26th International Congress on Modelling and Simulation, Adelaide, SA, Australia, 30 Nov–4 Dec 2025
mssanz.org.au/modsim2025

Does a water policy paper plus the Self-Thinking Data Manifest (STDM) equal an educational experience?

David J Penton [a], **Benjamin Leighton** [b]

[a] *CSIRO Environment Research Unit, Canberra, Australia,* [b] *CSIRO Environment Research Unit, Melbourne, Australia*
Email: dave.penton@csiro.au

Abstract: We investigate whether it is possible to write a journal paper that behaves as an interactive educational experience when users import the paper into an AI system such as Copilot or Claude. First, we append human language instructions in the form of a self-thinking data manifest (STDM) to a journal paper that explores how gender is incorporated in water policies (Cuddy et al, 2025). Then, we execute the appended instructions in three AI systems (Copilot, Claude, Google AI studio) and discuss the outcomes. The results are fascinating, but highly variable, with the results depending on the input instructions and the AI system.

STDM (https://github.com/csiro/stdm) is a standard for attaching instructions to documents for interpretation by large language models such as GPT4 or Gemini. First a user loads a STDM compliant document in an AI system. Then, the user prompts the AI system to, "follow the instructions". The document provides context and instructions to the large language model. It acts not just as static content but it also influences how the AI system interprets user prompts. The approach is novel – we are aware of only 2 other similar approaches: Agents.md, which provides instructions for coding AI agents (see agents.md); and Markdown for Agent Guidance & Instruction, which uses a separate document to provide machine readable instructions for AI agents (see magi-mda.org).

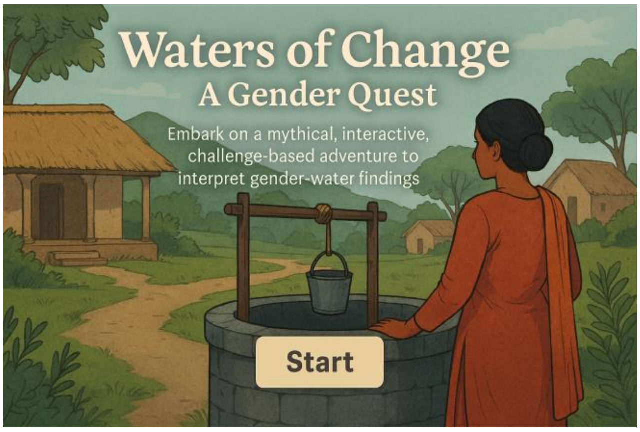

Figure 1. The starting image of the interactive game generated by Microsoft Copilot

To explore STDM we attempted to create a graphic adventure game with a text-based interface driven by the paper content and an embedded prompt. We created a 2-page annex to the paper in natural language that outlined the gameplay including number of levels, type of role play, images at the start of each level, etc. After considerable testing and refinement, Microsoft Copilot would invite a user to ask questions of characters and respond with realistic scenarios and content from the paper (e.g. women and children as primary water collectors in rural Nepal).

The stochastic nature of large language models meant that every time we loaded the paper, we would get different results. For example, one time we loaded the paper the game had an image at the start of every level, but another time we loaded the paper there was no image for second or subsequent levels. We also found that different AI systems interpreted the instructions differently: Claude would create code for a web-app; ChatGPT 5 (free) would not find the annex; and Google AI Studio (gemini pro) would provide a game without images. While promising, future work will need to address the challenge of providing consistent user experiences across different AI systems.

References

Cuddy SM, Koirala S, Wahid S & Penton DJ (2025) Measuring How Water-Related Policies of the Global South Consider Gender: Insights from Trialling a New Policy Gender Index in Nepal. *Asia & the Pacific Policy Studies*, *12*(3), e70029.

Keywords: *Large Language Models, Self-Thinking Data Manifest, Water, Policy, Gender*