

WoT AI Services and Gaps

Michael McCool
Intel Corporation

AI Services

- Ollama LLM Service
 - JSON object in
 - JSON object(**s**) out (***optional streaming – not supported in WoT TDs!***)
 - Fortunately streaming not essential to access functionality
- Piper Text-to-Speech Service
 - Text in (plain; URL variable or in POST body, NOT in JSON)
 - Audio out (WAV file, binary)
- Whisper Speech-to-Text (transcription) Service
 - Audio in (WAV, binary) but also other arguments (language, optional)
 - Multipart/form-data: ***not supported in WoT TDs!***

<https://github.com/w3c/wot-testing/tree/main/events/2024.11.Munich/Documentation/Intel>

Multipart/form-data

- Needed for Whisper
 - Technically, “faster-whisper” service
 - <https://github.com/fedirz/faster-whisper-server>
- Send audio/wav and text/plain (for language)
 - <https://github.com/w3c/wot-testing/blob/main/events/2024.11.Munich/TDs/Intel/whisper.td.jsonld>
 - ***Can't fully describe API in WoT TD!***
- Suggested resolution here:
 - <https://github.com/w3c/wot-thing-description/issues/1464>

JSON Object Streams

- Ollama API
 - Streaming JSON objects for real-time updates
 - Has a “done” flag in the last object
 - Not technically needed, closing connection signals end, too.
 - Objects separated by newlines
 - Objects all share data schema, same as single-object schema
 - Also optional, have “stream” boolean flag in request
 - Variable/dependent output data schema?
 - <https://github.com/w3c/wot-testing/blob/main/events/2024.11.Munich/TDs/Intel/ollama.td.jsonld>
 - Models <https://github.com/ollama/ollama/blob/main/docs/api.md>
- Added comment to this issue for followup:
 - <https://github.com/w3c/wot-thing-description/issues/1044>
- Fortunately, we don’t really NEED to use streaming in this API...
- But might be essential in other use cases
 - Streaming sensor updates
 - Incremental TTS output for Piper TTS service (e.g. sentence-by-sentence).

Other Observations

- Using MediaStream API in a web site *requires* https
 - Access to the microphone is considered a security risk
 - Option to allow it in insecure contexts in Chrome no longer works!
 - I spent *way* too much time finding this out...
 - <https://mdn.github.io/dom-examples/media/web-dictaphone/>
 - <https://developer.mozilla.org/en-US/docs/Web/API/MediaStream>
- Also experimented with Frigate
 - Real-time object identification in camera streams
- We still need to figure out how to use
 - Websockets (alternative to JSON object streaming, but for binary data)
 - WebRTC (IP cameras and other media streams)
 - RTSP (for IP cameras)
 - Reusable Connections (pending feature) may help here