Exploring AI to summarize evidence
WFP’s experience (so far…)

Aurélie Larmoyer, WFP Office of Evaluation
Background for my experience

**At the start was a challenge...**

Some people wish to receive evidence that can be directly useful to them.

We can recycle the evidence that exists, and repurpose it, BUT....

Retrieving, exploring or summarizing pieces of evidence manually is **slow and costly** and allows considering only subsets of evidence, as the human eye can grasp.

**AI a solution?**

We are exploring AI to develop an automated search solution tailored to generate insights from existing evaluation evidence.

➔ **STEP 1:** Extraction of relevant text passages from searching through all of our “universe” (published reports)

➔ **STEP 2:** Generate new text from the extracted text passages, to provide insights
Off-the-Shelf Apps: limitations and risks

Limit on size of the universe to explore (max 10 MB / 5 reports)

Accuracy of response: goes to the closest – risk of misinterpretations!

Data protection: where does the data go?
Off-the-Shelf Apps: limitations and risks

Limit on size of the universe to explore (2000 words at once)

Accuracy of response: asking 100 words -> gives me 148...

Bing Chat Enterprise (Open AI / Chat GPT)
Most large language models (LLMs) – behind Google’s “Gemini” or Open AI’s “GPT” – reflect ‘readings’ of the world that

1) Require substantial ‘fine-tuning’ to match our international development lingo – e.g. ‘Social protection’

2) Are biased towards the audiences of these searches over the period it has existed, i.e. channel ‘Western’ perspectives

3) May imply undue costs compared to the need - as when using a Ferrari to go grocery shopping...