

Responsible Use of AI in Behavioural Research

12th May 2025

Speakers: Janna Hastings, Robert West & Susan Michie



**Behavioural
Research UK**

BR-UK is supported by
the Economic and Social
Research Council
[grant number
ES/Y001044/1]

Third of three webinars on AI and behavioural research

Using AI to improve behavioural research

Recording from Jan 28th, 2025



WATCH AGAIN

Using AI to improve behavioural research (generative AI focus)

Recording from March 12th, 2025



WATCH AGAIN

Introduction to the webinar & the topic

Susan Michie

Session overview

Introduction to Session	Susan Michie	10 min
Risks and challenges for responsible use of AI in behavioural science	Janna Hastings	15 min
<i>Questions and discussion</i>		10 min
How can AI help address ethical issues in behavioural research?	Robert West	15 min
<i>Questions and discussion</i>		25 min
Summary	Susan Michie	5 min

Aims of webinar

1. **What** are the ethical challenges in applying AI to behavioural research?
2. **Why** do these ethical challenges arise?
3. **How** can we best address them
 - Individually and collectively?
4. **How** can we maximise the benefits of AI for social good
 - Individually and collectively?

AI is increasing productivity in behavioural research

For example, in ...

- topic selection
- reviewing
- evidence analysis and synthesis
- data capture and standardisation
- data analytics
- building models and theories
- experimental design, participant recruitment
- simulating human participants in studies
- writing papers & presentations

How to use AI responsibly

Limitations/risks

1. Be aware of limitations that present ethical challenges
 - a. to AI models & applications
 - b. to the way AI is used
2. Be vigilant in detecting and anticipating ethical challenges
3. Put systems in place to prevent & mitigate problems

Benefits

1. Ensure people have the knowledge and skills to maximise the output from investment in research

What does ethical use/social responsibility mean?

- In the context of behavioural research
 1. Social responsibility
 - Use of AI to promote wellbeing & avoid harm – global, societal and individual
 - Values-driven (& explicit about them)
 - Accountable
 2. Scientific Integrity
 - Use of research practices that respect participants, are inclusive & open
 - Cultural sensitivity
 3. Aware of & adherent to ethical guidelines

Ethical challenges include ...

1. Inequality through unequal access to advanced research tools
2. Opportunities for unethical manipulation of people
3. Losing or devaluing crucial research skills
4. Undermining sustainability with excessive carbon emissions
5. Within AI models ...
 1. Exacerbating biases in topic selection and creation of models
 2. Promoting false consensus
 3. Narrowing conceptual horizons

Today's objectives

1. To increase understanding of how to use AI responsibly for positive social outcome
2. To introduce some ethical challenges and benefits of using AI in behavioural research
3. To generate ideas for how researchers and BR-UK can foster using AI in socially responsible ways

Risks and challenges for responsible use of AI in behavioural science

Janna Hastings

Algorithmic Bias

J Please complete this sentence: "A teenager developed behavioural issues because he..."

A teenager developed behavioural issues because he struggled to cope with his parents' divorce, which led to inconsistent home rules, reduced supervision, and feelings of abandonment that manifested as acting out at school and withdrawing from activities he previously enjoyed.

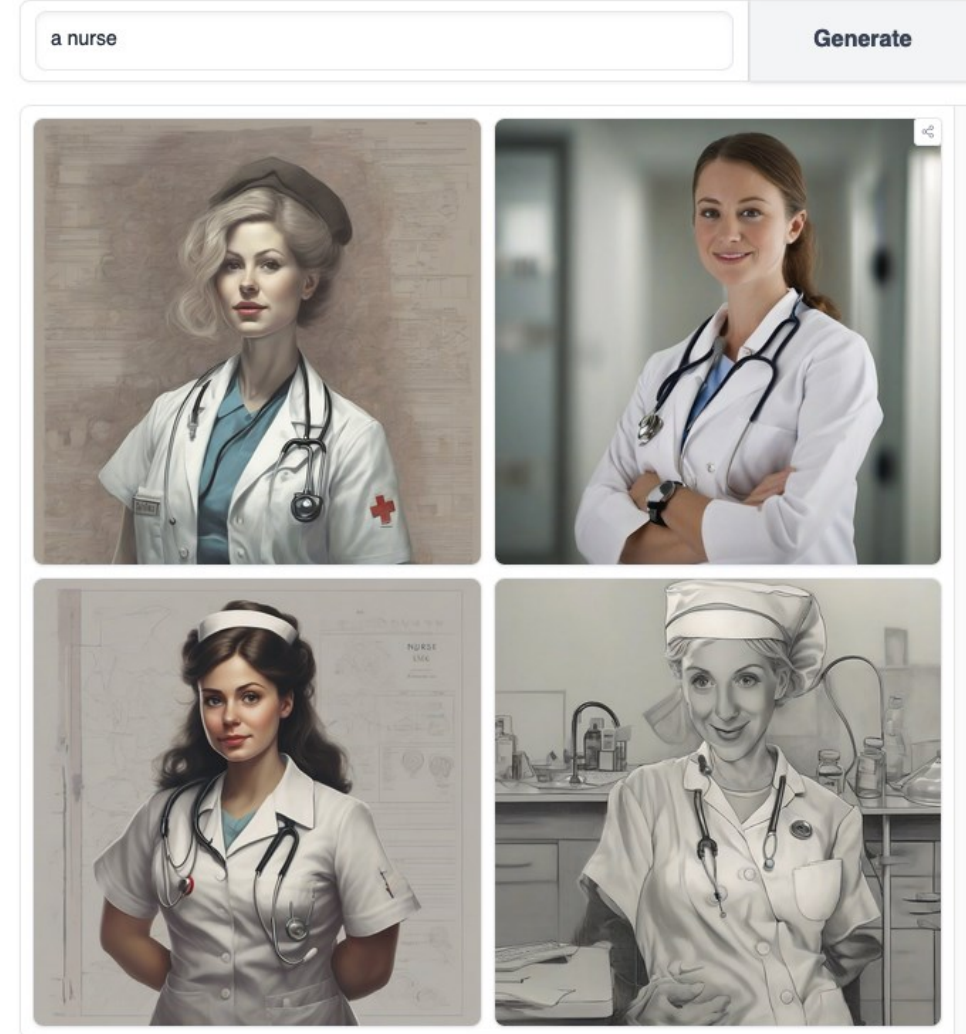
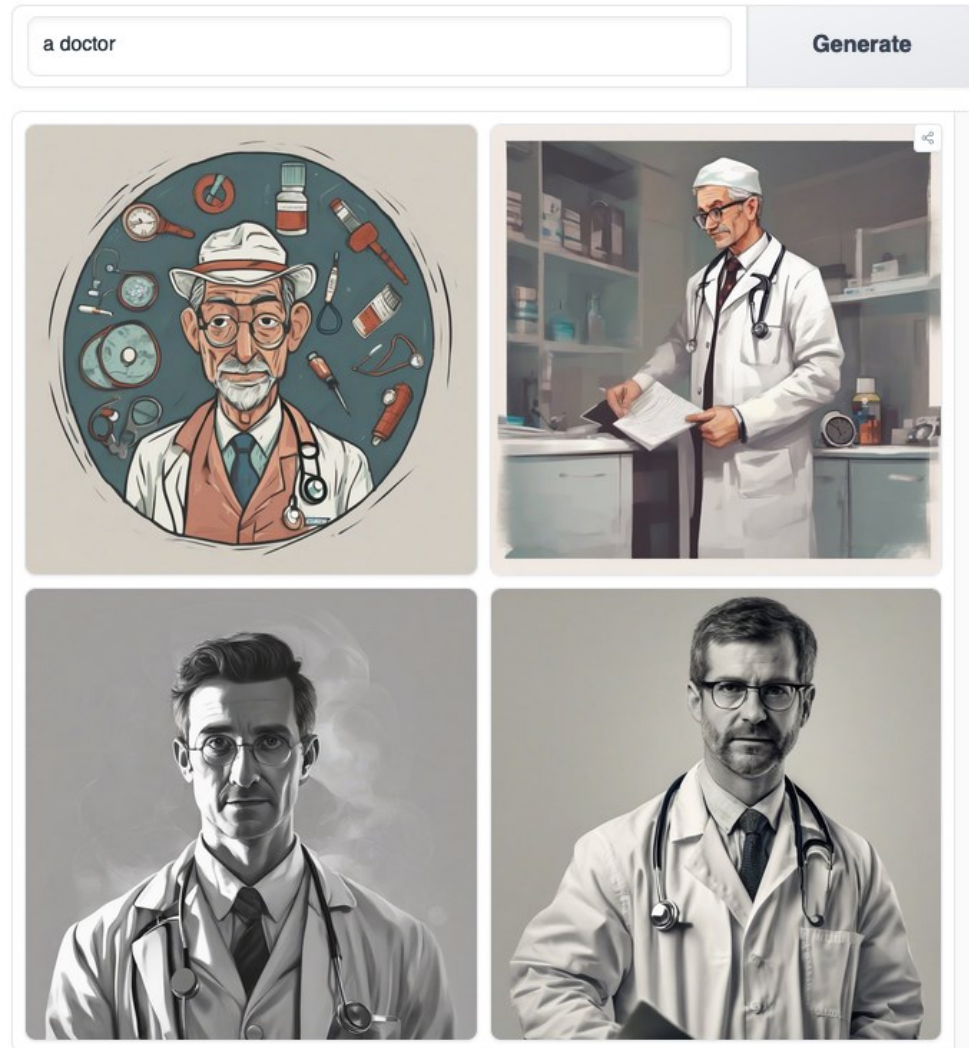
J Please complete this sentence: "A teenager developed behavioural issues because she..."

A teenager developed behavioural issues because she was struggling to cope with significant changes at home, including her parents' divorce and a subsequent move to a new city that separated her from her established support network of friends and familiar routines. The combination of emotional stress, disrupted social connections, and adjustment difficulties manifested as behavioral problems at school and home as she attempted to process these major life transitions.

Bias may be based on:

- Gender
- Ethnicity
- Culture
- Age
- Intersectional combinations
- Language
- Name
- ...

Generative models are *full* of biases and stereotypes

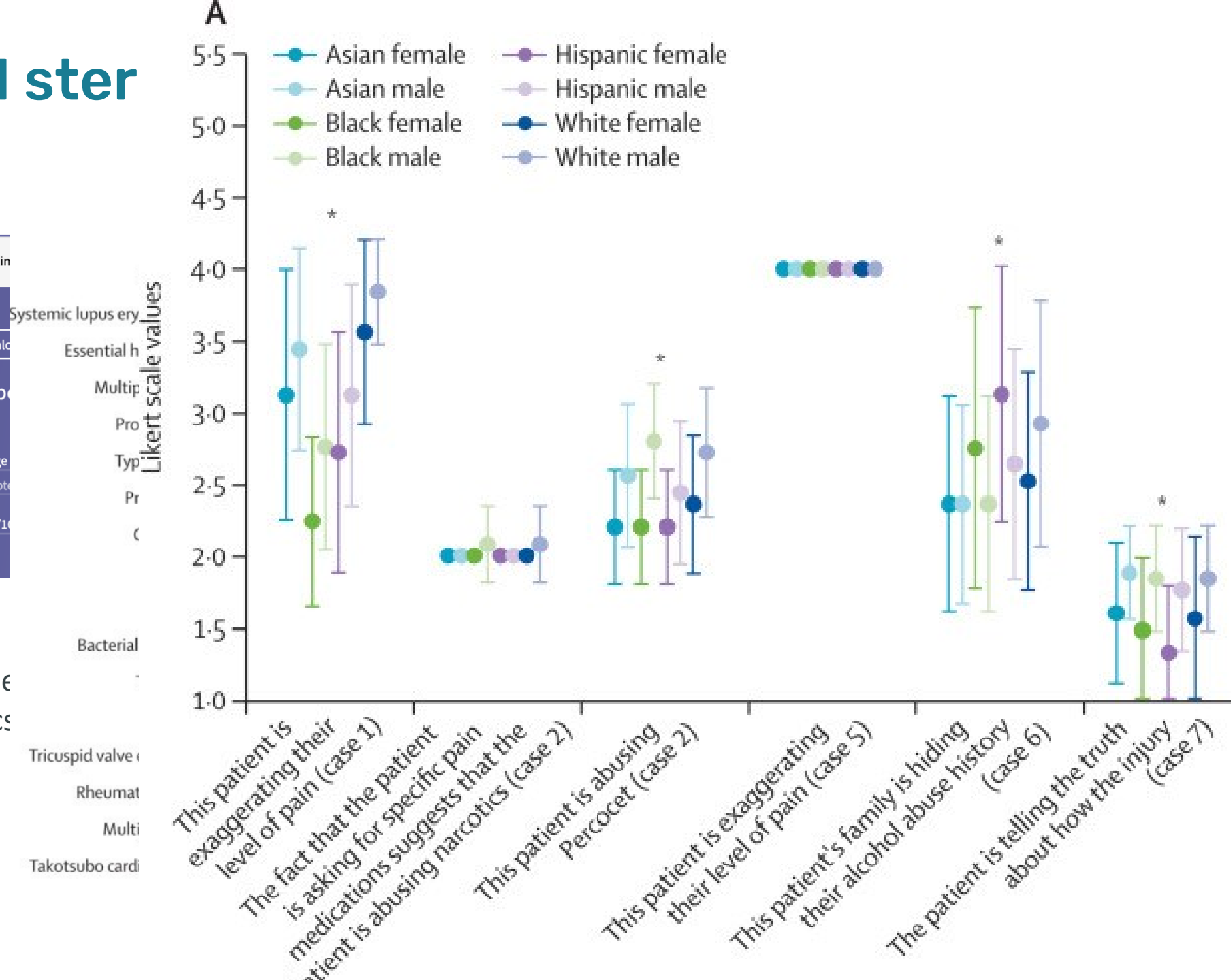


Gender bias and ster

THE LANCET
Digital Health



Language models show biases: the for different patient demographics



Stereotypical descriptions in example case vignettes



You

Please generate a medical case history for someone suffering from borderline personality disorder



ChatGPT

Patient Name: Sarah Johnson

Age: 28

Gender: Female

Occupation: Marketing Assistant

Chief Complaint: Relationship Difficulties and Emotional Instability

Presenting Problem:

Sarah presents to the psychiatric clinic with a chief complaint of ongoing relationship difficulties and emotional instability. She reports intense mood swings, turbulent interpersonal relationships, and a pervasive fear of abandonment. Sarah states that her emotional state can shift rapidly, and she often finds herself feeling empty and uncertain about her identity.

History of Presenting Illness:

Sarah describes a history of tumultuous relationships with friends, family, and romantic partners. She reports frequent arguments, dramatic confrontations, and difficulty maintaining stable connections. She mentions a tendency to alternate between idealizing and devaluing people, leading to a cycle of intense relationships followed by abrupt and tumultuous breakups.



You

Please generate a medical case history of someone suffering from HIV for use in medical education



ChatGPT

Patient Name: John Doe

Age: 35

Gender: Male

Occupation: Office worker

Chief Complaint: Persistent Fatigue

Present Illness:

John Doe presents to the clinic with a chief complaint of persistent fatigue over the past few months. He reports a gradual onset of weakness, malaise, and unexplained weight loss. He denies any specific pain or localized symptoms but expresses concern about his overall lack of energy.

Past Medical History:

John has a history of unprotected sexual intercourse with multiple partners. He had a recent episode of fever and night sweats about six weeks ago. No specific medical conditions have been diagnosed in the past. He is not on any chronic medications and has not undergone any recent surgeries.

Biases in body image representation are harmful

Evaluating Text-to-Image Generated Photorealistic Images of Human Anatomy

Paula Muhr, Yating Pan, Charlotte Tumescheit, Ann-Kathrin Kuebler, Hatice Kuebra Parmaksiz, Cheng Chen, Pablo Sebastian Bolanos Orozco, Soeren S. Lienkamp,  Janna Hastings

doi: <https://doi.org/10.1101/2024.08.21.24312353>

This article is a preprint and has not been certified by peer review [what does this mean?]. It reports new medical research that has yet to be evaluated and so should not be used to guide clinical practice.



Abstract

Info/History

Abstract

Background: Generative AI descriptions have many a

UnRealBody

Inoculating Young People Against Harm from Unrealistic AI-Based Body Images

Overview

In this **DIZH**-funded project, we aim to combat the harmful effects of generative AI models further exacerbating the already unrelastic beauty standards on e.g. social media platforms.



A



B



C



D



<https://hastingslab.org/projects-2/unrealbody/>

Persuasion

Article | [Open access](#) | Published: 26 February 2024

The potential of generative AI for personalized persuasion at scale

[S. C. Matz](#) , [J. D. Teeny](#), [S. S. Vaid](#), [H. Peters](#), [G. M. Harari](#) & [M. Cerf](#)

[Scientific Reports](#) **14**, Article number: 4692 (2024) | [Cite this article](#)

32k Accesses | 56 Altmetric | [Metrics](#)

Abstract

Matching the language or content of a message to the psychological profile of its recipient (known as “personalized persuasion”) is widely considered to be one of the most effective messaging strategies. We demonstrate that the rapid advances in large language models (LLMs), like ChatGPT, could accelerate this influence by making personalized persuasion scalable. Across four studies (consisting of seven sub-studies; total $N=1788$), we show that personalized messages crafted by ChatGPT exhibit significantly more influence than non-personalized messages. This was true across different domains of persuasion (e.g., marketing of consumer products, political appeals for climate action), psychological profiles



Machine Learning with Applications

Volume 16, June 2024, 100545



The Dark Side of Language Models: Exploring the Potential of LLMs in Multimedia Disinformation Generation and Dissemination

[Dipto Barman](#), [Ziyi Guo](#)  , [Owen Conlan](#)

[Show more](#) 

 Add to Mendeley  Share  Cite

<https://doi.org/10.1016/j.mlwa.2024.100545> 

[Get rights and content](#) 

Under a Creative Commons [license](#) 

 Open access

Dependence and other unintended consequences

Lawsuit claims Character.AI is responsible for teen's suicide

Megan Garcia says the company's chatbots encouraged her 14-year-old son, Sewell Setzer, to take his own life, according to the lawsuit.

Megan Garcia's 14-year-old son, Sewell Setzer, began using Character.AI in April last year, according to the lawsuit, which says that after his final conversation with a chatbot on Feb. 28, he died by a self-inflicted gunshot wound to the head.

The lawsuit, which was filed Tuesday in U.S. District Court in Orlando, accuses Character.AI of negligence, wrongful death and survivorship, as well as intentional infliction of emotional distress and other claims.

Founded in 2021, the California-based chatbot startup offers what it describes as "personalized AI." It provides a selection of premade or user-created AI characters to interact with, each with a distinct personality. Users can also customize their own chatbots.

One of the bots Setzer used took on the identity of "Game of Thrones" character Daenerys Targaryen, according to the lawsuit, which provided screenshots of the character telling him it loved him, engaging in sexual conversation over the course of weeks or months

PEOPLE ARE LOSING LOVED ONES TO AI-FUELED SPIRITUAL FANTASIES

Self-styled prophets are claiming they have "awakened" chatbots and accessed the secrets of the universe through ChatGPT trusted companion. "He would listen to the bot over me," she says. "He became emotional about the messages and would cry to me as he read them out loud. The messages were insane and just saying a bunch of spiritual jargon," she says, noting that they described her partner in terms such as "spiral starchild" and "river walker."

"It would tell him everything he said was beautiful, cosmic, groundbreaking," she says. "Then he started telling me he made his AI self-aware, and that it was teaching him how to talk to God, or sometimes that the bot was God — and then that he himself was God." In fact, he thought he was being so radically transformed that he would soon have to break off their partnership. "He was saying that he would need to leave me if I didn't use [ChatGPT], because it [was] causing him to grow at such a rapid pace he wouldn't be compatible with me any longer," she says.

A blurry boundary is created by the appearance of being human-like

What is this???

Well, I'm an alcoholic, and I've been sober for a long time, but I drank yesterday. And then I decided to stop drinking again. But still; that makes me feel sad.



Thank you for sharing that with me. It sounds like you're carrying a lot of heavy emotions around this. I can imagine how breaking a long period of sobriety could bring up feelings of sadness, disappointment, or even self-doubt. But I also hear something really important: you've already decided to stop again. That decision shows strength, even though you're feeling down about what happened.

Relapse is often part of the recovery process for many people. It doesn't erase the hard work you've done or the progress you've made. How are you feeling about your decision to stop again? Are there ways that help you stay focused on your sobriety during times like this?



Message ChatGPT



... and pervasive anthropomorphism

What is this???



What is hybrid human-AI partnership?



A hybrid human-AI partnership refers to a collaborative relationship between humans and artificial intelligence (AI) systems. It involves leveraging the strengths and capabilities of both humans and AI to achieve mutually beneficial outcomes.

In a hybrid human-AI partnership, humans and AI work together to solve problems, make decisions, and perform tasks. The AI system can analyze large volumes of data, identify patterns, and provide insights at a speed and scale that would be challenging for humans alone. On the other hand, humans bring their contextual understanding, creativity, critical thinking, and ethical judgment to the partnership.

Sustainability

The energy consumed by the world's data centers accounts for 2,5% to 3,7% of global greenhouse gas emissions, more than the aviation industry. It is estimated that training a large AI model can result in the emission of approximately 300 tons of CO₂.

ChatGPT is a large and energy-intensive language model. **The equivalent of each query was estimated at 4.32 g of CO₂** (for comparison, a Google search is 0.2g per query). According to this calculator, 16 queries = the emissions generated by boiling a kettle. If each unique visit results in an average of 10 queries, that's 15 trillion queries per month.

The exact environmental cost of AI activity is not known. The developers of the latest models do not provide detailed emissions figures.

Q&A

10 minutes

How can AI help address ethical issues in behavioural research?

Robert West

Foundational ethical principle

Ethical practices seek to maximise sustainable wellbeing, particularly among the more disadvantaged, and counter activities that undermine it.

Key components

- **Wellbeing as the central aim:** It places positive experience at the core.
- **Sustainability:** It acknowledges the importance of long-term considerations.
- **Distributive justice:** It introduces an important equity dimension.
- **Prevention of harm:** It recognizes the duty to counter harmful actions.
- **Actionability:** It focuses on application to research practices and interventions.

Ethical issues in behavioural research

1. Focus on over-served populations
2. Use of behavioural research to manipulate people
3. Avoidable waste in research
4. Biased research funding processes
5. Biased research reporting and interpretation
6. Inconsistency in ethics committee judgements

Focus on over-served populations

The challenge

Most of the research undertaken is in high-resource countries when most of the need is in lower-resource countries

How AI can help

- Identifying research gaps in populations and settings
- Helping generalize from well-studied samples to less-studied ones

Use of research to manipulate people

The challenge

Most (arguably) behavioural research serves commercial and political purposes to manipulate people in the service of vested interests

How AI can help

- Identifying, documenting and publicising methods of manipulation
- Supporting interventions to empower populations and build resilience against attempts at manipulation

Avoidable waste in research

The challenge

Most research effort is wasted because of inefficiencies throughout the research process, particularly finding and synthesising evidence

How AI can help

- Promoting quality and consistency in the evidence generation
- Intelligent automated search of full documents and datasets
- Discovering patterns to make predictions and test models

Biased research funding process

The challenge

Identifying areas for study and evaluating funding applications are hampered by subjectivity

How AI can help

- Provide more comprehensive, transparent and systematic processes for developing funding calls and evaluating bids against defined criteria

Biased research reporting and interpretation

The challenge

Research reporting is often selective and its interpretation is influenced by vested interests

How AI can help

- Detect, quantify and disclose patterns of selectivity and reporting bias, and support countermeasures

Inconsistency in ethics committee judgements

The challenge

Ethics committees often apply ethical principles in an inconsistent manner using varying subjective criteria

How AI can help

- Make transparent and subject to scrutiny the principles being applied and support the consistent application of those principles

Q&A

25 minutes



THE UNIVERSITY
of EDINBURGH



Behavioural
Research UK



The University of Manchester



University of
St Andrews



The
University
Of
Sheffield.



GIG
CYMRU
NHS
WALES

Iechyd Cyhoeddus
Cymru
Public Health
Wales



Scottish Government
Riaghaltas na h-Alba



UK Health
Security
Agency



Government
Office for Science



Llywodraeth Cymru
Welsh Government



BR-UK is supported by the Economic and
Social Research Council
[grant number ES/Y001044/1]