WoT AI Services and Gaps

Michael McCool

Intel Corporation

AI Services

- Ollama LLM Service
  - JSON in
  - JSON out (***optional streaming – not supported in WoT!***)

- 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!***

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

Multipart/form-data

- Needed for Whisper

- Send audio/wav and text/plain (for language)

- 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
  - Objects separated by newlines
  - Also optional, have "stream" flag in request
    - Means would need a variable/dependent output data schema

Other Observations

- Using Media APIs in a web site *requires* https
  - Apparently access to the mic 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...
- Also experimented with Frigate
  - Real-time object identification in camera streams
- We still need to figure out how to use
  - Websockets
  - WebRTC
  - RTSP (for cameras)
  - Reusable Connections (pending feature) may help here