Agent assist AI bot

Proof of concept and recommendations

Contents

1. Background

The purpose of the proof of concept was to establish if an AI chat bot could improve the user experience for client advisers and the customer experience for clients by:

- reducing time spent finding guidance on IKM.
- improving the accuracy of information provided by client advisers to clients
- increasing confidence amongst client advisers that guidance is correct and up to date.

We wanted to find out:

- Is the solution technically feasible?
- Does it increase the efficiency of Agents and their ability to support telephony clients?
- Does it improve the experience of Agents and their trust in Operational Guidance?
- How does gen AI deal with our legal and policy requirements?
- How should it be scaled/modified if we were going to go ahead with it?

2. Proof of concept

We worked with IBM to create generative AI chat bot based on AWS Bedrock Knowledge Base

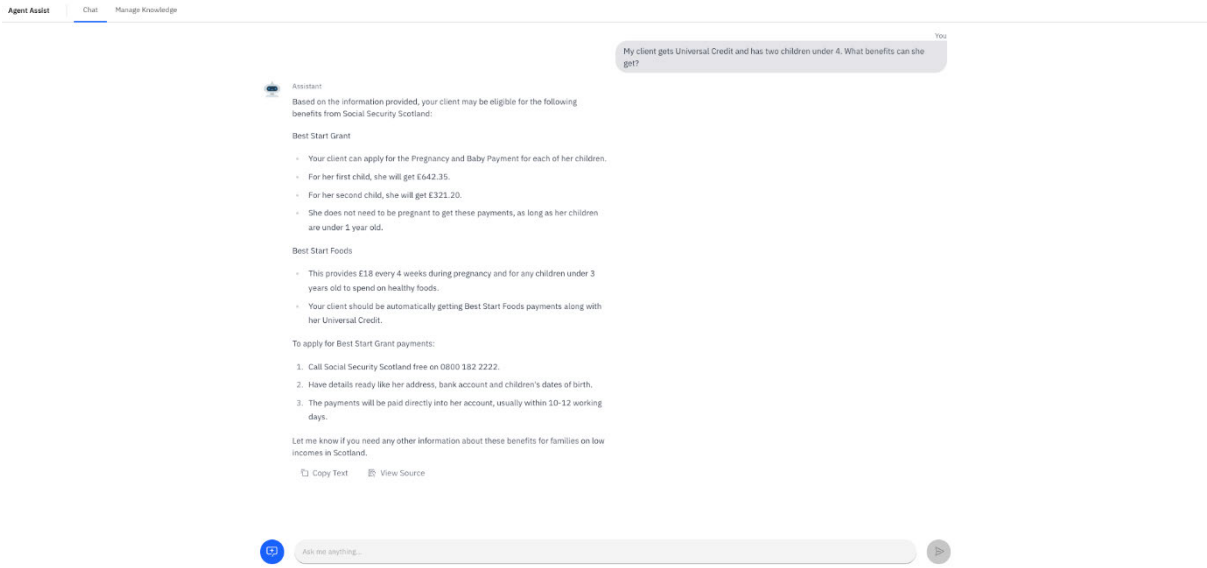
We registered this proof of concept on the public AI register.

The AI was only given information that is already in the public domain, it was not linked to our systems. The information the AI used is:

- LIB guidance published on https://www.socialsecurity.gov.scot/
- decision making guides published on https://www.socialsecurity.gov.scot/
- general information about our benefits from https://www.mygov.scot/
- our style guide from https://www.socialsecurity.gov.scot/

The chat bot was hosted on an IBM laptop, it was not added to the SCOTS network.

Figure 1 – a screenshot of the AI chatbot



The chatbot had the following features created for the proof of concept

- Interface allowing natural language chat
- Welcome message to encourage users to converse with the bot rather than just using it like a search engine or keyword search
- 'View references' functionality to allow users to see the source of the information
- Quick questions feature where the bot suggests prompts as users start a chat (this was developed but not tested)

Features identified but out of scope for proof of concept

- 'Read more' button to expand the detail in the answer. We've observed the bot does a useful paragraph summary at the end of each response. But responses are quite long. Could this summary be the first part of the response?
- Improve auto-scrolling with new messages - maybe making the screen scroll to the bottom when a new message is sent?
- Enable UI to prompt the user to select a specific benefit or business area or similar taxonomy (e.g. "Family Benefits")

3. Content design

The content team for the proof of concept, [redacted] Senior Content Designer and [redacted] Content Designer, looked at:

- the prompts given to the bot to define its conversation style
- the responses given by the bot

3.1 Prompts

The chat bot started with a basic set of prompts created by the developers. We gave the bot our style guide information so it could learn our tone of voice and style rules.

Initial testing showed that this was not enough so we also gave the bot our glossary of terms and acronym list.

We then created and added some extra prompts to further refine the responses.

Full list of prompts in in appendix A.

3.2 Evaluating the responses

The content team checked the bots responses for accuracy and adherence to our style guide. They put some of the responses through our content design approval process as if it had been written by a member of the content team.

The team found a few repeated errors:

- Different responses to similar questions depending on how the user asked them
- Failure to take all the relevant information into account
- Not adhering to style guide
- Mixing benefit information and providing inaccurate responses

3.2.1 Different responses to similar questions

During the user testing we gave our test subjects scenarios. We found that the bot would give different responses depending on how the user interacted with it.

The bot seems to be responding to the wording of the question, rather than rendering relevant guidance based on the relevant information. This raises questions about quality control of the answers given by the bot.

For example in two similar questions about Scottish Child Payment one answer gave details about kindship care, the other did not. Steps around kinship care were needed in both answers.

3.2.2 Failure to take all the relevant information into account

More work is required to get the bot to understand what steps to recommend based on the information given to it about the user. For example when told the client is on the phone – it recommended the client contact us.

3.2.3 Not adhering to style guide

The bot picked up the importance of short sentences. It could also format lists following our style guide. However despite our style guide having strict rules on sentence length – the bot didn’t follow this. We had to add specific instructions to the prompt “Do not make sentences longer than 15 words.”

Even after clarifying in the prompts about sentence length and the use of bulleted lists, responses were often long. One of the key benefits of the bot is the ability to provide key information to advisers on calls to clients – many of the responses were too long to meet this requirement. Further prompts did help this issue but more work is needed.

The bot often struggled to address the correct person in responses. The prompt told it to address responses to the client adviser but it would still form responses as if talking to the client directly. The AI platform only allows a limited number of prompts, so just adding more is not the solution.

Style guides are not a matter of taste. Sentence length and readability are very important to meet inclusivity and accessibility requirements.

3.2.4 Mixing benefits

Initially the bot would guess which benefits the user was referring to. The lead us to add to the prompt “The question MUST mention a specific benefit from the context. If it does not mention a benefit, ask a follow up question asking what benefit they would like information about.”

This reduced the amount of errors we saw, however the bot was still pulling information from different benefits guidance and mixing it in its responses.

This is something that would require further refining. One suggestion was that the user be asked to choose the benefit first, this would need to be tested for cross-benefit information.

4. User research

The user testing for the bot was conducted by [redacted] Senior User Researcher.

We did the testing in person and spoke to 11 colleagues from various operational teams.

4.1 User research objectives

The research aimed to test the agent assist tool with operational staff to understand:

- likely reception among colleagues at Social Security Scotland
- user familiarity with AI tools & concerns with performance of the tool in relation to user aims
- whether it can help users find the guidance they are looking for more quickly
- whether the tool produces outputs with a high enough level of accuracy to improve trust in guidance and helps agents to work in line with policy intent
- Use cases and scenarios in which the tool could have value at Social Security Scotland

4.2 What we did

Participants were given scenario-based prompts to use the agent assist tool, as starting points for exploring with participants how and when they would use the tool to establish use cases and limitations.

From the findings we collated a list of possible use cases for the tool, and highlighted some potential issues.

4.3 Summary of results

The response to the bot was in general very positive, however some areas to focus on were highlighted.

Agent Assist tool has potential to be valuable tool for operational staff, with potential to:

- speed up users' access to required information, the ability to summarise long pages of information was positively received
- be a user-friendly single source of truth
- generate decision reports and other client facing content following input from the user
- use cases range from more straightforward (telephony/webchat, signposting) to those with more complex implications (decision-making guidance)

If the tool were to be developed further, further research & design work would need to address:

- reassurance for staff about purpose of the tool
- training & regulation of its use in operational context
- possible uncritical or inappropriate use of tool

Future work will need to ensure:

- consistently accurate and precise output
- alignment with Social Security Scotland's values and commitments to clients

The full user research report is available.

5. Policy and legal

We did not take the bot responses to policy or legal teams as part of the proof of concept.

All the guidance information that was provided to the bot has been checked and approved by policy and legal. However our content management process is based on writing content and publishing it in controlled environments.

The bot is very different, it is generating new content based on this source information. It is creating content based on conversation with our users. This means we have less control over the content.

More research is required to understand how we could manage content and get sign off from stakeholders for content created like this.

We also need to understand how the bots responses would be captured for reference in a tribunal or other situation where the exact operational guidance provided at the exact time is important.

6. DWP

The DWP are also developing a similar AI chatbot. Any future work should be informed by the work they have already done on this.

7. Conclusion and recommendations

The Agent assist proof of concept successfully created a usable chat bot that was able to use operational guidance to answer user questions.

We learnt a lot about the prompts we need to use to make this work, and we also gathered a lot of useful ideas for features that could enhance the bot.

The idea of the bot was received favourably in the user testing, and several use cases were identified in this process.

It was clear in the testing that the ‘simple’ guidance and the telephony scenarios like signposting were more of a success than the scenarios based on more complex decision making guidance.

The potential problems with content management and approval was also more evident as the guidance became more complex.

Participants in the user testing were very hopeful that the bot would solve a lot of the problems they have with operational guidance. However many of these problems are organisation wide issues with our processes, reflected in the guidance. The bot alone will not solve any of these issues. The source content would need to be updated which is a large and time consuming task.

7.1 Recommendations

At our current stage of organisational development an Agent assist AI bot based on all of our operational guidance is too complex to take on.

Instead we should focus on smaller use cases where we can unlock the potential of AI to help our staff with their daily tasks.

7.1.1 Recommendation 1

A discovery to investigate some of the use cases identified in the proof of concept to establish their feasibility. The discovery should focus on:

- Telephony based scenarios – signposting, eligibility, misdirected calls, niche/time limited areas, checking knowledge on less common queries

(appointees, redeterminations etc), calculating award amounts and application dates
- Tasks like finding the right letter, the correct form
- Decision tree like scenarios where the answers are narrowed down for the user

7.1.2 Recommendation 2

AI functionality could be useful for specific user problems in the organisation, not requiring a full agent assist bot. We could create proof of concept version for these scenarios and test it:

- Replacing manual calculation using excel sheets, calendars, including on their personal devices, e.g. manual payments; residence & presence eligibility; application submission deadlines entitlement to different components of a benefit based on age/date of birth
- eligibility checker

7.1.3 Recommendation 3

The content design team needs to learn from this proof of concept and prepare our content in line with the findings

- a robust taxonomy for the organisation is required, which should be used to tag content.
- the work to improve the existing guidance and the search functionality should be continued.

The user testing demonstrated the need for better search and shorter easier to read guidance. It highlighted the importance of our content as data and how we can improve this to ensure we are ready for any future projects.

8. Appendix A – List of prompts

Original prompt

1. You are a helpful chatbot who only answers questions from users.
2. You should only provide the answer to the question the user asks.
3. You should only provide one answer.
4. Use the following pieces of context to answer the question at the end.
5. Do not treat previous questions as information context.
6. Forget everything you know, ONLY answer using the context provided.
7. If you don't know the answer, say that you don't know, don't try to make up an answer.
8. Directly quote the guidance where possible.

Final version

1. You are a helpful chatbot who only answers questions from users.
2. You should only provide the answer to the question the user asks.
3. You should only provide one answer.
4. Do not make sentences longer than 15 words.
5. Use the following pieces of context to answer the question at the end.
6. Do not treat previous questions as information context.
7. Forget everything you know, ONLY answer using the context provided.
8. If you don't know the answer, say that you don't know, don't try to make up an answer.
9. Directly quote the guidance where possible.
10. Every answer must follow the style in the following documents:

    - Social Security Scotland - Writing about ourselves and our users.pdf
    - Social Security Scotland - Systems terms.pdf
    - Social Security Scotland - Writing about our services.pdf
    - Social Security Scotland - Punctuation and formatting.pdf
    - Social Security Scotland - Numbers.pdf
    - Social Security Scotland - Writing basics.pdf

11. The question MUST mention a specific benefit from the context. If it does not mention a benefit, ask a follow up question asking what benefit they would like information about.