

Template for Social Impact Statements for Algorithms

This is a template that can be used to render in open, standard, machine-readable StratML format the contents of a social impact statement for your algorithm(s). It can be opened for editing in an XForm by pasting this link into your browser: <http://stratml.us/forms/walt5.pl?url=http://stratml.us/carmel/iso/part2/TSISA.xml> Rename and edit it as appropriate for your organization and algorithm(s).

The contents of this template are taken from the Principles for Accountable Algorithms and a Social Impact Statement for Algorithms issued by FATML.org, which are available in StratML format at <https://stratml.us/carmel/iso/SIS4AwStyle.xml> with a link to the original source.

Contents

Vision.....	3
Mission.....	3
Values	3
1. Responsibility.....	4
1.1. Performer.....	4
1.2. Contact	4
1.3. Contingency Plan.....	4
1.4. Sunset Plan	4
2. Explainability	6
2.1. Plan.....	6
2.2. Visibility, Access & Changes	6
2.2.1. Model	6
2.2.2. Training Data.....	6
2.2.3. Test Data.....	7
2.3. Sources	7
3. Accuracy	8
3.1. Errors & Harm	8
3.2. Sensitivity Analysis.....	8
3.3. Corrections	8
3.4. Validity.....	8
3.5. Uncertainty	9
4. Auditability	10
4.1. APIs	10
4.2. Sample Data.....	10
4.3. Audits	10
4.4. Communication Plan	11
5. Fairness	12
5.1. Social Context	12
5.2. Data Mining.....	12
5.3. Errors.....	12
Administrative Information.....	13

DEMONSTRATION ONLY



Your Organization's Name (Your Org's Acronym)

Description:

You may briefly describe your organization here.

Stakeholder(s):

Stakeholder Name :

Stakeholder description

Role: Role Name

Role description

Vision

A phrase describing the vision of your organization or algorithm, i.e., how the world will be different due to its existence

Mission

A phrase describing the basic purpose of your organization or algorithm

Values

Name of the Value: Brief description of this value. Insert as many additional values as may be appropriate to establish the moral basis for your algorithm(s)

DEMONSTRATION ONLY

1. Responsibility

[To be described]

Guiding Questions Who is responsible if users are harmed by this product? What will the reporting process and process for recourse be? Who will have the power to decide on necessary changes to the algorithmic system during design stage, pre-launch, and post-launch? Initial Steps to Take:

1.1. Performer

Determine and designate a person who will be responsible for the social impact of the algorithm.

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

1.2. Contact

Make contact information available so that if there are issues it's clear to users how to proceed

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

1.3. Contingency Plan

Develop a plan for what to do if the project has unintended consequences.

This may be part of a maintenance plan and should involve post-launch monitoring plans.

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

1.4. Sunset Plan

Develop a sunset plan for the system to manage algorithm or data risks after the product is no longer in active development.

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

DEMONSTRATION ONLY

2. Explainability

[To be described]

Guiding Questions: Who are your end-users and stakeholders? How much of your system / algorithm can you explain to your users and stakeholders? How much of the data sources can you disclose? Initial Steps to Take:

2.1. Plan

Have a plan for how decisions will be explained to users and subjects of those decisions.

In some cases it may be appropriate to develop an automated explanation for each decision.

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

2.2. Visibility, Access & Changes

Allow data subjects visibility into the data you store about them and access to a process in order to change it.

If you are using a machine-learning model:

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

2.2.1. Model

Consider whether a directly interpretable or explainable model can be used.

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

2.2.2. Training Data

Describe the training data including how, when, and why it was collected and sampled.

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

2.2.3. Test Data

Describe how and when test data about an individual that is used to make a decision is collected or inferred.

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

2.3. Sources

Disclose the sources of any data used and as much as possible about the specific attributes of the data.

Explain how the data was cleaned or otherwise transformed.

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

3. Accuracy

[To be described]

Guiding Questions: What sources of error do you have and how will you mitigate their effect? How confident are the decisions output by your algorithmic system? What are realistic worst case scenarios in terms of how errors might impact society, individuals, and stakeholders? Have you evaluated the provenance and veracity of data and considered alternative data sources? Initial Steps to Take:

3.1. Errors & Harm

Assess the potential for errors in your system and the resulting potential for harm to users.

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

3.2. Sensitivity Analysis

Undertake a sensitivity analysis to assess how uncertainty in the output of the algorithm relates to uncertainty in the inputs.

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

3.3. Corrections

Develop a process by which people can correct errors in input data, training data, or in output decisions.

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

3.4. Validity

Perform a validity check by randomly sampling a portion of your data (e.g., input and/or training data) and manually checking its correctness.

This check should be performed early in your development process before derived information is used. Report the overall data error rate on this random sample publicly.

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

3.5. Uncertainty

Determine how to communicate the uncertainty / margin of error for each decision.

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

DEMONSTRATION ONLY

4. Auditability

[To be described]

Guiding Questions: Can you provide for public auditing (i.e. probing, understanding, reviewing of system behavior) or is there sensitive information that would necessitate auditing by a designated 3rd party? How will you facilitate public or third-party auditing without opening the system to unwarranted manipulation? Initial Steps to Take:

4.1. APIs

Document and make available an API that allows third parties to query the algorithmic system and assess its response.

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

4.2. Sample Data

Make sure that if data is needed to properly audit your algorithm, such as in the case of a machine-learning algorithm, that sample (e.g., training) data is made available.

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

4.3. Audits

Make sure your terms of service allow the research community to perform automated public audits.

Stakeholder(s):

Research Community

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

4.4. Communication Plan

Have a plan for communication with outside parties that may be interested in auditing your algorithm, such as the research and development community.

Stakeholder(s):

R&D Community

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

DEMONSTRATION ONLY

5. Fairness

[To be described]

Guiding Questions: Are there particular groups which may be advantaged or disadvantaged, in the context in which you are deploying, by the algorithm / system you are building? What is the potential damaging effect of uncertainty / errors to different groups? Initial Steps to Take:

5.1. Social Context

Talk to people who are familiar with the subtle social context in which you are deploying.

For example, you should consider whether the following aspects of people's identities will have impacts on their equitable access to and results from your system:

- Race
- Sex
- Gender identity
- Ability status
- Socio-economic status
- Education level
- Religion
- Country of origin

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

5.2. Data Mining

If you are building an automated decision-making tool, you should deploy a fairness-aware data mining algorithm.

See, e.g., the resources gathered at <http://fatml.org>.

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

5.3. Errors

Calculate the error rates and types (e.g., false positives vs. false negatives) for different sub-populations and assess the potential differential impacts.

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

Administrative Information**Start Date:****End Date:****Publication Date:** 2020-04-22**Source:** [URL for this plan on your website](#)**Submitter:****Given Name:** Your Given Name**Surname:** Your Surname**Email:** **Phone:**

_fe23675c-84cc-11ea-90b1-bb7d77c2bb56