

## Briefing Document: AI in Behavioural Research Webinar (January 2025)

### 1. Introduction & Webinar Objectives

- **Focus:** The webinar introduces AI and behavioural research, with a specific focus on *generative AI*.
- **Series:** This is the first of a planned series of three sessions. Future sessions will cover:
  - AI-supported data analysis (analytical AI)
  - Responsible use of AI in behavioural research
- **Aims:** Increase understanding of AI and how it can improve behavioural research.
- Introduce specific AI tools for assisting in behavioural research
- Demonstrate how AI can assist with various tasks related to behaviour research and intervention design.
- Explore potential biases in AI and ways to mitigate them

### 2. Generative AI vs. Analytical AI

- **Generative AI:** Creates new content (e.g., text, research studies, images).
- Examples: ChatGPT, Claude, Ellicit, Site, Deep Seek
- Purpose is to generate new content.
- **Analytical AI:** Classifies, reasons with, and makes predictions from existing information.
- Examples: IBM Watson
- Purpose is to provide labels, predictions or decisions such as in fraud detection, weather forecasting, and medical diagnosis.
- **Webinar Focus:** The session focuses primarily on the use of generative AI in behavioural research.

### 3. Historical Context and Technical Aspects of Generative AI

- **AI as an Umbrella:** AI is a broad term encompassing various technologies, many of which are still used today.
- **Logic-Based AI:** Early AI systems (e.g., 1990s) used logic-based approaches, particularly in medicine.
- Key features: Strict rules, absolute truths.
- **Modern Machine Learning:** Contemporary AI, including generative AI, relies on deep neural networks.
- **Deep Neural Networks:** 'Deep' refers to the multiple layers within the network.
- These networks are very large with billions of parameters that are trained.
- Deep neural networks can be used for analytical purposes (e.g. image labelling), but in recent years, new use of these networks have developed, that is generative capabilities.

- Generative AI models convert different types of information into numeric representation that can be used to generate information between languages and modalities.

#### 4. How Generative AI Models Work

- **Model Instruction:** Users instruct the model on how it is supposed to behave, e.g., “you are a helpful assistant”.
- **Context:** This includes user-provided information and prior interactions within the same conversation.
- **Prompt Engineering:** Users can guide models using specific prompts to generate different outputs. For example, by:
  - Requesting a specific length of response (e.g., "one paragraph," "concise").
  - Instructing the model how to process a task e.g. "think step by step".
  - Asking for summaries of previous interactions.
  - Requesting generated examples based on previous interaction.
- **Examples:** The presenters use the example of asking an AI about "behaviour change techniques to support healthy eating" and then refining the prompt to produce a concise bullet point list and then examples of these techniques.
- **New Models:** New models are being developed and released very rapidly and now include models such as DeepSeq.

#### 5. Practical Applications of Generative AI in Behavioural Research

- **General Tasks:** Summarising documents.
- Preparing slides/bullet points for presentations.
- Generating ideas.
- Structuring topics.
- Drafting papers.
- Rewriting text in different styles/voices for different audiences.
- **Specific Tasks:** Thematic analysis of text.
- Scoping reviews.
- Literature searching and reviews.
- Refining methodology.
- **Key Benefit:** A key benefit is its ability to generate ideas that might not occur to a human, saving time and getting projects started.
- **Important Note:** It's critical to have subject matter knowledge to identify errors and vagueness in AI-generated content.

- **Interrater reliability:** Using multiple models to evaluate interrater reliability is useful, with the human still having a vital role to play in ensuring validity.
- **Heat parameter:** General models use a 'heat parameter' which results in variability in output. More focused models that are tuned for research or specific application purposes may be more consistent and helpful for specific use cases.
- **Science-based AI tools:** Tools like Ellicit and Site are more reliable for research purposes than general-use language models, as they are often based on research studies and provide citations.
- **Prompt specificity is key:** It is very important to be very specific in the questions that you ask as it can dramatically change the output that you receive.

## 6. Using AI for Intervention Design

1. **Process:Generate Long List:** Use a general large language model (e.g., ChatGPT, Claude) to generate a broad list of potentially relevant intervention components.
  2. **Select Plausible Items:** Identify items from the long list that seem reasonable or evidence-based.
  3. **Focused Models:** Put those items into science-based language models (e.g. Ellicit, Site) for more refined results.
  4. **Vary Prompts:** Explore different prompts with varying degrees of specificity regarding the behaviour, population and setting.
  5. **Apply to BCW:** Apply the same questions across all the BCW (Behaviour Change Wheel) criteria: Acceptability, practicability, effectiveness, affordability, spillover effects, and equity.
  6. **Human in the loop:** Refine your results based on your experience and knowledge
- **Example:**The presenter provides an example of using this process to identify components for an intervention that is designed to reduce bullying in schools.
  - **Key Benefit:** This approach facilitates rapid scoping reviews, potentially within a morning, that would have previously been infeasible.

## 7. Limitations and Cautions

- **Evidence Base:** General models may include non-evidence-based information, so caution is needed.
- **Accuracy:** Models can produce different responses, requiring multiple iterations and human review to confirm consensus.
- **References:** General language models can struggle to provide accurate references, so it is necessary to use science-based tools.
- **Evolving Technology:** AI is rapidly advancing, with new features and improved performance emerging regularly.
- **Use AI with a critical eye:** Always apply critical reasoning to outputs and use AI as a tool and not a replacement for your own expertise.

- **Ethical Considerations:** There are many potential ethical considerations which will be discussed in the third seminar.

## 8. Addressing Bias and Other Concerns

- **Optimism Bias:** General AI models tend to exhibit optimism, and users can prompt them to provide more skeptical answers.
- **Bias Reflection:** AI models don't introduce bias; they reflect biases already present in their training data.
- **Mitigating Bias:** To mitigate biases you need to:
  - Be aware of biases.
  - Use diverse prompting strategies.
  - Use external sources of information to augment the models
  - Employ human oversight in order to ensure validity of results
- **Carbon Neutrality:** There are two trends in the development of AI, one is to build bigger models to increase performance, the other is to increase performance while reducing the size of the model. The latter is necessary to ensure the technology is not detrimental to the environment.
- **Data Security & Ethics:** Be aware of data protection implications and ethical considerations when using AI tools, especially with sensitive data. Data may be held by corporate companies whose servers are based outside of the jurisdiction that you are working in. In these cases it is necessary to have a secure local installation of generative AI.
- **Journal Requirements:** Journals are beginning to formulate guidelines around the use of AI in submitted research; therefore, it is essential to check journal specific requirements.
- **Affordability:** The cost of using some AI-based tools is increasing and needs to be addressed. However, there are open source versions of these tools that are continually being developed.

## 9. The Future of AI in Behavioural Research

- **Democratization of Knowledge:** AI has the potential to empower a wider range of people by making behavioural science information more accessible.
- **Role of the Behavioural Scientist:** The role of the behavioural scientist will evolve. Instead of being the gatekeeper of knowledge, they will become facilitators in the understanding of the science of behaviour.
- **Synthetic Participants:** There is emerging research in using AI to create synthetic participants to test behavioural interventions. However, these types of participants may not provide representative results. It is necessary to compare synthetic participant responses to real participant responses.

## 10. Key Questions and Takeaways

- **Health records analysis:** AI can be applied to health records to identify discussions about specific health behaviors, but prompts need to be specific.

- **Messaging campaign development:** AI can be used to generate messaging campaigns, but they might need refining.
- **Ethical Committees:** Ethical approval for data may not be covered if you use models that are based outside of your jurisdiction. Therefore you may need to use a secure local installation.
- **Equality, Diversity & Inclusion:** AI can help generate ideas in relation to ED&I but quantifying the health inequalities impact requires the use of analytical AI and human analysis, which will be covered in the next session.
- **Open-Source:** There are open-source versions of AI models that are continually being developed, that may mitigate cost and some data privacy concerns.

## 11. Next Steps & Future Webinars

- **Analytical AI:** The next webinar will focus on analytical AI and machine learning for data analysis, making predictions, and calculating a number.
- **Responsible Use:** The third session will explore ethical considerations, bias mitigation, and the responsible use of AI in behavioural science.
- **Resources:** Slides and additional resources will be available on the BRUK website, and there are plans for a blog post on these topics.