



FactSheet for AI Systems

This FactSheet provides transparency regarding the Artificial Intelligence (AI) product offered by Rep'd. It details system purpose, training data, model architecture, and risk management strategies in alignment with GovAI Coalition standards.

1. System Overview

Vendor Name	Rep'd
System Name	Rep'd
Overview	Rep'd is an AI-powered video platform for local government. We provide municipalities with an AI-powered Q&A system that answers resident questions instantly through pre-recorded staff videos and generative AI text responses. All sources are immediately visible to residents when a question is answered via AI-generated text, and every video is reviewable by staff before being published or shared with residents.
Purpose	Rep'd helps municipalities: • Build Trust: Rep'd "humanizes" local government by introducing names, faces, and context to sensitive or complex issues. • Stop Misinformation: Rep'd provides local government with a megaphone to correct misinformation before it spreads. • Save Time: AI answers simple resident questions without staff involvement, and prevents misunderstandings from devolving into time-consuming crises.
Intended Domain	Improving local government communications by way of saving municipality staff time while providing critical context and a human tone to complex or sensitive local issues.

2. Technical Specifications & Data

Training Data	Our AI systems are bespoke for each municipal partner. Each system is trained via a combination of publicly-available municipal website URLs, domains, and documents. All sources are identified, approved, and provided to us by our municipal partners and are legally obtained. We add data to the training set as instructed by our municipal partners. Most partners provide the majority of source material upfront to minimize the need for constant retraining.
Test Data	We run extensive testing on every aspect of the platform, including AI-powered translation, transcription, sentiment identification, and automated categorization. Testing is typically conducted in a private setting using specific source material provided by municipal partners (websites, domains, documents) prior to launch.
Model Information	The AI system utilizes a multi-model architecture including: • AWS: Neural machine translation (AWS Translate) and speech-to-text processing (AWS Transcribe) for transcription and translation. • OpenAI: Hybrid model approach using GPT-4o for primary answer generation and quality assessment, with GPT-3.5-turbo for sentiment analysis, question categorization, and content classification. Text embeddings use text-embedding-3-small.
Update Procedure	Municipal data is continuously updated through automated scraping processes that parse sitemaps, crawl website content, and process documents (PDFs, Excel files). New content is chunked and embedded into client-specific vector databases. OpenAI and AWS models are managed via API versioning with no direct client control. Staff can review and override AI responses before publication.

Inputs and Outputs	Inputs: Text (user questions), video/audio files (for transcription), municipal website URLs, documents (PDFs, Excel files), and structured government data. Outputs: AI-generated text answers with source citations, transcribed speech, translated text, sentiment scores (1–3 scale), question categories, confidence scores (0–1 scale with good/bad grades), and social media content.
---------------------------	--

3. Performance & Risk Management

Performance Metrics	• Confidence Scores: AI-generated responses receive confidence scores (0–1 scale) with "good"/"bad" grade classifications. • Vector Similarity: Semantic search uses cosine similarity scores for source matching and relevance ranking. • Token Usage: API token consumption tracked for cost management. • Source Traceability: All answers include URL citations to original sources. • User Feedback: Thumbs up/down mechanisms with optional rejection notes stored for quality improvement. • Monitoring: Continuous API response validation and manual audits.
Bias Mitigation	Translation and transcription biases are addressed through human review. Sentiment analysis uses standardized prompts with GPT-3.5-turbo. Government data retrieval is strictly limited to factual, sourced information from official municipal sources to avoid misrepresentation.
Robustness	The system flags low confidence scores for human review via feedback mechanisms (thumbs up/down). Deep search functionality performs independent web verification for rejected answers. Continuous monitoring and manual validation enhance system robustness.
Optimal Conditions	Operates optimally with structured inputs, well-formatted text, and clear audio. High-quality structured source material significantly improves retrieval accuracy.
Poor Conditions	Performance declines with noisy audio or ambiguous text. Translation accuracy may lower for highly technical idioms. Hallucinations are minimized via RAG , but may occur if government source data lacks relevant information.
Explanation & Transparency	Each AI response includes confidence scores and source traceability. All Government data retrieval outputs link back to original official sources , ensuring transparency. Logic is continuously refined for user comprehension.

Jurisdiction & Data Protection	Data is processed in compliance with applicable privacy regulations. AWS services (Translate, Transcribe) and OpenAI APIs are used with appropriate data handling practices. Municipal data is stored in client-specific database schemas for data isolation. Security protocols safeguard sensitive information across all processes.
---	--

4. Impact Assessment

Monitoring & Human Oversight	AI-generated responses are logged and periodically reviewed. Human reviewers can override outputs. While overrides do not currently retrain models automatically, they are used in manual audits to calibrate future system parameters.
Bias Management	AWS and OpenAI apply built-in mitigation. We rely on verified government sources for data retrieval to minimize misinformation. Sentiment analysis and categorization use standardized prompts to ensure consistency across different question types.
Flagging Issues	Users can report AI performance issues through integrated feedback mechanisms. Reports are manually reviewed, and adjustments are made to filters and overrides.