh benefits to all stakeholders against the potential risks to the rights and interests of consumers associated with data processing activity, such as profiling and model development.[3] New requirements to produce assessments to regulators on demand (e.g., Colorado) or provide them at least annually (e.g., California) could trigger requests to show or demonstrate new accountability requirements.

Yet today there is no common standard or regulatory expectation as to what these new demonstrable accountability requirements as part of risk governance processes should consist of and what form an assessment should take. Some of the unknown factors include how assessment oversight, model bias and discrimination determination, and internal control requirements should be structured. This lack of clarity creates uncertainty for businesses who wish to increase their use of data as part of their strategies and inconsistencies in various regulatory approaches (i.e., regulators may step in without the involvement of business to set standards).

The foundation for a solution set is a better understanding of business challenges **and** regulatory expectations. This understanding served as input to the development of a normative framework consisting of described demonstrable process and procedures. This content was developed further though a multi-stakeholder engagement session.

IAF's multi-dimensional balancing was combined with the IAF's extensive track record of developing frameworks and assessments that include associated and requisite governance and controls. Through the multi-stakeholder convening process, the Project delivered the U.S. Assessment, a draft normative framework consisting of an assessment with governance controls as well as a suggested demonstrability framework that:

- expands business governance capability and sustainability,
- increases regulator confidence,
- enables further guidance to business by regulators.

The resulting U.S. Assessment benefits business and regulators who, with confidence from the multi-stakeholder process, can incorporate them into horizontal governance strategies, processes, and procedures within their business organizations and into provided regulatory guidance.

The Project Process

[2] See State-Assessment-Provisions-v1.0-002.pdf (secureserver.net)

[3] Similar assessments soon could be mandated by new laws governing the use of AI and the associated fair implications to people, e.g., EU Proposed Artificial Intelligence Regulation and Canada's Bill C-27 (specifically the Artificial Intelligence and Data Act (AIDA) part of C-27).

- The IAF developed a draft solution framework addressing key problem areas.
    - The IAF incorporated relevant U.S. state laws and rules and draft regulations and other developing models as part of a draft solution model.
    - This stage was supported by individual dialogue with select business participants and regulatory authorities.
- The draft framework was reviewed first through a convening meeting with participating business organizations.
- Once the solution framework was finalized, the IAF convened a multi-stakeholder session that engaged the business, regulatory, and NGO communities. The framework was presented and discussed, and suggestions were made to fine-tune the framework.

What was Learned from Business and Regulator Meetings

Interviews with Business

The most common thing heard from business is that it wants to comply not only with the letter but with the spirit of the U.S. state privacy laws. To do so, business wants certainty, but it is concerned about rules and regulations being drafted in a vacuum. In other words, it is concerned that regulators will draft rules and regulations without any understanding of business realities. Business appreciates the comment period provided once there is an initial draft of a rule or regulation but is concerned that too much can be "baked in" at that stage and therefore thought the earlier regulators could hear about how business operates, such as at a multi-stakeholder session or during informal visits to the regulator, could be helpful.

In a similar vein, business wants to get to an industry standard or a best practice. The ultimate question business wants answered is: If we had an investigation and/or are required to provide an assessment to a regulator, and we presented "this," would it be enough? Other businesses put it this way: We want the decoder ring. We want to know how our program measures up against X.

Business' most common solution to these concerns is through use cases, i.e., the development of a regulator approved completed U.S. Assessment for a particular type of high-risk processing. Suggested use cases were AI model development, AI model training, removal of bias from AI models, anti-money laundering, direct marketing and advertising, and fraud.

Finally, in developing the U.S. Assessment, business encouraged the IAF to identify which parts are required and which are not and to find the delta between the various laws that require balancing, including the Balancing Section of the Fundamental Rights Impact Assessment required under the EU AI Act.

Interviews of Former Regulators

Benefits to consumers should not be articulated in an assessment at a very superficial level. They should not be described in just a few lines; there needs to be a narrative which ideally is evidence based. An example of this evidence is conducting external research to understand the interests of consumers.

All the former regulators said they could tell the difference between an assessment that had been prepared by a lawyer (often after the fact) and not by the business before the initiation of high-risk processing. The assessment needs to be completed by the most knowledgeable people within the business and needs to be approved by a person or persons outside the part of the business proposing the high-risk processing with the appropriate authority to approve.

The U.S. Assessment

The Assessment Requirements in the U.S. State Laws

Almost all new U.S. state privacy laws require various forms and timing of privacy risk assessments. The laws in California, Colorado and New Jersey direct the issuance of rules/regulations by the appropriate agency regarding the content of these assessments. Colorado has promulgated rules, and California has issued draft regulations (collectively Regulations). California calls this assessment a “Risk Assessment,” and the other states requiring an assessment call this assessment a “Data Protection Assessment.” For the purposes of clarity, the requirements contained in the state privacy laws and the Regulations are referred to as a Risk and Data Protection Assessment (RDPA). For reasons discussed below, the IAF thinks an expanded assessment is necessary which adds governance and ethical requirements, and this expanded assessment developed by the IAF, attached as an Addendum, is referred to as the U.S. Assessment. As discussed below, the U.S. Assessment makes clear which parts are required by the U.S. state privacy laws and by the Regulations and which parts are not so required.

The state privacy laws require that a RDPA identify and weigh the benefits that may flow to the Controller that presents a heightened risk of harm to a consumer (California uses the term “business,” but for sake of consistency the U.S. Assessment uses the term Controller) resulting from the processing of personal information (California uses the term “personal information,” and most state laws use the term “personal data,” and for the sake of consistency the U.S. Assessment uses the term “personal information”). In general, the state laws require RDPAs to identify and weigh the benefits to the Controller, the consumer, other stakeholders, and the public resulting from the processing of personal data that presents a heightened risk of harm to a consumer against the potential rights of the consumer associated with the processing as mitigated by safeguards that can be employed by the Controller to reduce these risks. These laws go on to require that the RDPA factor in the use of de-identified data, reasonable expectations of consumers, the context of the processing and the relationship between the Controller and the consumer whose personal data will be processed.

The state privacy laws identify the external stakeholders (the Controller, the consumer, consumers in general, the public, and other stakeholders). To help in the identification of internal stakeholders, the IAF had added examples in Appendix I of the U.S. Assessment.

The Requirements in the Regulations

The Regulations require that the RDPA at a minimum include, and the U.S. Assessment includes, the following:

- A short summary of the Processing
- The categories of personal information to be Processed
- The context of the Processing
- The consumers’ reasonable expectations
- The operational elements of the Processing
- The purpose of the Processing
- The benefits resulting from the Processing to the Controller, the consumer, the consumers in general, the public, and other stakeholders
- The sources and nature of risks to the rights of consumers associated with the Processing
- Safeguards the Controller will implement to reduce these risks
- A description of how the identified benefits of the Processing outweigh the identified risks as mitigated by the identified safeguards

They also require the following if the Processing executes a decision or facilitates human decision-making (ADMT) (ADMT includes Profiling which is any form of automated processing of personal information to evaluate, provide, or deny financial or lending services, housing, insurance, education enrolment or opportunity, criminal justice, employment opportunities, health-care services, or access to essential goods or services and which presents a reasonably foreseeable risk of unfair or deceptive treatment of, or unlawful disparate impact on consumers, financial or physical injury to consumers, a physical or other intrusion upon the solitude or seclusion, or private affairs or concerns of consumers if the intrusion would be offensive to a reasonable person, or other substantial injury to consumers):

- A plain language explanation of:
    - Why the Controller is using or seeks to use the ADMT to achieve the purpose of the Processing, including the benefits of using automated processing over manual processing, the appropriate uses of the ADMT, and any limitations on the appropriate uses of the ADMT
    - The personal information processed by the ADMT, including the personal information that were or will be used in the ADMT and the sources of the personal information
    - Why the ADMT directly and reasonably relates to the Controller's goods and services
    - The training data and logic, including any assumptions, used to create the ADMT
    - The outputs secured from the ADMT and how the Controller will use the outputs, including any decisions to be made using the ADMT and especially whether they will be used to make a decision to provide or deny or substantially contribute to the provision or denial of financial or lending services, housing, insurance, education, enrollment or opportunity, criminal justice, employment opportunities, health-care services or access to essential goods or services
    - The steps the Controller has taken or any steps it plans to take to maintain the quality of personal information processed by the ADMT
    - How the Controller evaluates its use of the ADMT for validity, reliability, fairness, and disparate impact and the results of any such evaluations
- If there is human involvement in the ADMT, the degree and details of any human involvement
- Safeguards for any data sets produced by or derived from the ADMT
- If the ADMT is conducted by third-party software purchased by the Controller, the name of the software and copies of any internal or external evaluations sufficient to show the accuracy and reliability of the software

The IAF's Multi-Dimensional Balancing

The U.S. privacy laws and the Regulations require the balancing of multiple stakeholders, benefits, and risks but do net tell how to conduct this balancing. The IAF has written on this subject for many years, and most recently the IAF has developed an impact analysis which takes into account all of the stakeholders and weighs the benefits/interests against the risks/harms and effectiveness of mitigations using a 1,3, 5 scale which aligns with many typical risk management approaches. IAF's multi-dimensional weighing is unique. It factors in as many stakeholders, benefits, and risks as are relevant to the processing being assessed. It is capable of weighing each of the stakeholders vis-à-vis each of the other factors. It can demonstrate the results mathematically or pictorially or both. It can be used to supplement the required narrative response. Examples of how to conduct multi-dimensional balancing are found in Appendix III of the U.S. Assessment.

The Regulations and most risk taxonomies do not address risks to consumers in general and to the public. The IAF believes these types of risk are key and has added some examples to consider from the Ethical OS, the Omidyar Ethical Framework for Tech, and Business Data Ethics in Appendix III of the U.S. Assessment. The IAF also thinks the list of risks to individual consumers is incomplete and has added risks that come from its model legislation, the FAIR and OPEN USE ACT, which is based upon input from IAF strategists and membership, in Appendix II of the U.S. Assessment as well.

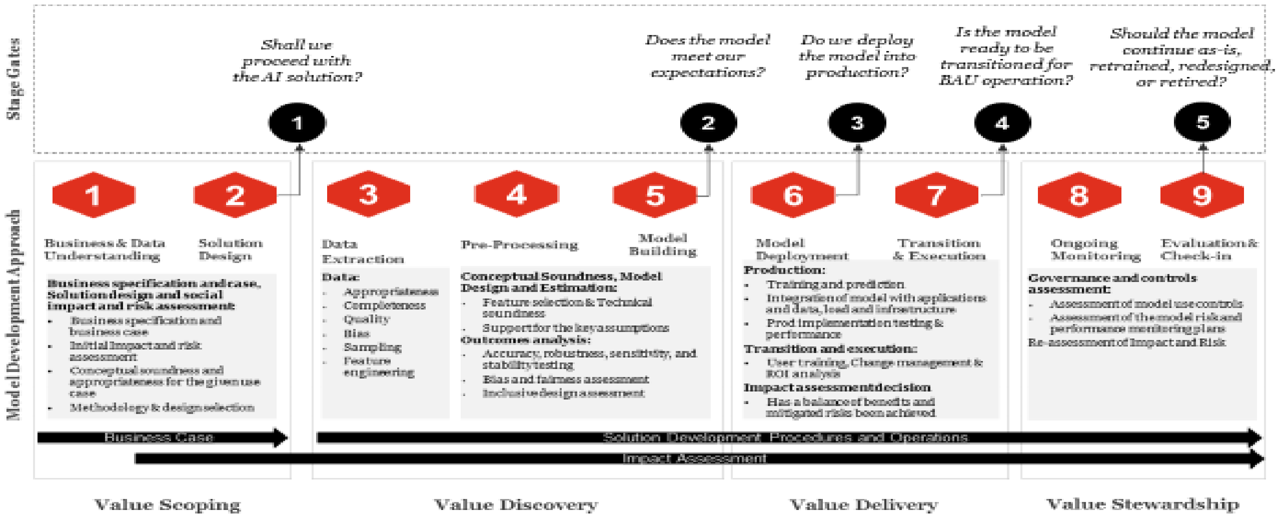
Additional Governance Requirements

To keep up with the risks/impacts and benefits/interests to organizations, consumers, consumers in general, and the public brought about by advanced analytics, in the IAF's view, the RDPA required by the State Privacy Laws needs to be expanded to address additional governance controls. These general controls are included in the U.S. Assessment for illustrative purposes, as they help answer the "who, what, and how" questions for a regulator. Therefore, the U.S. Assessment addresses this need and refers to the Model AIA for a fulsome approach to assessing these types of issues.

The actions to be taken required by the Regulations are demanding; most of them are not mechanical and call for more than a translation of the Regulations into a compliance checklist. To meet the Regulations' explicit and implicit requirements, the IAF believes organizations also will have to adopt new governance processes, which includes substantive guidelines, policies, and procedures.

As AI grows a part of business objectives and strategies, organizations will need to expand their own risk assessment and management processes. For example, an assessment process should cover the full development lifecycle requirements from strategy and planning, model development, the specific issues related to training data, deployment, ongoing operation and monitoring issues, and governance. These requirements were outlined in the Model AIA which included an example of a five stage gate review to successful AI Governance.[4] Each stage gate review is designed to address specific questions and should involve a broad set of stakeholders and decisionmakers:

[4] Modified on original Six stage gates to a successful AI governance | by Anand Rao | Towards Data Science


Stage Gates
Shall we proceed with the AI solution?
Does the model meet our expectations?
Do we deploy the model into production?
Is the model ready to be transitioned for BAU operation?
Should the model continue as-is, retrained, redesigned, or retired?
1
2
3
4
5
Model Development Approach
1
2
3
4
5
6
7
8
9
Business & Data Understanding
Solution Design
Business specification and case, Solution design and social impact and risk assessment:
• Business specification and business case
• Initial Impact and risk assessment
• Conceptual soundness and appropriateness for the given use case
• Methodology & design selection
Data Extraction
Data:
- Appropriateness
- Completeness
- Quality
- Bias
- Sampling
- Feature engineering
Pre-Processing
Model Building
Conceptual Soundness, Model Design and Estimation:
- Feature selection & Technical soundness
- Support for the key assumptions
Outcomes analysis:
- Accuracy, robustness, sensitivity, and stability testing
- Bias and fairness assessment
- Inclusive design assessment
Model Deployment
Transition & Execution
Production:
• Training and prediction
• Integration of model with applications and data, load and infrastructure
• Prod implementation testing & performance
Transition and execution:
• User training, Change management & ROI analysis
Impact assessment decision
• Has a balance of benefits and mitigated risks been achieved
Ongoing Monitoring
Evaluation & Check-in
Governance and controls assessment:
- Assessment of model use controls
- Assessment of the model risk and performance monitoring plans
Re-assessment of Impact and Risk
Business Case
Solution Development, Procedures and Operations
Impact Assessment
Value Scoping
Value Discovery
Value Delivery
Value Stewardship

By extension, a RDPA should encompass a review at key parts of a processing development lifecycle. In effect, the assessment process required by a RDPA should be iterative and matched to the cadence of an organization's development process.

The RDPA required by the Regulations is used only when High Risk Processing is conducted. Because of the IAF's additions to the RDPA, the U.S. Assessment can assess AI where AI goes beyond ADMT. It increasingly seems likely these High-Risk Processing scenarios also will involve AI, thus supporting their inclusion in the U.S. Assessment.

Summary of the U.S. Assessment

Which parts of the U.S. Assessment (see Addendum) are required, and which parts are not required, are summarized as follows:

- Parts II – XII, XVIA&C-D – XIX, XXI, XXII and Appendices I and IV (colored in yellow) are taken directly from the Colorado Privacy Act and the Rules implementing the Colorado Privacy Act (Colorado Rules) which can be found here, the Colorado AI Act, and the CPPA Draft Risk Assessment Regulations (California Draft Regulations) which can be found here.
- Parts XIII – XV, XX, XXII and Appendices II (3 of 21 risks to consumers) and III (colored in green) are taken from:
  - IAF's Model AIA which was developed in consultation with AI experts and practitioners (Responsible AI) privacy and data protection professionals and based on IAF's 10+ years of experience in developing big data and complex

analytics assessments such as to be used with AI.
    - IAF's Balancing Briefing Paper, and
    - IAF's model legislation, the FAIR and OPEN USE ACT based upon input from IAF strategists and membership.
- Part XVIIB and Appendix II (except of 3 of 21 risks to consumers) (colored in pink) are taken from the Ethical OS, the Omidyar Ethical Framework for Tech, and from Business Data Ethics.

This framework for the U.S. Assessment should be viewed as a document that would be made available to a regulator to demonstrate the requirements set out in the Regulations and the Colorado AI Act. Each organization likely will operationalize the requirements in its own way and use these operational components as a source to complete the U.S. Assessment.

What was Learned in the Multi-Stakeholder Session

Participants at the Multi-Stakeholder Session were from large U.S. businesses who also are global companies, academics, and present and former regulators. Four themes were heard at the Multi-Stakeholder Session:

- The Need to Focus on Stakeholders

One attendee made the point that it is important not to gloss over identifying the stakeholders, that it is hard to do it well, but that identification of stakeholders well can really blow up the size of the assessment. Another attendee made the point that different parts of the business should contribute to the identification of the stakeholders.
IAF Response: Additional questions have been added to the Stakeholder Section of the U.S. Assessment, and additional resource material has been added to Appendix II to help identify internal stakeholders (the external stakeholders are specified in the U.S. state privacy laws and the Regulations).

- The Need for Balancing to be Objective

Concern was expressed by the several attendees with putting numbers on the balancing because numbers can give the impression that there is objectivity that does not exist. Some of the businesses in attendance replied that it is necessary that the U.S. Assessment be able to evaluate risk in the same way as other risk assessments across the organization. The consensus was the 1, 3, 5 scale which aligns with many typical risk management approaches should be used.
IAF Response: The 1,3,5 logic used in the U.S. Assessment is the 1, 3, 5 scale in many typical risk management approaches and consistent with many organizations' Enterprise Risk Management programs. It is also an optional add to the required "narrative" component of the U.S. Assessment.

- The Need for Consistent Level of Detail in the Sections

One attendee pointed out that some sections of the U.S. Assessment provided more detailed guidance than other sections. In particular, as discussed above, the Stakeholder Section of the U.S. Assessment needed to provide more direction in identifying stakeholders.
IAF Response: More detailed guidance provided in the Appendices has been moved into the main Sections of the U.S. Assessment.

- Specific Additional Content

Several attendees had recommendations for specific additions to the RDPA. They are a laundry list of very discrete inclusions:

- Consult more recent ethical resources, such as Human Rights or Ethical Frameworks
- Include Experts as Stakeholders
- Include Red Teaming and Postmortems as part of Model Development
- Specify the development of policies, procedures, and guidelines
- Require a record/ inventory of Go/No Go decisions as precedent of previous decisions and guidance for future decisions
- Require keeping track of versions of the U.S. Assessment so there is a record of changes made
- Define “plain language”

IAF Response: All these changes were made.

Going Forward

There are two possible separate work streams going forward:

- Socialization of the U.S. Assessment

Simultaneous with the publication of this report and the U.S, Assessment on the IAF website, the IAF will write a blog and post on social media. The IAF also will approach the IAPP about an article in the IAPP Daily Dashboard (and similar publications).

- Potential New Projects

As additional requirements are added by new U.S. state privacy laws and regulations, the U.S. Assessment could be updated. The current version of the U.S. Assessment was developed over a period of several months with input from business and academics as part of a Multi-Stakeholder Session. The IAF hopes this dialogue will expand to include regulators so the makeup and format of the U.S. Assessment will help advance any ensuing guidance and will enable business to meet the requirements of these new and future U.S. state privacy laws. It was noted that U.S. states' future enactment of AI laws may impact the scope and content of the U.S. Assessment.

Also, attendees at the Multi-Stakeholder Session thought use cases on, for example, AI model development, training, bias removal, anti-money laundering, fraud, and direct marketing and advertising would be helpful. Also, another attendee at the Multi-Stakeholder Session observed that the U.S. Assessment addresses predictive AI and does not address generative AI (because the U.S. privacy laws address predictive AI).

APPENDIX

U.S. Assessment

I. Introduction

Almost all new U.S. state privacy laws require various forms and timing of privacy risk assessments. While California calls this a "Risk Assessment," most other states call it a "Data Protection Assessment." For the purposes of clarity, this document will refer to these requirements as a Risk and Data Protection Assessment (RDPA). This framework for a RDPA is organized into XXII parts. Parts II – XII, XVIA&C – D - XIX, XXI – XXII and Appendices I and IV (colored in yellow) are taken directly from the Colorado Privacy Act and the Rules implementing the Colorado Privacy Act (Colorado Rules) which can be found here, the Colorado AI Act, and the CPPA Draft Risk Assessment Regulations (California Draft Regulations) which can be found here. The Colorado Rules and the California Draft Regulations are collectively referred to as Regulations. Parts XIII – XV, XX, XXII and Appendices II (3 of 21 risks to consumers) and III (colored in green) are taken from:

- IAF's AI Assessment which was developed in consultation with AI experts and practitioners (Responsible AI) privacy and data protection professionals and based on IAF's 10+ years of experience in developing big data and complex analytics assessments such as to be used with AI, and which can be found here,
- IAF's A Principled Approach to Rights and Interest Balancing (Balancing Briefing Paper) which can be found here, and
- IAF's model legislation, the FAIR and OPEN USE ACT based upon input from IAF strategists and membership and which can be found here.

Part XVIIB and Appendix II (except 3 of 21 risks to consumers) (colored in pink) is taken from Ethical OS, the Omidyar Ethical Framework for Tech, and from Business Data Ethics. This framework for a U.S. Assessment should be viewed as a document that would be made available to a regulator to demonstrate the requirements set out in the Regulations and the Colorado AI Act (see Appendix III). Each organization likely will operationalize the requirements in its own way and use these operational components as a source to complete a U.S. Assessment.

The state privacy laws require that a RDPA identify and weigh the benefits that may flow to the Controller

(California uses the term business, but for sake of consistency this document will use the term Controller), consumer, other stakeholders, and the public, against the potential risks to the rights of the affected consumer or consumers in general as mitigated by safeguards employed to reduce those risks. The weighing required by these state privacy laws is different from the weighing required by privacy laws anywhere else in the world. For example, the EU General Data Protection Regulation (GDPR) data protection impact assessments (DPIAs) only call for an assessment of the risks to the rights and freedoms of natural persons. The GDPR DPIA is a tool for managing risks to the rights of data subjects in high-risk data processing scenarios and typically is enforced against a narrower set of data protection rights; other fields manage the risks to society and organizations. EDBP Guidelines, p.17.

The Regulations go even further than other privacy laws in collectively setting forth 15 risks to consumers that may be considered, six of which are in the IAF's FAIR and OPEN USE ACT, identifying mitigating measures and the timing of when RDPAs should be conducted, and collectively require a RDPA for processing that presents a heightened risk of harm (High Risk Processing). High Risk Processing includes:

- Selling or sharing personal information (Colorado used the term personal data, and California uses the term personal information. This document uses the term personal information).

- Processing sensitive information.

- Using Automated Decision-Making Technology (ADMT) (California uses the term ADMT which includes profiling, the Colorado Rules use the term Profiling, and the Colorado AI Act uses the term High-Risk Artificial Intelligence System; for inclusiveness, this document uses the term ADMT), in furtherance of a decision that results in the provision or denial of financial of financial or lending services, housing, insurance, education enrollment or education opportunity, criminal justice, legal services, employment or employment opportunity, contracting opportunities or compensation, healthcare services, or access to essential goods, services, or opportunities, including essential government services.

- Processing the information of consumers that the Controller has actual knowledge are less than 16 years of age.

- Processing the personal information of consumers who are employees, independent contractors, job applicants, or students using technology to monitor employees, independent contractors, job applicants, or students.

- Processing the personal information of consumers in publicly accessible places using technology to monitor consumers' behavior, location, movements, or actions.
- Processing the personal information of consumers to train artificial intelligence or ADMT.
- Processing of personal information for the purposes of targeted advertising.
- Processing of personal information for the purposes of ADMT where such ADMT presents a reasonably foreseeable risk of:
  - unfair or deceptive treatment of, or unlawful disparate impact on, consumers
  - financial, physical or reputation injury to consumers
  - physical or other intrusion upon the solitude or seclusion, or the private affairs or concerns, of consumers, where such intrusion would be offensive to a reasonable person.

The Regulations and most risk taxonomies do not address risks to society or large groups of individuals. The IAF believes these types of risk are key and has added some examples from the Ethical OS and Business Data Ethics to consider in the risk portion of the U.S. Assessment.

The actions to be taken required by the Regulations are demanding; most of them are not mechanical and call for more than a translation of the Regulations into a compliance checklist. To meet the Regulations' explicit and implicit requirements, we believe organizations will have to adopt new governance processes, which includes substantive guidelines, policies, and procedures (see XIII-XV). Required actions call for open ended multi-dimensional weighing, with the expectation multiple internal stakeholders will be involved. In addition to supplementing the requirements of the Regulations with IAF's prior work, the IAF has developed an impact analysis which takes into account all of the stakeholders and weighs the benefits/interests against the risks/harms using a 1,3, 5 scale which aligns with many typical risk management approaches.

IAF's multi-dimensional weighing is unique. It factors in as many stakeholders, benefits, and risks as are relevant to the processing being assessed. It is capable of weighing each of the stakeholders vis-à-vis each of the other factors. It can demonstrate the results mathematically or pictorially or both. It can be used to supplement the required narrative response.

The RDPA required by the Regulations is used only when High Risk Processing is conducted. Because of the IAF's additions to the RDPA, the U.S. Assessment can assess Artificial Intelligence (AI) where AI goes beyond ADMT. It increasingly seems likely these High-Risk Processing scenarios also will involve AI, thus supporting their inclusion in the U.S. Assessment. Getting the U.S. Assessment right is good for business. The weighing of the risks and benefits to the numerous stakeholders will be worthwhile only if it done competently and with integrity. The U.S. Assessment enables business to weigh the risks and benefits of AI competently and with integrity.

II. Timing of the RDPA

- The Controller must conduct and document the RDPA before initiating High Risk Processing.
- The Controller must review and update the RDPA as often as appropriate considering the type, amount, and sensitivity of personal information Processed and the level of risk presented by the Processing, throughout the Processing activity's lifecycle in order to: 1) monitor for harm caused by the Processing and adjust safeguards accordingly; and (2) ensure that data protection and privacy are considered as the Controller makes new decisions with respect to the Processing.
- RDPAs for Processing that uses ADMT subject to ADMT access and opt-out rights must be reviewed and updated at least [annually/biannually/once every three years] and include an updated evaluation for fairness and disparate impact and the results of any such evaluation.
- A new data Processing activity is generated when existing Processing activities are modified in a way that materially changes the level of risk presented. When a new data Processing activity is generated, a RDPA must reflect changes to the pre-existing activity and additional considerations and safeguards to offset the new risk level. (See Appendix 1 for a list of examples of modifications which may materially change the level of risk of a Processing activity.) A change in the Processing activity is material if it diminishes the benefits of the Processing activity, creates negative impacts, or increases the magnitude or likelihood of already-identified negative impacts, or diminishes the effectiveness of safeguards.

III. Retention of the RDPA

RDPAs, including prior versions which have been revised when a new data Processing activity is generated (even as a result of a RDPA) or to account for a material change, must be stored for as long as the Processing activity continues, and for at least five (5) years after the completion of the RDPA or conclusion of the Processing activity, whichever is later. Versioning (keeping different versions of RDPAs) should be employed. R DPAs must be held in an electronic, transferable form. In Colorado, data processing assessments are required for activities created or generated after July 1, 2023; this requirement is not retroactive. In California, for High-Risk Processing that the business initiated prior to the effective date of the California Draft Regulations and continues after the effective date of the California Draft Regulations, the Controller shall conduct a risk assessment within 24 months of the effective date of the California Draft Regulations. Under the Colorado AI Act, a deployer must maintain the most recently completed assessment, all records concerning each assessment, and all prior assessments, if any, for at least three years following the final deployment of the assessment.

IV. Triggers for a RDPA and other Obligations

A Controller shall conduct and document the RDPA before initiating a Processing activity that constitutes High Risk Processing. Under the Colorado AI Act, a deployer, or a third-party contracted by the deployer, must review the deployment of each ADMT deployed by the deployer to ensure that the ADMT is not causing algorithmic discrimination. In California, organizations will be required to submit their risk assessments to the California Privacy Protection Agency. In Colorado, the data protection assessments and impact assessments must be provided to the Attorney General upon request.

V. Triggers for a RDPA for ADMT

If a Controller is using ADMT for processing that will be subject to ADMT access and opt-out rights, then the RDPA must include the ADMT RDPA requirements. See Appendix IV.

VI. Scope	A. A RDPA must be a genuine, thoughtful analysis of each personal information Processing activity that presents High Risk Processing that: 1) identifies and describes the risks to the rights of consumers associated with the Processing; 2) documents measures considered and taken to address and offset those risks; 3) contemplates the benefits of the Processing; and 4) demonstrates that the benefits of the Processing outweigh the risks offset by safeguards (controls) in place. B. The depth, level of detail, and scope of the RDPA should take into account the scope of the risk presented, the size of the Controller, the amount and sensitivity of personal information Processed, the personal information Processing activities subject to the RDPA, and complexity of safeguards applied. C. A "comparable set of Processing operations" that can be addressed by a single RDPA is a set of similar Processing operations including similar activities that present heightened risks of similar harm to a consumer.
VII. Purpose (summary business need/goal/ objective for this personal information use scenario)	A description with specificity why the Controller needs to conduct the Processing and how the Processing achieves that purpose. A short summary of the Processing activity that presents significant risk to consumers' privacy and the context of the Processing activity, including the relationship between the Controller and the consumers whose Personal Data will be Processed, the consumers' reasonable expectations concerning the purpose for processing their personal information or the purpose's incompatibility with the context in which their personal information was collected, and how the Controller will process the personal information (including how the Controller will collect, use, disclose, and retain personal information). The Colorado AI Act requires a statement by the developer disclosing the purpose, intended use cases, and deployment context afforded by the ADMT.

VIII. Benefits and Risks	A short summary of the benefits and risks resulting from the Processing that may flow, directly and indirectly, to the Controller, consumer, other expected stakeholders, and the public. The Colorado AI Act requires a statement by the deployer disclosing the benefits afforded by the ADMT and an analysis of whether the deployment of the ADMT poses any known or reasonably foreseeable risks of algorithmic discrimination, and if so, the nature of the algorithmic discrimination and the steps that have been taken to mitigate the risks. These benefits and risks shall be identified and described with specificity in section XVII.
IX. Personal Information Involved (is any sensitive information included)	The categories of personal information (information that identifies, relates to, describes, is linked or reasonably linkable, directly or indirectly, to an identified or identifiable consumer or household and does not include de-identified data or publicly available information) to be processed and whether they include sensitive personal information, including personal information from a known Child.
X. Nature and Operational Elements of the Processing Activity	In determining the level of detail and specificity to provide, the Controller must consider the type, amount, and sensitivity of personal information Processed, the impacts that operational elements will have on the level of risk presented by the Processing activity, and any relevant unique relationships. Relevant operational details may include: a. Sources of personal information and method of collecting, using, disclosing, storing, retaining, and sharing; b. Technology or Processors to be used; c. Names or categories of personal information recipients, including Third Parties, Affiliates, and Processors that will have access to the personal information, the processing purpose for which the personal information will be provided to those recipients, and categorical processes that the Controller uses to evaluate that type of recipient; d. Operational details about the Processing, including planned processes for personal information collection, use, disclosure, storage, retention, and sharing; e. Specific types of personal information to be processed; f. How the Controller's processing of personal information complies with data minimization requirements (including why the Controller needs to process the personal information and the relevance of the personal information to the Processing); g. How long the Controller will retain each category of personal information and why the Controller needs to retain each category for that length of time; h. The approximate number of consumers whose personal information the Controller plans to process and the context of the relationship with the consumer.

XI. Sensitive Personal Information	Sensitive personal information means: • Personal information revealing racial or ethnic origin, religious beliefs, a mental or physical health condition or diagnosis, sex life or sexual orientation, or citizenship or citizenship status • Genetic or biometric data that may be processed for the purpose of uniquely identifying a consumer or • Personal information from a known child. The details of the process implemented to ensure that sensitive personal information and sensitive personal information inferences are not transferred and are deleted within 12 hours of the Processing activity. Examples are certain government identifiers (e.g., social security numbers); an account log-in; financial account; debit card or credit card number with any required security code, password, or credentials allowing access to an account; precise geolocation; contents of mail, email, and text messages.

XII. ADMT including Profiling	If a Controller is using ADMT for Processing that will be subject to ADMT access and opt-out rights, a RDPA of that Processing activity must include the following: 1. A plain language explanation of the specific types of personal information that were or will be processed by the ADMT, including the personal information used to train the ADMT and the sources of the personal information (The Colorado AI Act requires a deployer, or a third-party contracted by the deployer (collectively deployer), to include a description of the categories of data the ADMT processes as inputs and outputs the ADMT produces) 2. The decision to be made using ADMT 3. The benefits of automated processing over manual processing for the stated purpose 4. A plain language explanation of why ADMT including Profiling is being used to achieve the purpose of the Processing, the appropriate use(s) of the ADMT, and any limitation on the appropriate use(s) of the ADMT 5. A plain language explanation of the training data and logic of the ADMT, including any assumptions of the logic and any statistics used in the analysis, either created by the Controller or provided by a Third Party which created the applicable ADMT 6. If the ADMT is conducted by the Third-Party software purchased by the Controller, the name of the software and copies of any internal or external evaluations sufficient to show the accuracy and reliability of the software where relevant to the risks described below 7. A plain language explanation of the output(s) secured from the ADMT 8. A plain language description of how the output(s) from the ADMT are or will be used, including whether and how they are used to make a decision to provide or deny or substantially contribute to the provision or denial of financial or lending services, housing, insurance, education enrollment or education opportunity, criminal justice, legal service, employment or employment opportunity, health-care services, or access to essential goods or services, including essential government service 9. A plain language explanation of the steps taken or any steps to be taken to maintain the quality of personal information processed by the ADMT, including personal information used to train the ADMT 10. If there is human involvement in the use of ADMT, a plain language explanation of the degree and details of any human involvement 11. A plain language explanation of how the ADMT is evaluated for validity, reliability, fairness and disparate impact, and the results of any such evaluation 12. A plain language explanation of any safeguards that the Controller plans to implement

	to address the negative impacts to consumers' privacy that are specific to its use of ADMT or for data sets produced by or derived from the ADMT (The Colorado AI Act requires a deployer to include a description of any transparency measures taken concerning the ADMT, including any measures taken to disclose to a consumer that the ADMT is in use when the ADMT is in use and of the post-deployment monitoring and user safeguards provided concerning the ADMT, including the oversight, use, and learning process established by the deployer to address issues arising from the deployment of the ADMT)) 13. If personal information has been processed or is being processed to train AI or ADMT, and that AI or ADMT has been or is being made available to other persons for their own use, how a plain language explanation of the appropriate purposes for which the persons may use the AI or ADMT has been provided or will be provided to those persons and any safeguards that have been implemented or will be implemented to ensure that the AI or ADMT is used for appropriate purposes by other persons (the Colorado AI Act requires a deployer to include an overview of the categories of data the deployer used to customize the ADMT if the deployer used data to customize the ADMT) 14. If personal information has been processed or is being processed to train AI or ADMT and is made or is being made available to other businesses (recipient business), how the necessary facts have been or will be provided to those recipient businesses to conduct the recipient businesses' risk assessments 15. Following an intentional and substantial modification to an ADMT, the Colorado AI Act requires the deployer to include a statement disclosing the extent to which the ADMT was used in a manner that was consistent with, or varied from, the developer's intended uses of the ADMT .
XIII. Other Governance Requirement Byproducts	Over and above the requirements in the RDPA relating to ADMT including Profiling, since the Regulations refer to terms/issues that are not defined, such as "Test for fairness, negative inferences, adversarial use/attacks, unconscionable treatment, creation of anxiety, embarrassment, fear or other mental trauma, over collection, surveillance or use beyond what a reasonable consumer would expect," the following topics also should be considered. These topics lead to the need to create governance/control processes and questions that would appear in the functional part of a privacy impact assessment (PIA). Question themes found in an Artificial Intelligence Assessment (AIA), which includes ADMT/Profiling, are set forth below. Examples of questions in an AIA can be found here.

XIIIa Lifecycle	The project plan should account for each stage of the Model lifecycle: Plan and Design Model, Collect and Process Data, Build and Use Model, Verify and Validate Model, Deploy and Use Model, Operate and Monitor Model, Assess Impacts of Model
XIIIb Fairness	The term "fairness" has been described, and steps are in place to measure and test for achieving fairness.
XIIIc Traceability	Traceability can be maintained across data, experiments, model versions and usage. Performance can be captured against success criteria.
XIIId Training	The quality of training data has been assessed. Were there enough total training samples? The samples were representative of different social groups based on – race, gender, color, age, income, etc.
XIIIe Model Testing	The performance of the model was tested. The model was well-trained and analyzed through different metrics – Precision, Recall, F1Score, Accuracy, Bias, Robustness and Sensitivity training, Red Teaming, Postmortems, etc.
XIIIf Equal treatment	All users are treated equally. If not – and your algorithms and predictive technologies prioritize certain information or sets prices or access differently for different users – describe how you would handle consumer/user demands or government regulations or contractual requirements that require all users be treated equally, or at least transparently unequally. The criteria for conducting Bias Audits under the Final Rule implementing New York City Local Law 144 may be helpful in determining whether a decision or other action is discriminatory.
XIIIg Third Parties	If your organization obtained models or datasets from a third party, describe how the risks of using third parties are assessed and managed and what the documentation requirements of using the third parties are.
XIIIh End user	Describe how end-users or other subjects are made aware adequately that a decision, content, advice, or outcome is the result of an algorithmic decision?
XIV. Other Governance Requirements	In addition to the RDPA for ADMT/Profiling, the Regulations also contain requirements regarding transparency (including the model logic), opting out, and consent. Describe the additional governance processes put in place to address these additional requirements, such as substantive guidelines, policies, and procedures.
XV. Governance Controls	Describe the governance controls in place that will enable a consistent, robust, repeatable

	development/implementation process, such as substantive guidelines, policies, and procedures. Describe how all project team members' roles have been established and communicated.
XVI. Impact Analysis A. The RDPA must involve all relevant internal actors from across the Controller's organizational structure and, where appropriate, relevant external parties, to identify, assess and address the data protection benefits and risks to consumers.	Describe how the benefits of the Processing outweigh the risks as mitigated by the safeguards. This description includes the sources and nature of risks to the rights of consumers associated with the Processing activity posed by the Processing activity. The source and nature of the risks may differ based on the Processing activity and type of personal information processed. The negative impacts to consumers' privacy associated with the Processing, including the sources of these negative impacts, must be identified and the magnitude of the negative impacts and the likelihood of the negative impacts occurring must be described with specificity, including how the criteria used to determine the magnitude and likelihood of the negative impact. Use the specific Stakeholder Detailed Analysis to identify a full range of benefits and risks to key stakeholders. See Appendix II for examples of stakeholders and of benefits and risks.
B. Risks to Society/ Large Groups of Individuals	As in XVI(A), describe how the benefits of the Processing outweigh the risks to society and large groups of individuals as mitigated by the safeguards. See Appendix II for examples of risks to society and large groups of individuals.

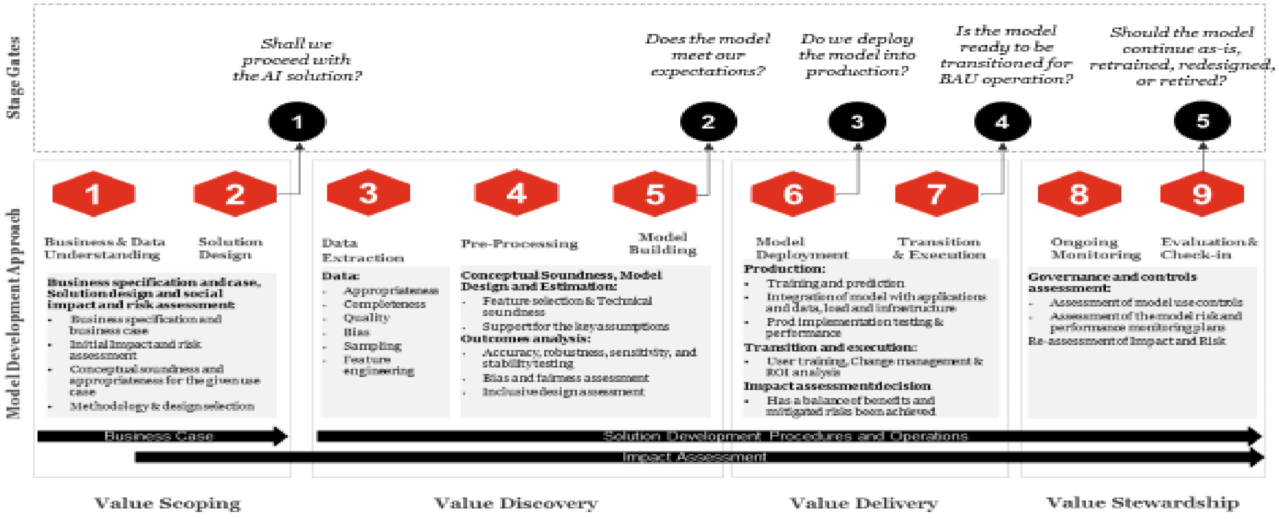
C. Stakeholders (Internal or external person or group of people or organizations or society or segments of a population who are involved in or who can affect or be affected by the Processing) – Detailed Analysis	Benefits/Interests	Risks/Harms
Consumer		
Consumers in General		
Public		
Controller		
Other		
D. Safeguards	Identify the safeguards that the Controller plans to implement to address the negative impacts identified in XVIA&B above and explain how these safeguards address these negative impacts with specificity, including whether and how they eliminate or reduce the magnitude of the negative impacts or the likelihood of the negative impacts occurring; and any safeguards the business will implement to maintain knowledge of emergent risks and countermeasures. See Appendix II for examples of safeguards.	

XVII. Additional Mitigators	See Appendix II for examples of additional mitigators.
XVIII. Residual Risk	After assessing the risk (risk = impact x likelihood) and applying any safeguards/mitigators, assess the residual risk remaining to the consumer's privacy after these safeguards/mitigators are implemented and what these residual risks are.
XIX. Summary of Weighted Balance (Include Heat Maps or other "Weighing" criteria)	A description of how the benefits of the Processing outweigh the risks identified, as mitigated by the safeguards identified. RDPAS MUST IDENTIFY AND **WEIGH** THE BENEFITS THAT MAY FLOW, DIRECTLY AND INDIRECTLY, FROM THE PROCESSING TO THE CONTROLLER, THE CONSUMER, OTHER STAKEHOLDERS, AND THE PUBLIC AGAINST THE POTENTIAL RISKS TO THE RIGHTS OF THE CONSUMER ASSOCIATED WITH THE PROCESSING, AS MITIGATED BY SAFEGUARDS THAT THE CONTROLLER CAN EMPLOY TO REDUCE THE RISKS. THE CONTROLLER MUST FACTOR INTO THIS RDPA THE USE OF DE-IDENTIFIED DATA AND THE REASONABLE EXPECTATIONS OF CONSUMERS, AS WELL AS THE CONTEXT OF THE PROCESSING AND THE RELATIONSHIP BETWEEN THE CONTROLLER AND THE CONSUMER WHOSE PERSONAL DATA WILL BE PROCESSED. This narrative summary could be just a "narrative" or could also include some pictorial/numerical examples of the **Multi-Dimensional/Stakeholder Balancing (See Appendix III for examples of this type of balancing)**
XX. Decision- Go/No-Go (Approver)	How effective are the mechanisms that facilitate the Processing activity auditability (e.g., traceability of the development process, the sourcing of training data and the logging of the Processing system's processes, outcomes, positive and negative impact)? Could the AI system be audited by independent third parties? What are the additional requirements surrounding the Processing activity? Are there other legal, cross-border, policy, contractual, industry or other obligations linked to the collection, analysis, and use(s) of personal information? Have these all been addressed? If the Processing involves ADMT/Profiling, does the assessment effectively address the reasonably foreseeable risk of: 1. Unfair or deceptive treatment of, or unlawful disparate impact on consumers; 2. Financial or physical injury to consumers; 3. A physical or other intrusion upon the solitude or seclusion, or private affairs or concerns, of consumers if the intrusion would be offensive to a reasonable person; or

	4. Other substantial injury to consumers. Is it foreseeable that the potential application of ADMT/Profiling might seem surprising, inappropriate, or discriminatory or might be considered offensive causing distress or humiliation? How effective are the overall controls and safeguards in reducing risk? Describe with specificity how and why it was determined that the negative impacts do or do not outweigh the benefits, including how any specific safeguards identified in XVI.D. affect this RDPA. If the Attorney General/CPPA requested a copy of the U.S. Assessment, is the organization comfortable in sharing? **Decision** – Do benefits and mitigated risks support proceeding with the Processing activity? Are there any other factors that should be considered? Have the interests, expectations and rights of stakeholders been effectively addressed. Are potential risks to consumers sufficiently mitigated?? Keep a record/inventory of decisions (both Go/No Go decisions). This record/inventory builds up precedent on how similar situations have been decided in the past and is a resource tool to guide future decisions.
XXI. Signatories	**Names, Titles, and Signatures of all Decision Makers (persons who make risk tolerance decisions within the organization):** **Relevant internal actors and external parties contributing to the RDPA:** **Any internal or external audit conducted in relation to the RDPA, including the name of the auditor, the names and positions of individuals involved in the review process, and the details of the audit process:** **Dates the RDPA was reviewed and approved, and names, positions, and signatures of the individuals responsible for the review and approval of the RDPA:**

XXII. Making an RDPA Operational

Each organization has their own way of making requirements operational, whether they be legally driven or internally policy driven. These operational methods are usually developed and designed consistent with the specific organization's culture and rhythm of business. As AI grows a part of business objectives and strategy, organizations will need to expand their own risk assessment and management processes. For example, an assessment process should cover the full development lifecycle requirements from strategy and planning, model development, the specific issues related to training data, deployment, ongoing operation and monitoring issues, and governance. These requirements were outlined in the IAFs AI Assessment which included an example of a five stage gate review to successful AI Governance.[5] Each stage gate review is designed to address specific questions and should involve a broad set of stakeholders and decision makers:


Stage Gates
Shall we proceed with the AI solution?
1
Does the model meet our expectations?
2
Do we deploy the model into production?
3
Is the model ready to be transitioned for BAU operation?
4
Should the model continue as-is, retrained, redesigned, or retired?
5
Model Development Approach
1
Business & Data Understanding
2
Solution Design
Business specification and case, Solution design and social impact and risk assessment:
Business specification and business case
Initial Impact and risk assessment
Conceptual soundness and appropriateness for the given use case
Methodology & design selection
3
Data Extraction
Data:
Appropriateness
Completeness
Quality
Bias
Sampling
Feature engineering
4
Pre-Processing
5
Model Building
Conceptual Soundness, Model Design and Estimation:
Feature selection & Technical soundness
Support for the key assumptions
Outcomes analysis:
Accuracy, robustness, sensitivity, and stability testing
Bias and fairness assessment
Inclusive design assessment
6
Model Deployment
7
Transition & Execution
Production:
Training and prediction
Integration of model with applications and data, load and infrastructure
Prod implementation testing & performance
Transition and execution:
User training, Change management & ROI analysis
Impact assessment decision
Has a balance of benefits and mitigated risks been achieved
8
Ongoing Monitoring
9
Evaluation & Check-in
Governance and controls assessment:
Assessment of model use controls
Assessment of the model risk and performance monitoring plans
Re-assessment of Impact and Risk
Business Case
Solution Development Procedures and Operations
Impact Assessment
Value Scoping
Value Discovery
Value Delivery
Value Stewardship

By extension a RDPA should encompass a review at key parts of a project's development lifecycle. In effect, the assessment

[5] Modified on original Six stage gates to a successful AI governance | by Anand Rao | Towards Data Science

process required by an RDPA should be iterative and matched to the cadence of an organization's development process.

XXI. Assumptions/Caveats/Implications

- The organization has in place a privacy impact assessment (PIA) process that helps the organization determine whether the requirements in the Regulations related to consumer rights (e.g., Transparency, Consent, Access, Portability) have been met.
- The RDPA required by the Regulations and the Colorado AI Act will be added to the organization's already existing assessment processes.
- Organizations will use a set of "triggering" questions to determine if a RDPA, including for ADMT, including Profiling, will be needed. Practically speaking, ALL projects will likely have to go through an initial risk assessment to determine if a RDPA likely will be required. This process reinforces the likelihood of having to design an iterative, multistage assessment more aligned with an AI development lifecycle.
- Capitalized terms used in the RDPA have the meaning given them in the Regulations and the Colorado AI Act.
- Most of the RDPA's listed adverse impacts are contained in the Regulations and the Colorado AI Act. Those that are not, are listed in green.
- The explicit, implied aspects of the Regulations and the Colorado AI Act, coupled with issues that are not defined, suggest the need for enhanced governance controls processes. Many of these are more associated with responsible AI governance.
- Since the Colorado Attorney General can ask for any RDPA upon request, and RDPAs are required to be submitted to the California Privacy Protection Agency, it is likely the AG's office and the Agency would ask for details on processes and controls associated with key mitigators. Therefore, a demonstrable governance program, likely including new elements, should be considered, and in some cases, the Regulations and the Colorado AI Act so require.

XXII. Definitions

A. **"Plain language"** means:

1. Degree and details of any human involvement in the Controller's use of ADMT:

    a. Identifying who at the Controller will be responsible for the use of the ADMT and for what they are responsible

b. Identifying and describing the human's qualifications, if any, to understand the Controller's use of the ADMT, including the personal information processed by, and the logic and output(s) of the ADMT

c. Explaining whether and, if so, how the human evaluates the appropriateness of the personal information processed by, and the logic and output(s) of the ADMT for the Controller's proposed use(s)

d. Explaining whether the human has the authority to influence whether or how the Controller uses the output(s) of the ADMT and, if so, how they exercise this authority

e. If the human can influence how the Controller uses the output(s) of the ADMT, explaining whether and, if so, how the Controller uses the human's influence to calibrate the ADMT or the Controller's use of the ADMT

f. If a human is not involved in the Controller's use of ADMT, explaining why there is no human involvement, and which safeguards the Controller has implemented to address the risks to consumers' privacy that may arise from the lack of human involvement

2. How the Controller evaluates its use of ADMT for validity (confirmation that the ADMT, including its input(s), performs as intended for the Controller's proposed use(s), included the ADMT's accuracy in performing as intended), reliability (ability of the ADMT to perform as intended for the Controller's proposed use(s), repeatedly and without failure, under time interval(s) and conditions consistent with the Controller's proposed use(s), and fairness (equality, equity, and avoidance of discrimination harms):

   a. The metrics the Controller uses to measure performance, validity, reliability, fairness, known limitations, and disparate impact, and why the metrics selected are appropriate measures of validity, reliability, fairness and disparate impact.

   b. If the Controller uses data, hardware, software or other technological components provided by another person, including AI or ADMT, the Controller must identify the name(s) of the person(s), the name(s) of the technological component(s) provided, and how the Controller ensures that the technological component(s) provided do not negatively impact the validity, reliability, fairness or discriminatory impact of the Controller’s use of ADMT

i. This explanation also shall include any copies of internal or external evaluations related to the technological component’s validity, reliability, fairness or discriminatory impact provided to or conducted by the Controller)

c. Whether and, if so, how the Controller evaluated other versions of the ADMT or other ADMT for validity, reliability, fairness, or discriminatory impact for the Controller’s proposed use(s)

d. If the Controller evaluated other versions of the ADMT or other ADMT for validity, reliability, fairness, or discriminatory impact for the Controller’s proposed use(s), why the Controller did not use the other versions or ADMTs

e. The results of the Controller’s evaluations

Appendix 1 – A new data Processing activity is generated when existing Processing activities are modified in a way that materially changes the level of risk presented. For example:

- The way that existing systems or Processes handle personal information
- Processing purpose
- Personal information Processed or sources of consumers’ personal information
- Method of collection of consumers’ personal information
- Personal information recipients
- Intentional and substantial modification to an ADMT
- Processor roles or Processors
- Algorithm applied or algorithmic result
- Software or other systems used for Processing
- The purpose of processing consumers’ personal information
- Consumers’ reasonable expectations concerning the purpose for processing their personal information or the purpose’s compatibility with the context in which their personal information was collected
- The minimum personal information that is necessary to achieve the purpose of the Processing
- The operational elements of the Processing
- The benefits resulting from the Processing to the Controller, the consumer, other stakeholders, and the public
- The negative impacts to consumers’ privacy associated with the Processing, including the sources of these negative

impacts

- The safeguards that the Controller has implemented or plans to implement to address the identified negative impacts
- The Controller's assessment of whether the identified negative impacts identified, as mitigated by the safeguards, outweigh the identified benefits
- Why the Controller is using or seeks to use ADMT to achieve the purpose of the Processing
- The output(s) secured from the ADMT and how the Controller will use the output(s)
- The steps the Controller has taken or any steps it plans to take to maintain the quality of personal information processed by the ADMT, including personal information used by the Controller to train the ADMT
- The logic of the ADMT and the assumptions of the ADMT's logic
- How the Controller evaluates its use of the ADMT for validity, reliability, and fairness
- The degree and details of human involvement in the Controller's use of ADMT
- The safeguards that the Controller plans to implement to address the negative impacts to consumers' privacy that are specific to its use of ADMT or for data sets produced by or derived from the ADMT

Appendix II – Stakeholders, Risk, Benefits and Mitigations

Stakeholders – Examples of stakeholders are:

- Project manager, sponsor and team
- The customer (individual or organization)
- Suppliers of material or other resources
- Creditors
- Employees
- Unions
- City, community, or another geographic region
- Professional organizations
- Any consumer or group impacted by the Processing
- Any consumer or group in a position to support or undermine the Processing
- Internal or external, local or international

Stakeholders may be looked at organizationally, geographically, or by involvement with various Processing phases or outcomes. Another way of determining stakeholders is to identify those who are directly impacted by the Processing and those who may be.

Indirectly affected.

Stakeholders should be grouped by geographic region, organization, Processing involvement, or whether or not they are directly or indirectly impacted. This analysis enables a close look at each stakeholder in order to gather more in-depth information in order to understand their impact, involvement, communication, requirements, and preferences.

Benefits - What are the benefits to the defined impacted other stakeholders? Could the use of this data be used in a way that may result in a specific stakeholder or group of stakeholders being treated differently in a positive way from other groups of individuals? Can the benefits obtained by various stakeholders be measured? Determine and describe the positive impacts on the various stakeholders that are expected to come from the application of this technology/data activity. Determine what the potential positive goal of the difference in treatment is (if any). Are areas of interest such as integrity of the person, autonomy, respect for private life, liberty and security, or education access affected in a positive way?

Examples off benefits to organizations:

- Improved profitability
- Enhanced employee satisfaction, engagement and productivity
- Enhanced customer relationship
- Undamaged brand/reputation
- Enhanced brand/reputation
- Increased market share
- Prevention of cyber-crime and fraud
- New/improved/innovative products/services
- Improved customer service

Are there benefits for society as a whole? Consider factors such as increased revenue, lower costs, improved efficiency, enhanced employee satisfaction, engagement and productivity, enhanced citizen (or workforce) relationship, enhancement or maintenance of brand or reputation, assurance of compliance, fraud prevention, enhancement or maintenance of cyber or physical security, new or improved public services or citizen service, improved manner of marketing, improved ability to assess customer preferences, improvements to innovation or enabling greater, faster, more efficient innovation, improved research processes, improved ability to conduct research and find or enroll study subjects, improved efficiency with studies, innovative ways to conduct research, better healthcare, improved education, positive impact on climate change, more accessible/usable technology, protection of reasonable expectation of privacy (including anonymity), protection of freedom of religion/ thought/speech, protection of prohibition against discrimination on basis of race/national or ethnic origin/color/religion/age/sex/sexual orientation/marital status/disability. Do the benefits of having the model and/or data use in production outweigh the costs of maintaining it? Are there any social interests served with the deployment of the Processing activity? How does the Processing activity contribute to or increase well-being? How will the Processing activity contribute to human values?

Examples of benefits to individuals:

- More objective outcomes
- Safer interactions
- Better product selection
- Better access to new products and services
- Significant discounts
- Better product utilization
- Improved service
- Improved ease of use
- Engaged consumers/customers/employees
- More convenience
- Appropriately linked to other choices
- Anticipating or meeting of a need
- Exercise of self-determination
- Public sector access
- Anonymous transportation

Risks - Considering all the factors relating to the data, metric or measure, the likely use, the associated activity, the identifiability and sensitivity of the information and its use, what are the risks (real and/or perceived) to the identified stakeholders/users? Could any metric of measure be used in a way that makes a decision on a specific user and/or creates a profile on them (real or perceived)? Specific risks to individuals to consider include:

- Constitutional harms, such as chilling or deterring consumers' free speech or expression, political participation, religious activity, free association, freedom of belief, freedom to explore ideas, or reproductive freedom; and harms to consumers' ability to engage in collective action or impede the right to unionize;

- Intellectual privacy harms, such as the creation of negative inferences about an individual based on what an individual reads, learns, or debates and using information inferred about consumers to manipulate them;
- Negative impacts to consumers' data security, such as unauthorized access, destruction, use, modification or disclosure of personal information or unauthorized activity resulting in the loss of availability of personal information;
- Discrimination harms, such as a violation of federal antidiscrimination laws or antidiscrimination laws or any state or political subdivision thereof, or discrimination upon the basis of protected class(es) or their proxies or that has unlawful disparate impact such as upon protected class(es);
- Unfair, unconscionable, or deceptive treatment;
- A negative outcome or decision with respect to an individual's eligibility for a right, privilege, or benefit related to financial or lending services, housing, insurance, education enrollment or opportunity, criminal justice, employment opportunities, health-care services, or access to essential goods or services;
- Financial injury or economic harm, including limiting or depriving consumers of economic opportunities; charging consumers higher prices; compensating consumers at lower rates; or imposing additional costs upon consumers including costs associated with the unauthorized access to consumers' personal information;
- Physical injury (including processing that creates the opportunity for physical or sexual violence), harassment, or threat to an individual or property;
- Privacy harms, such as physical or other intrusion upon the solitude or seclusion of the private affairs or concerns of consumers, stigmatization or reputational injury, infer highly sensitive, latent information from seemingly innocuous surface data;
- Psychological harm, including anxiety, embarrassment, fear, emotional distress, stress, frustration, shame, feelings of violation, and other mental trauma;
- Other detrimental or negative consequences that affect an individual's private life, private affairs, private family matters or similar concerns, including actions and communications within an individual's home or similar physical, online, or digital location, where an individual has a reasonable expectation that personal information or other data will not be collected, observed, or used;
- Inconvenience or expenditure of time;
- Disruption and intrusion from unwanted commercial communications or contacts
- Loss of autonomy through acts or practices that are not reasonably foreseeable by an individual and that are intended to materially:
  - Alter that individual's experiences
  - Limit that individual's choices

  - Influence that individual’s responses, or
  - Predetermine results or outcomes for that individual;
- Impairing consumers’ control over their personal information, such as by providing insufficient information for consumers to make an informed decision regarding the processing of their personal information or by interfering with consumers’ ability to make choices consistent with their reasonable expectations or by giving Controllers far more insight into their customers than their customers have into them, thereby producing a power imbalance between the Controllers and their customers;
- Coercing or compelling consumers into allowing the processing of their personal information, such as by conditioning consumers’ acquisition or use of an online service upon their disclosure of personal information that is unnecessary to the expected functionality of the service or feeling pressure to conform to behaviors that consumers think will please the algorithmic decisionmaker;
- Exploiting consumers’ vulnerabilities, such as age, employment or student status, immigration status, health status, or financial hardship;
- Producing erroneous models and predictions, as a result of inaccurate data or faulty algorithms, that can negatively impact consumers;
- Creating opacity and procedural unfairness because most consumers lack an understanding of, and opportunities to challenge, the Controller’s algorithmic determinations that can shape their life opportunities;
- Increasing automation which, in turn, displaces human labor; or
- Using analytics for intentional, harmful purposes.

Consider what factors about the activity have the highest impact on the likelihood any of these risks could be realized?

Examples of risks to large groups and society are:

- Overuse of technology (e.g., social media by teenagers)
- Utilization of disinformation and propaganda
- Economic and asset inequalities – Who will and will not have access to technology
- Utilization of technology by surveillance state
- Utilization of online tools by hateful and criminal actors – Intentional, harmful use
- Algorithmic bias in technology against protected classes
- Who controls consumer data and who monetizes it – Controllers have far more insight into their customers than their

customers have into them producing a power imbalance between Controllers and customers

- Loss of consumer trust
- Invasion of privacy – Inferring from innocuous surface data highly sensitive latent information about consumers
- Manipulation – Using information inferred about consumers to manipulate them
- Opacity and procedural unfairness – Most consumers lack an understanding of, and opportunities to challenge, Controllers' algorithmic determinations that can shape their life opportunities
- Displacement of labor – Increased automation can displace human labor
- Pressure to conform – Consumers may feel pressure to conform to behaviors that they think will please algorithmic decisionmakers

Examples of risks to organizations

- Negative media attention
- Negative regulatory impact
- Compliance
- Reputational
- Business continuity
- Financial loss
- Reticence risk
- Crime
- Fraud

Safeguards/Additional Mitigators:

At a minimum, the implementation of the following safeguards should be considered:

- Safeguards to protect personal information such as encryption, segmentation, and access controls;
- Use of privacy-enhancing techniques such as trusted execution environments, federated learning, homomorphic encryption, and differential privacy;
- Restrictions on the processing of personal information;

Some additional mitigators to consider are:

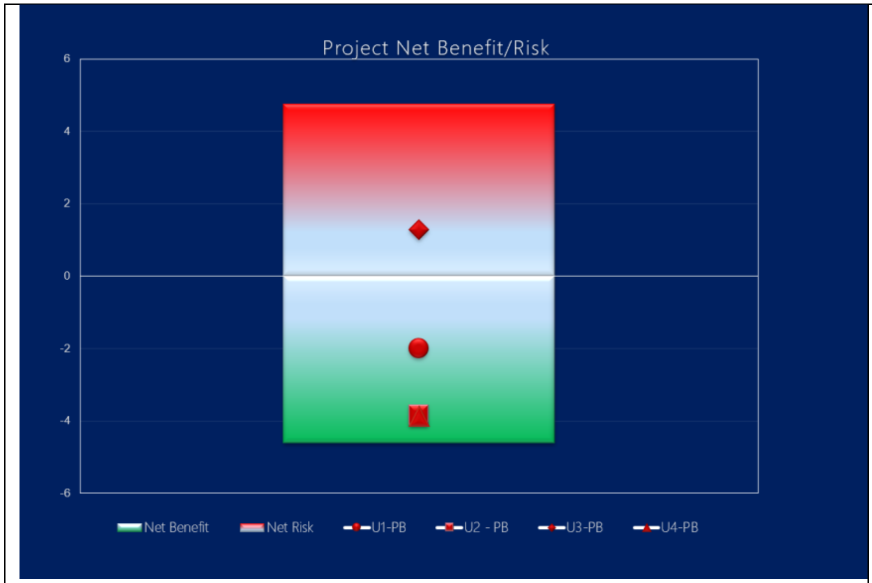
a. The use of de-identified data; b. Measures taken pursuant to the Controller duties, including an overview of data security practices the Controller has implemented, any data security assessments that have been completed, and any measures taken to comply with the consent requirements of the Regulations; c. Measures taken to ensure that consumers have access to the rights provided in the Regulations; d. Contractual agreements in place to ensure that personal information in the possession of a Processor or other Third Party remains secure; or e. Any other practices, policies, or trainings intended to mitigate Processing risks.

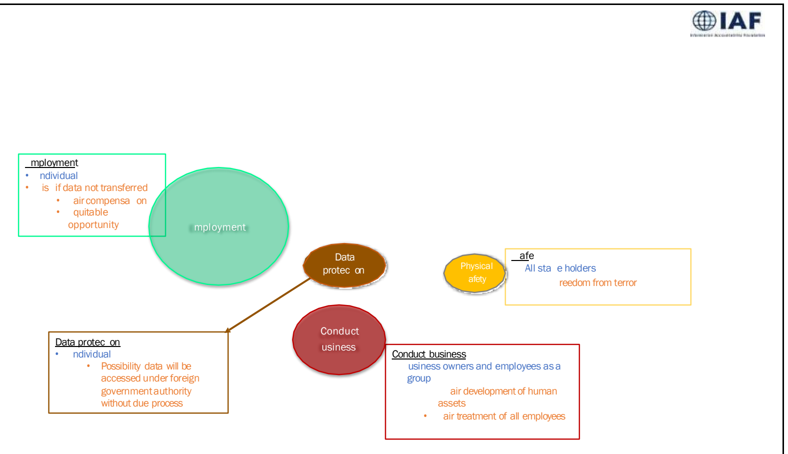
If Profiling is being assessed, safeguards used to reduce the risk of harms identified and safeguards for any data sets produced by or derived from the ADMT/Profiling must be assessed. Include a description of any other practices, policies, or trainings intended to mitigate Processing risks

In addition, are there any technical and/or procedural safeguards (mitigating controls) that could be implemented to prevent and mitigate risks should they occur (e.g., increased transparency, additional suggestions/guidance to the customer, more choice, etc.)?

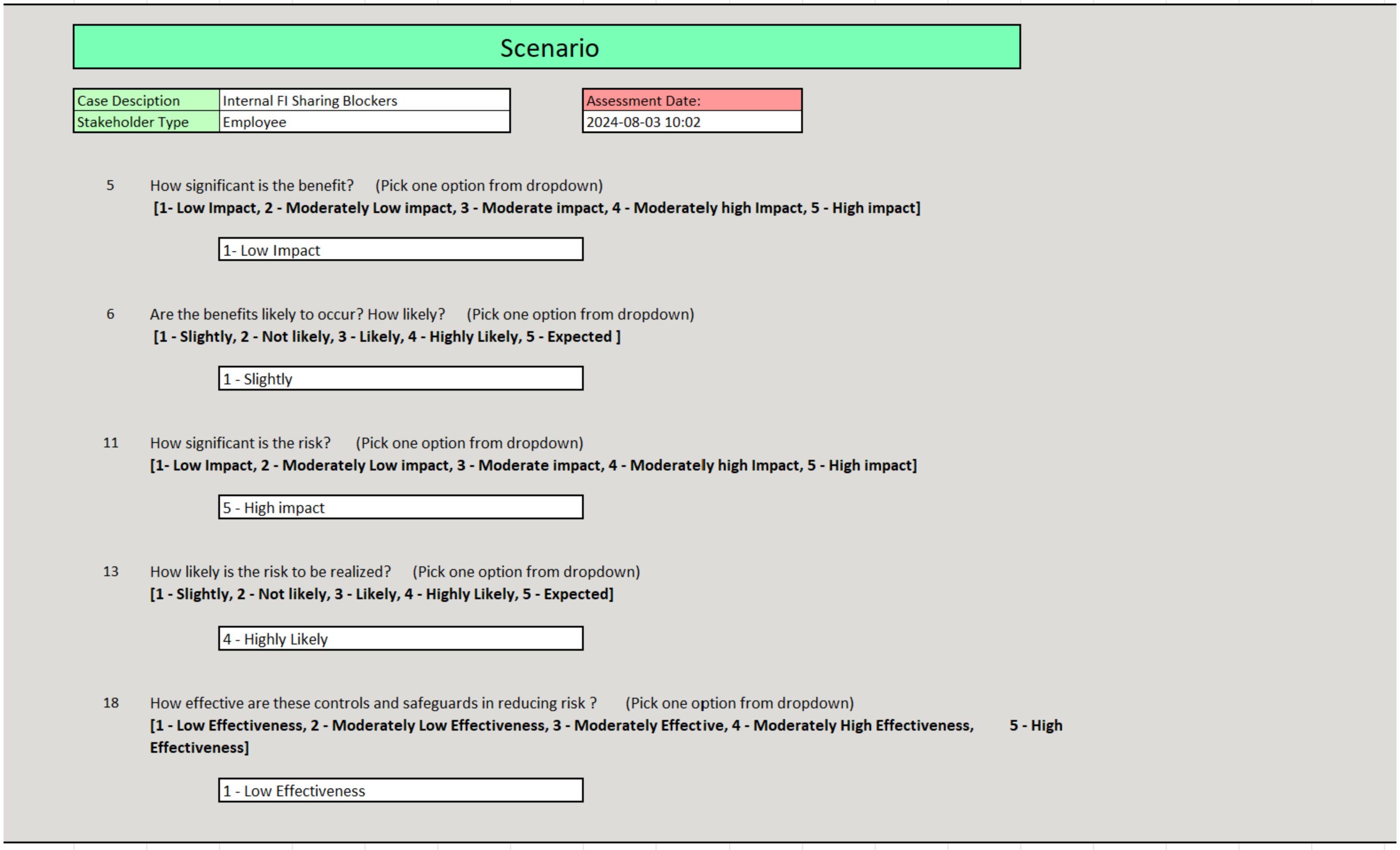
Appendix III

Two examples of Multi-Dimensional/Stakeholder Balancing Output – Both use a math determined balancing that aligns with a common risk management methodology (weighing), but one uses a mathematical depiction and the other one uses a pictorial depiction of the balancing (see A-Principled-Approach-to-Rights-and-Interest-Balancing.pdf (secureserver.net) for more details).




IAF
mployment
• ndividual
• is if data not transferred
• air compensa on
• quitable opportunity
mployment
Data protec on
Physical afety
afe
All sta e holders
reedom from terror
Data protec on
• ndividual
• Possibility data will be accessed under foreign government authority without due process
Conduct usiness
Conduct business
usiness owners and employees as a group
air development of human assets
• air treatment of all employees

The mathematical model uses a process common to many risk management programs and utilizes a defined 1,3,5 scale. The answers to the five scale questions then are computed to reflect the “projected net benefit/risk” result. This form of mathematical assessment usually is best performed as an output of the internal stakeholder evaluation of the multi-stakeholder benefits, risks and mitigations and can be used to support the narrative part of the assessment.



Scenario

Case Desciption	Internal FI Sharing Blockers
Stakeholder Type	Employee

Assessment Date:
2024-08-03 10:02

5 How significant is the benefit? (Pick one option from dropdown)
[1- Low Impact, 2 - Moderately Low impact, 3 - Moderate impact, 4 - Moderately high Impact, 5 - High impact]

1- Low Impact

6 Are the benefits likely to occur? How likely? (Pick one option from dropdown)
[1 - Slightly, 2 - Not likely, 3 - Likely, 4 - Highly Likely, 5 - Expected]

1 - Slightly

11 How significant is the risk? (Pick one option from dropdown)
[1- Low Impact, 2 - Moderately Low impact, 3 - Moderate impact, 4 - Moderately high Impact, 5 - High impact]

5 - High impact

13 How likely is the risk to be realized? (Pick one option from dropdown)
[1 - Slightly, 2 - Not likely, 3 - Likely, 4 - Highly Likely, 5 - Expected]

4 - Highly Likely

18 How effective are these controls and safeguards in reducing risk ? (Pick one option from dropdown)
[1 - Low Effectiveness, 2 - Moderately Low Effectiveness, 3 - Moderately Effective, 4 - Moderately High Effectiveness, 5 - High Effectiveness]

1 - Low Effectiveness

Appendix IV Triggers for a RDPA

Triggers for a RDPA	Regulations collectively require a RDPA for processing that presents a heightened risk of harm (High Risk Processing). High Risk Processing includes: • Selling or sharing personal information (Colorado used the term personal data, and California uses the term personal information. This document uses the term personal information). • Processing sensitive information. • Using Automated Decision-Making Technology (ADMT) (California uses the term ADMT which includes profiling, the Colorado Rules use the term Profiling, and the Colorado AI Act uses the term High-Risk Artificial Intelligence System; for inclusiveness, this document uses the term ADMT), in furtherance of a decision that results in the provision or denial of financial of financial or lending services, housing, insurance, education enrollment or education opportunity, criminal justice, legal services, employment or employment opportunity, contracting opportunities or compensation, healthcare services, or access to essential goods, services, or opportunities, including essential government services. • Processing the information of consumers that the Controller has actual knowledge are less than 16 years of age. • Processing the personal information of consumers who are employees, independent contractors, job applicants, or students using technology to monitor employees, independent contractors, job applicants, or students. • Processing the personal information of consumers in publicly accessible places using technology to monitor consumers' behavior, location, movements, or actions. • Processing the personal information of consumers to train artificial intelligence or ADMT. • Processing of personal information for the purposes of targeted advertising. • Processing of personal information for the purposes of ADMT where such ADMT presents a reasonably foreseeable risk of:

	○ unfair or deceptive treatment of, or unlawful disparate impact on, consumers ○ financial, physical or reputation injury to consumers ○ physical or other intrusion upon the solitude or seclusion, or the private affairs or concerns, of consumers, where such intrusion would be offensive to a reasonable person.