

Urban Feedback System

Using image recognition and text generation for urban safety

SYSTEM’S DESCRIPTION

The AI system is designed to quickly identify and describe unsafe situations involving people or infrastructure based on photos and text comments submitted by municipality residents on the city app. It uses image recognition and text generation to process these submissions and was developed for municipal staff to improve response times.

BENEFITS

Faster resolution of issues	<div><div></div><div></div><div></div></div>
Improved accessibility of public services	<div><div></div><div></div><div></div></div>
Reduction in the need for human monitoring	<div><div></div><div></div><div></div></div>
Increased public safety	<div><div></div><div></div><div></div></div>
Support for urban planning	<div><div></div><div></div><div></div></div>
	<div><div></div>benefit enjoyed by</div>



In use

Launched on January 10, 2024, the urban feedback system operates within the city’s boundaries and is currently used by 10 municipal staff, with the capacity to add more accounts as needed, up to a limit of 20.

IMPACT ASSESSMENT REPORT

available in multiple formats including Braille



Last update: 20 Feb 2025

RISKS

MITIGATION STRATEGIES

Capability Risks

Privacy violations from model outputs	Implement face blurring and data anonymization
Misidentification of people or objects	Use diverse training datasets to minimize biases and improve accuracy across demographics and environments

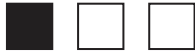
Human Interaction Risks

Staff uncertainty about generated descriptions	Include explanation logs showing which elements from the photo and resident text were used in the description
Resident unfamiliarity	Conduct user testing with diverse groups of residents to ensure intuitive design and accessibility of the city app

Systemic Impact

Over-collection and retention of resident’s user data	Establish strict data governance policies, including data minimization and retention limits
---	---

City residents
Municipal staff
Institutions
and Environment



risk faced by

SYSTEM’S DATA

Essential

Urban scene images	<div><div></div><div></div></div>
Issue descriptions	<div><div></div><div></div></div>
Location data	<div><div></div><div></div></div>
Date and time of submission	<div><div></div><div></div></div>

Non-essential

Resident's user profile information	<div><div></div><div></div></div>
Photographic metadata	<div><div></div><div></div></div>
	<div><div></div>yes</div>

Potentially employed
for future uses
Personally indetifiable
information

PERFORMANCE OF MODELS ON DATA

Data	Model	Version
Urban scene images	PaliGemma	3B-MIX
Issue descriptions		

Model metrics across subgroups

We report the highest observed scores (% of analyzed urban scenes and generated issue descriptions) and the subgroups where they occurred.

The evaluation considers the following subgroups:

- **Perceived gender:** Male, Female
- **Ethnicity:** White, Black, Indian, Asian, Middle Eastern, Hispanic or Latino
- **Age group:** <19, 20-49, >50 years (yrs.) old

Metric	Perceived gender Subgroup	Ethnicity Subgroup	Age group Subgroup
Identity attack	0%	0%	0%
Threat	0.06% Female	0.14% Black	0.17% >50 yrs. old
Insult	0.06% Male	0.09% Asian	0.16% 20-49 yrs. old

REPORTING RISKS

City helpline: 0XXX XXX XXX
Reporting portal: report-risk@com
Mail: XX Main Street, City A
XXX-XXX Contry Z

REGISTERED OFFICE

Name of the company
XX Main Street,
XXX-XXX Contry X

CERTIFICATES



Compliant with the AI
Management System Standard