

AI FactSheet – Madison AI

Vendor Name	Madison AI
System Name	Madison AI
Overview	Internal, Closed, Custom LLM
Purpose	<ol style="list-style-type: none">1. Internal knowledge assistant for local governments based on the public record.2. Report writer using this same information for such reports as staff reports and fiscal notes.
Intended Domain	<p>Madison AI serves as an internal knowledge assistant specifically designed for local government operations. Specifically, Madison AI is an internal, closed LLM. We have developed specific knowledge domains where the data source is the jurisdictions data in a closed, dedicated index. We have specific models for the following knowledge domains:</p> <ul style="list-style-type: none">• Agency Governance & Management• Planning & Zoning• Budgeting & Finance• Internal Policies <p>Broadly, the first difference between Madison AI and MS Co-Pilot is Madison AI is ingesting, converting, indexing and RAGing thousands of files. This matters because AI does not read many file types. In order to achieve high accuracy and recall, the agency’s data must be clean – which is the first step of our process. Then, we have created specific data retrieval and processing agents specific to local government use cases – like a complete voting history on a policy, code review, and board/staff report/memo drafting. All of these pieces are specific to the agency instead of an off-the-shelf solution, such as Co-Pilot & GPT. That said, we are using the Azure Microsoft components as the foundation of Madison AI.</p>
Training Data	Each Madison AI customer has its own dataset (or multiple datasets), created in collaboration with the agency’s AI Team. Usually, the dataset consists of local government’s public records, master plans, strategic plans, as well as zoning and code documents.

Test Data	There is no test data as this is not a fine-tuned model.
Model Information	Madison AI leverages a large language model (LLM) architecture, specifically utilizing a variant based on the GPT-4.1 model.
Update procedure	The source data set is updated monthly to ensure Madison AI has access to the most current and relevant information.
Inputs and Outputs	<p>Data Inputs:</p> <ul style="list-style-type: none"> - Base data set is provided by the agency, or we create the dataset in partnership with the agency. This dataset is refreshed weekly, through automated scripts. - Users are able to upload their own files to use in their work. Those files are temporarily stored for 1 week. <p>User Query Inputs:</p> <ul style="list-style-type: none"> - User inquiries or questions related to the source data, submitted through the system's interface. <p>Outputs:</p> <ul style="list-style-type: none"> - Responses generated by the knowledge assistant based on user inquiries or questions. - Staff reports generated based on user-provided input, leveraging the same dataset used by the knowledge assistant.
Performance Metrics	<p>RAG applications have standard performance measures, but not necessarily targets, yet. We have established our own baseline and targets noted below. The current numbers are as of 1/1/25.</p> <ol style="list-style-type: none"> 1. Faithfulness: This checks how reliable the solution is in avoiding false or made-up information. In other words, it ensures the answers are based on real facts and knowledge. Current: 83% Target: 96% 2. Precision: This assesses whether the solution provides answers that are not only correct and accurate but also aligned with what the question specifically asks for. Current: 79.5% Target: 90% 3. Recall: This measures how much relevant information from all the available facts is successfully included in the answers, ensuring important details aren't left out. Current: 89.4% Target: 90%

Bias

Madison AI leverages OpenAI's GPT-4.1 model, which has undergone extensive efforts to mitigate biases related to gender, race, socioeconomic status, disability, culture, age, or other protected classes. However, the system's responses are grounded in the source documents it accesses local government public records, master plans, strategic plans, and zoning/code documents. These documents may contain inherent biases reflective of historical and societal contexts.

The Madison AI RAG methodology is built to ingest and process thousands of files in the data retrieval process. By doing this, the AI's response is built on numerous cross-referenced files. We also prioritize the official record, such as laws, code, the voting record. We are not stopping here, but this is the current state as of Jan 2025.

Robustness

Madison AI is designed to prioritize the most recent data updates, ensuring it uses the latest information for generating responses. Users can provide feedback through a thumbs up/down mechanism, specifying issues like missing citations or irrelevance, which is used to improve the system over time. Monthly dataset updates ensure ongoing accuracy and relevance, and the system's logic helps manage data redundancy effectively. Although the model itself isn't fine-tuned based on user input, feedback is utilized to enhance search algorithms and overall system reliability, ensuring robust handling of various data and user interactions.

The user feedback is stored in a Cosmos DB. Once we have enough user feedback from that specific jurisdiction (each jurisdiction has a dedicated environment, so it is just that agency's feedback) we aggregate the data and use that processed feedback to update the RAG logic.

Optimal Conditions

Madison AI performs best when provided with clear, accurate, and up-to-date source data. While there is no stringent requirement for the quantity of records/observations, higher data quality can enhance the system's performance.

Conditions that favor optimal performance include:

1. Well-structured and comprehensive documents.
2. Regular updates to ensure the latest information is accessible.
3. Clear and specific user queries.

Poor Conditions

Madison AI will perform poorly if the data is of very low quality, meaning it does not contain the relevant information the user is looking for or the information is structured in a way that it loses contextual meaning.

The agent will return "This information is not contained in the documents available to me."

Explanation	Madison AI provides explainability through context-centric outputs, generating responses grounded in original source documents like local government public records and strategic plans. It includes citations whenever possible, allowing users to trace the origin of information, and integrates user feedback via a thumbs up/down mechanism to enhance performance. Outputs are consistently formatted to meet the expectations of local government officials, ensuring usability. Human subject matter experts (SMEs) can review outputs to provide additional context or corrections. Designed for ease of understanding, the generated responses are accessible to individuals with varying expertise levels, thereby ensuring transparency, justifiability, and comprehensibility of Madison AI's predictions and outputs.
--------------------	---

Jurisdiction-specific Considerations	N/A
---	-----

Algorithmic Impact 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 system tracks the responses of the users to identify correct responses and incorrect responses. Those responses are used to train the system going forward, by user. These responses can be used to improve the system in the future if desired.
---	--

How is bias managed effectively?	Metrics for monitoring and improving bias are currently in production with the University of Nevada Reno.
----------------------------------	---

Have the vendors or an independent party conducted a study on the bias, accuracy, or disparate impact of the system? If yes, can the Agency review the study? Include methodology and results.	No, not yet.
--	--------------

How can the Agency and its partners flag issues related to bias, discrimination, or poor performance of the AI system?

The system provides the ability for the user to vote up or down the response provided. These responses are saved and are used to improve the search going forward.

For the responses that are down voted, the application asks why the user voted it down, with several choices (depending on the content). One of these options is related to bias as well as performance. We review this information quarterly, at the moment and this is included in the improvements to the model at an aggregate level.

How has the Human-Computer Interaction aspect of the AI tool been made accessible, such as to people with disabilities?

The system is embedded into a graphics user interface that is compliant with modern screen readers.

Please share any relevant information, links, or resources regarding your organization's responsible AI strategy.

Information about our responsible AI strategy is listed below:

Madison AI applies Responsible AI Principles using this functionality:

1. **Accuracy** is achieved by **only** using your data set.
2. **Completeness** of results by engineering long-form response.
3. **Full citation** of every data point to eliminate "AI drift" & hallucinations.
4. **Prioritization** of master docs and policies to "weight" authority content.
5. **Error checking** against relevant "master documents."
6. **Data is current** through monthly reindexing.