Vendor FactSheet for Algorithmic Systems

Please provide details regarding your algorithmic system product by filling out the FactSheet[1] template and Algorithmic Impact Assessment Questionnaire below. You can find an example of a completed FactSheet on page 3.

FactSheet

Vendor Name	*Your organization*
Model Name	*Name of the system*
Overview	*Briefly describe the system's purpose and function.*
Purpose	*Describe the intended use of the system.*
Intended Domain	*Describe the domain/context the system is intended to be used in.*
Training Data	*Describe the data used to train the model.*
Model Information	*Explain how the model works at a high-level.*
Inputs and Outputs	*Describe the data that is fed into the model and the data that is generated by the model.*
Performance Metrics	*Describe the accuracy of the system, preferably with numerical metrics (e.g. false positive rate, false negative rate, true positive rate, etc.). If the system predicts a continuous value (e.g., electricity prices, translation) provide relevant performance metrics for the use-case (e.g., average % error and mean-squared error from actual price, BLEU/NIST translation score or human judgement score) and cite the metric used. Specify if performance metrics were calculated on the same data used to train the model or not.*
Optimal Conditions	*What conditions are necessary for the system to perform optimally?*
Poor Conditions	*Under what conditions does the system's performance decrease in accuracy?*
Bias	*Have you tested this model for bias across dimensions such as race, gender, etc.? What did that testing entail and what were the results?*

[1] The FactSheet template is heavily inspired by the IBM Research AI FactSheets 360 project.

Test Data	*Describe the data used to test the model's performance.*

Algorithmic Impact Assessment Questionnaire

Accuracy	
Under what conditions/circumstances has the system been tested?	
Have the vendors or an independent party conducted and published a validation report (including the methodology and results) that audits for accuracy and discriminatory/disparate impact? If yes, can the City review the study?	
Will the model be learning from the information it gets in the field during deployment?	
Equity	
What quality control is in place to test and monitor for potential biases in the AI system (e.g., non-representative training data, overfitting, hard-coded rules)?	
How can the City and its partners flag issues related to bias, discrimination or poor performance of the AI system?	
Explainability	
What performance metrics were selected to judge the model's effectiveness? What is it optimizing for, and under what constraints?	
How are the outcomes of the AI system explained to subject matter experts, users, impacted individuals, or others?	

Example: FactSheet[2]

Vendor Name	XYZ Technologies, Inc.
Model Name	Image Caption Generator
Overview	This document is a FactSheet accompanying the Image Caption Generation model on IBM Developer Model Asset eXchange.
Purpose	This model generates captions from a fixed vocabulary that describe the contents of images.
Intended Domain	Computer Vision
Training Data	The model is trained on the COCO dataset.
Model Information	The model, named Show and Tell, is based on an encoder-decoder pattern.
Inputs and Outputs	Input: An image. Output: Description of the image
Performance Metrics	Model is assessed based on human assessment of the quality of the caption matched to the image (scored 0-5). The average score for captions reviewed by humans in 2021 was 3.7, with a standard deviation of 0.3. 95% of captions were scored between 3.1 and 4.3.
Optimal Conditions	• Model works well for inputs similar to the training dataset. • Images have good resolution and lighting.
Poor Conditions	• Images have poor resolution or lighting. • The input is from a different distribution than what the model is trained on. • The model is not trained for a specific class.
Bias	The Image Caption Generator was evaluated for bias for Male gender as against Female gender using the AIF360 toolkit. From the evaluations it was found that the model is 42.1% more biased towards generating male-specific caption words in images than female-specific gender caption words.

[2] The example FactSheet is taken from IBM Research AI Factsheet 360's Image Caption Generator sample.

Test Data	Test dataset provided by 2015 MSCOCO Image Captioning Challenge. More about the evaluation server can be found here.
Explanation	While the model architecture is well documented, the model is still a deep neural network, which largely remains a black box when it comes to explainability of results and predictions.

Example: Algorithmic Impact Assessment Questionnaire

Accuracy	
Under what conditions/circumstances has the system been tested?	The model has been tested on the testing data set, which includes professional stock photos of humans, animals, buildings, and nature. It has not been tested on images that have been taken by laypeople using their phone cameras.
Have the vendors or an independent party conducted and published a validation report (including the methodology and results) that audits for accuracy and discriminatory/disparate impact? If yes, can the City review the study?	Yes, you can find the validation report on our company website. Link here: https://....com
Will the model be learning from the information it gets in the field during deployment?	No, the model will not be re-calibrating in the field. This would require manual modification of the algorithm or re-training of the model with new data by our engineers.
Equity	
What quality control is in place to test and monitor for potential biases in the AI system (e.g., non-representative training data, overfitting, hard-coded rules)?	The training data set was created in consultation with a representative sample of the US population by race, gender, and age to mitigate any biases in the text generated. We understand that people of different race, gender, and age may describe an image differently than one another, and continue to modify our training data to feature captions representative of the US population.
How can the City and its partners flag issues related to bias, discrimination or poor performance of the AI system?	There is built-in functionality in our program for the user to report one-off incidents of an inaccurate/biased output. This feedback informs the updates our engineers make to the training

	data and algorithm design, which would be reflected in later versions of the software. To report concern of a more comprehensive, systematic bias of the model, please contact abcdef@gmail.com.
Explainability	
What performance metrics were selected to judge the model's effectiveness? What is it optimizing for, and under what constraints?	Match quality of caption to image (reported by human reviewer) on a scale from 0-5. The model is optimizing for accurate descriptions of the image, and is penalized for lengthy caption text.
Are the outcomes of the AI system understandable by subject matter experts, users, impacted individuals, and others?	While the model architecture is well documented, the model is still a deep neural network, which largely remains a black box when it comes to explainability of results and predictions.