Excerpt – GovGPT FAQs

What are the security, privacy and accuracy considerations?

Security, privacy and trust are top considerations for GovGPT. Defined, publicly facing websites are indexed to GovGPT's model, which are then ringfenced as a technical barrier to containerise the information GovGPT retrieves when forming its output. Indexing and ringfencing are mitigations for hallucinations that also support high accuracy outputs.

The model **does not** collect or store any data; and a pop-up disclaimer will be mandatory on the site to have users acknowledge the limitations of AI models and guidance to not input personally identifiable information.

An AI Impact Assessment has been conducted to ensure responsible development and governance of the model.

GPT-4o comes with an advanced set of content filtering for themes such as hate, violence, harm and sex. Additionally, GovGPT has been programmed to offer direction to 111 and/or Lifeline if it detects user attempts to chat across these themes.

GovGPT and the base model 4o have extensive security testing, threat monitoring and other measures that can be viewed in detail at /www.openai.com/gpt-4o-system-card/.