DPO Review: Environmental Services Division (ESD) is requesting an AI solution that can identify fullness and contents of trash cans to support reducing recycle bin contamination. This information will not be used levy fines or penalties. It will be used to inform future outreach, such as a follow up letter and/or other educational outreach as needed.

Factsheet was requested given computer vision component played significant role in application.

No PII is involved and AI risk is low given it will only inform general outreach strategy.

Approved.

AI FactSheet for Third Party Systems

Please provide details regarding your Artificial Intelligence (AI) product by filling out the FactSheet[1] template below. You can find an example of a completed FactSheet on page 3.

FactSheet

Vendor Name	Zabble Inc.
System Name	Zabble Zero Mobile Tagging
Overview	Zabble is a technology company that helps organizations efficiently manage their waste and workflows using an AI-powered platform, Zabble Zero. One of its products, Mobile Tagging, is a mobile- and web based app to conduct efficient audits using computer vision AI to measure fullness and detect contaminants and engage stakeholders with real-time alerts.
Purpose	Zabble's AI automatically identifies a bin's fullness and its contaminants in real-time through a mobile phone's camera via the Mobile Tagging app. This allows the users to not only spend less time documenting a bin's fullness and contaminants but also access aggregated insights across the inspections.
Intended Domain	Waste Management
Training Data	The model is trained using top-down images of waste receptacles from customer deployments of the platform in various environments. Images are labeled by professional annotators.

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

Test Data	The models are tested on a holdout set of customer images, labeled according to the same methodology as the training data. The models are implemented in Zabble Zero's mobile-based app. Prior to deployment, the models are tested in a test version of the app in a close proximity to the customer environment.
Model Information	Zabble's AI includes object detection and image classification models. The base model for object detection is YOLOv5. There are two image classification models, one for predicting interior fullness and one for predicted exterior fullness. The base model for fullness classification is ResNet18. All models are fine-tuned using Zabble's training dataset.
Update procedure	Models are trained twice a year when a critical mass of data has been added. Users are encouraged/required to use the latest version of the Zabble Zero app, so they do not have the option to use an older version of the models.
Inputs and Outputs	Inputs: A jpeg image file at least 480x680 pixels (images are resized by the models) Outputs: A JSON of the predicted fullness classes and probabilities; a JSON with a list of objects detected, their bounding coordinates, and the model confidence
Performance Metrics	The deployed models are evaluated on new customer images labeled according to the same methodology as the training and test sets. Zabble computes these metrics overall, by customer, and by receptacle type. For fullness, accuracy and adjacent accuracy (within 10% of the labeled fullness range) are evaluated. For object detection, Zabble reports metrics on the existence of an item in an image. Sensitivity and specificity are reported.
Bias	Human-factor bias is not relevant to Zabble Zero AI. Sampling bias may be present, as Zabble's training dataset relies on the images of its current and past customers.
Robustness	The Zabble app instructs users to review the outcomes of the AI's predictions and make adjustments where needed. Zabble's regular AI evaluations examine

	areas where the model is not performing well, and seeks to improve this performance in the next version of the model.
Optimal Conditions	Zabble AI performs best on clear images with sufficient lighting, no blurriness, when the top layer of the contents of the receptacle can be seen, and on receptacles where most of the edges of the container are visible.
Poor Conditions	Zabble AI does not perform well on low-light, blurry images, when the contents of the receptacle are otherwise difficult to see, or when it is difficult for a human to estimate the fullness of a container because the top edges are not visible or most of the receptacle is not in view.
Explanation	In live prediction mode before the photo is taken, the predicted fullness % and the boxes and labels of detected items are shown on the camera screen. After taking the photo, the app presents the fullness prediction in the form of a slider, which the user is expected to adjust when not accurate. For detected objects, the app provides these written instructions: "Confirm any items detected from the image or commonly tagged by other users."

Algorithmic Equity Assessment Questionnaire

How is the AI tool monitored to identify any problems in usage? Can outputs (recommendations, predictions, etc.) be overwritten by a human, and do overwritten outputs help calibrate the system in the future?	The Zabble app instructs users to review the outcomes of the AI's predictions and make adjustments where needed. Zabble's regular AI evaluations examine areas where the model is not performing well, and seeks to improve this performance in the next version of the model.
Have the vendors or an independent party conducted a study on the bias, accuracy, or disparate impact of the system? If yes, can the City review the study? Include methodology and results.	A summary of Zabble's most recent fullness model evaluation is available here: https://www.zabbleinc.com/blog-post/latest-zabble-zero-ai-fullness-predictions Zabble's evaluation of its previous object detection model is available here: https://www.zabbleinc.com/blog-post/zabblezero-

	contamination-object-detection-accuracy **NOTE: A newer report of the results has currently not been published.
Is the data used to the train the system representative of the communities it covers?	Zabble's training dataset uses the images of its current and past customers, which is generally representative of other images of waste receptacle data. However, there may be differences in the types of waste/recycling materials present and/or the types of containers holding these materials..
How can the City and its partners flag issues related to bias, discrimination or poor performance of the AI system?	Feedback can be provided by contacting the Zabble Customer Success Manager or by emailing support@zabbleinc.com.
How is the AI tool made accessible to people with disabilities?	People with mobility, auditory, speech, or mild cognitive disabilities can use the Zabble Zero platform with little to no accommodations. The Zabble app is designed for visual waste audits, therefore, people with significant visual impairments may be unable to use the technology.
What other human factors, if any, were considered for usability and accessibility of the system?	Zabble designs for ease-of-use to allow users with only basic training to use the mobile app.

Example FactSheet[2]

This is an example of the factsheet above completed by a fake company. This is only here for reference and does not need to be included in the completed form.

Vendor Name	XYZ Technologies, Inc.

[2] The example FactSheet is taken from IBM Research AI Factsheet 360's Audio Classifier sample.

System Name	Audio Classifier
Overview	This document is a FactSheet accompanying the Audio Classifier model on IBM Developer Model Asset eXchange.
Purpose	This model classifies an input audio clip.
Intended Domain	This model is intended for use in the audio processing and classification domain.
Training Data	The model is trained on the AudioSet dataset by Google.
Test Data	The test set is also part of the AudioSet data. There was a 70:20:10% split of the data into train:val:test. The ratio of samples/class was maintained as much as possible in all the splits.
Model Information	The audio classifier is a two-stage model: • The first model (MAX-Audio-Embedding-Generator) converts each second of input raw audio into vectors or embeddings of size 128 where each element of the vector is a float between 0 and 1. • Once the vectors are generated, there is a second deep neural network that performs classification.
Inputs and Outputs	Input: a 10 second clip of audio in signed 16-bit PCM wavfile format. Output: a JSON with the top 5 predicted classes and probabilities.
Performance Metrics	Metric Value Mean Average Precision 0.357 Area Under the Curve 0.968 d-prime 2.621
Bias	The majority of audio samples in the training data set represent voice and music content. Potential bias caused by this over-representation has not been evaluated. Careful attention should be paid if this model is to be incorporated in an application where bias in voice type or music genre is potentially sensitive or harmful.

Robustness	This audio classifier is not robust to the L-infinity and L2 norms for the HopSkipJump attack. L2 L-Infinity 5th Percentile 887.0 (200.9) 5.5 (4.9) 10th Percentile 1496.6 (720.6) 7.53 (5.73) 15th Percentile 3723.1 (4707.2) 52.8 (41.8) 25th Percentile 7187.9 (---) 187.6 (198.1) 50th Percentile 11538.6 (---) 502.8 (---) The susceptibility of the model to the two attacks. The parenthetical values in the table above represent the fitted curve evaluated at 11 iterations. (When we are unable to fit a curve, or the result is negative, we denote by ---.)
Optimal Conditions	• When the input audio contains only one or two distinct audio classes. • When the audio quality is high with lesser noise.
Poor Conditions	• When the audio contains more that two distinct classes. • When the audio quality is low with more noise.
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.

Algorithmic Equity Assessment Questionnaire

How is the AI tool monitored to identify any problems in usage? Can decisions be overwritten by a human, and do overwritten decisions help calibrate the system in the future?	The system can be monitored in usage, and audio classification decisions can be retroactively overwritten by a human. The overwritten decisions can help calibrate the system in the future if desired.
Have the vendors or an independent party conducted and published a validation report (including the methodology and results) that	Yes. Results from the third party study can be provided upon request.

audits for accuracy and discriminatory/disparate impact? If yes, can the City review the study?	
Is the data used to the train the system representative of the communities it covers?	The majority of audio samples in the training data set represent voice and music content from Google's AudioSet dataset. The dataset includes voices of all types, but may not be in a ratio reflective of the San José community. Potential bias caused by this over-representation has not been evaluated.
How can the City and its partners flag issues related to bias, discrimination or poor performance of the AI system?	The system provides a web portal to each customer to show the results of the system and its impact on transit performance in the form of reports and graphs.
How is the AI tool made accessible to people with disabilities?	The system is embedded into a graphics user interface that is compliant with modern screen readers, and provides the option for auto-generated dictation of text on the screen.
What other human factors, if any, were considered for usability and accessibility of the system?	Users can provide feedback on the interface through an online suggestion box, which has informed the evolution of the interface design over the last several years.