At the nadir of the cost of living crisis, millions of people in the UK struggled to access essentials such as food, warmth, and electricity. This surge in humanitarian need prompted the British Red Cross (BRC) to rapidly develop multiple new forms of aid, as well as scaling up existing support such as emergency cash assistance.

As the nation grappled with the devastating consequences of a decade of austerity policies, a pandemic, and the escalation of the conflict in Ukraine, BRC's senior leaders faced a daunting task. They needed to understand how the cost of living crisis might evolve over the coming years to make informed decisions about the levels and kinds of support the organisation would need to provide. However, time was of the essence, and conducting rigorous scenario development with the participation of affected communities was not an option.

Faced with this challenge, I turned to an unconventional group of virtual helpers: AI chatbots. Generative AI had already demonstrated impressive performance across various tasks, including strategy design, research, summarising evidence, identifying gaps in knowledge, and data analysis. These also happen to be some of the core skills for designing scenarios about how the future might unfold.

Drawing upon the framework developed by IARAN (the Inter-Agency Research and Analysis Network, a humanitarian foresight collective), I embarked on a series of experiments to test how generative AI tools could augment and enhance the scenario development process. The framework involved identifying drivers of change, judging their impact and uncertainty, formulating hypotheses, and ultimately writing scenarios based on selected hypotheses.

With the help of ChatGPT, I explored questions to ensure a comprehensive understanding of the system's dynamics. My team had already been monitoring and analysing the cost of living crisis, so mapping the architecture was a fairly quick job — but it was helpful to have ChatGPT on hand so I could explore questions such as, “What are the key drivers of living standards?” or, “What are the drivers of change for the cost of living?” to see if I had missed anything obvious. Even simple questions such as “What causes prices to rise?” can give helpful ideas to inspire deeper research.



Scoring the impact and uncertainty of each driver of change was a particularly interesting part of this experiment. I simulated how we would do this in a workshop with actual humans by engaging a diversity of AI chatbots: ChatGPT, Bing (which uses the same underlying large language model as ChatGPT), Perplexity, Google Bard (as it was known at the time), and Claude.

After a few rounds of prompt engineering, I convinced the chatbots to return structured output: tables of impact and uncertainty scores that I could manually analyse in Excel to aggregate the scores into a single impact-uncertainty matrix.



The quality of the AI-generated answers was surprisingly good. There were a handful of cases where an AI made odd choices — for example, Perplexity categorised unemployment as having a low impact on the cost of living crisis, which I (and the other AIs) disagreed with. Other than that, the scores were sensible and varied in useful ways across chatbots, prompting me to test my own assumptions and biases.

Generative AI models struggled with the next steps — incidentally, those requiring the most creativity: generating mutually exclusive hypotheses for how the high-impact, high-uncertainty drivers of change might evolve over the coming years, then writing short stories to illustrate each scenario. I mostly ended up writing the hypotheses and scenarios manually, although ChatGPT provided some support in thinking through options for scenarios.

Despite a few hiccups and disappointments, bringing together multiple generative AI tools greatly sped up the generation of a plausible set of scenarios for how the cost of living crisis could evolve in the UK over the coming half-decade, supporting rapid decision-making for BRC’s leadership.

**Avoiding bias and checking reliability**

Much of the bias-reduction in this experiment took place through my own subject-matter knowledge, having led deep-dive research and analysis into the cost of living crisis, including its drivers and humanitarian impacts. This research included learning from a series of focus groups with people experiencing poverty, which was the closest we could get to meaningfully involving people under such tight time and budget constraints.

**Addressing ethical concerns**

In this instance, the main ethical concern was around the lack of meaningful participation for people with lived experience of the issue at hand, as well as not being able to involve frontline volunteers and staff. At its best, foresight brings together diverse voices and perspectives to imagine ‘what if’ and collaboratively design desirable futures — which this experiment lacked.

When telling others about my experiment, I make it clear that this AI-augmented approach is no substitute for deeply engaging people in the process of strategic foresight. The value of scenario planning lies in including people in the development of scenarios rather than just presenting them with a finished product. AI, in this instance, can merely be a tool to help us think through possibilities.

Since this experiment only relied on publicly available evidence and data, there were no concerns around data privacy or similar risks that required mitigation.

**Impacts**

In the end, the scenarios were used for shorter-term decision-making; British Red Cross leadership ultimately decided to wind down our cost-of-living support. However, this AI-augmented approach to strategic foresight has garnered a great deal of interest within The Movement as well as among external foresight practitioners.

I wrote up my approach and learnings from this experiment in a blog post shared on medium.com (which, to date, has 266 views and 199 reads). My piece was republished on IARAN’s website, and has since been shared multiple times on LinkedIn. I also presented this experiment to members of the IFRC’s AI community of practice (which I also co-facilitate).

**What I learnt and what next**

Prompting well is vital. Although leveraging multiple generative AI models helped me deliver a decent set of scenarios rapidly, I still had to test and iterate several forms of prompts to get useful results. This pilot also required multiple manual steps that could now be automated, especially now that several large language models can return structured output, such as JSON.

A more comprehensive version of this would encapsulate all steps in an end-to-end manner, such as through a web app — which I have begun to design (and will share the code as well as the app itself publicly).

Since originally experimenting with this approach in July 2023, open models now possess the power that closed models such as ChatGPT and Claude demonstrated at the time. And AI models on the whole provide higher-quality responses across a range of domains. I am planning to test my approach again with these newer, more powerful models. And with a greater range of models now available, we have access to a greater diversity of AI ‘thought’ and ‘experience’.

Large language models have also become adept tool-users in the time since I conducted this experiment. They can now plug in to sources such as search engines, encyclopaedias, and media outlets, as well as generate content from custom knowledge bases — improving their accuracy and factual correctness. Many AI models have become expert computer programmers; one thing I am currently experimenting with is how we could use large language models to rapidly develop quantitative forecasts as well as qualitative scenarios.

Finally, I am exploring the potential for generative AI to support another area of strategic foresight: monitoring weak signals of change. These are early indicators of emerging trends or potentially significant changes that are currently less well-understood (for example, virtual citizenship). Leveraging generative AI models and plugging them into various sources (e.g. news, research journals, social media, etc.) could help foresight practitioners make sense of a deluge of information. Weak signals of change can also come from sources such as artists, musicians and filmmakers; multimodal AI models that can reason over images, video and audio, as well as text, could offer interesting avenues to further explore.

As we navigate the complexities of an ever-changing world, the integration of AI into strategic foresight offers a promising avenue for humanitarians to rapidly adapt and respond to emerging challenges. By harnessing the power of generative AI - and bringing it together into a collaboration with the expertise of affected communities and humanitarian responders - we can enhance our ability to anticipate and prepare for the future, ultimately enabling us to better serve those in need during times of crisis.