Consumer’s Cognitive and Affective Perceptions of Artificial Intelligence (AI) in Social Media: Topic Modelling Approach

Abstract: This study aimed to examine consumer perceptions of artificial intelligence (AI) by analyzing YouTube comments using topic modeling. A natural language processing model called BERTopic was employed to identify latent topics in a total of 93,729 English-language comments. These comments were posted on 20 popular AI-related videos, each of which had over 1 million views, between January and December 2021. The analysis of the comment data yielded five major themes, each with its corresponding percentage representation. The first theme, accounting for 27.2% of the comments, revolved around the fear of AI surpassing human control. The second theme, comprising 13.4% of the comments, focused on anthropomorphism and the uncanny valley effect, indicating mixed emotions towards human-like robots. The third theme, representing 31.7% of the comments, highlighted AI bias and hallucinations, shedding light on the impact of AI inaccuracies on trust. The fourth theme, accounting for 17.2% of the comments, addressed ethical and moral concerns, particularly concerning AI development and oversight. Lastly, the fifth theme, encompassing 10.5% of the comments, explored AI trustworthiness, emphasizing the reliability of AI in decision-making processes. Based on the findings, this study challenges previous research that primarily reported positive AI experiences. It quantitatively identifies complex negative emotions towards AI technologies across diverse user demographics. Furthermore, it expands our understanding of how anthropomorphism and AI "hallucinations" influence trust, which are critical issues as AI becomes increasingly integrated into society. The application of BERTopic enabled nuanced modeling of online discourse, bridging gaps in our knowledge regarding consumer perceptions of emerging technologies on a large scale. In terms of novelty, this study is the first to utilize BERTopic modeling to uncover consumer experiences with AI by analyzing social media comments on a large scale. The insights obtained from this approach offer compelling new perspectives compared to previous research in this field.

Keywords: Artificial intelligence, Consumer Experience, Machine Learning, Hallucination, Fear.

I. INTRODUCTION

As artificial intelligence (AI) enables the creation of personalized and tailored user experiences by learning from user behaviour, preferences, and context, AI systems can provide relevant content, suggestions, and interactions that enhance consumer cognitive and affective experiences [1]. These experiences can be mediated by various forms of AI embodiment, such as physical (e.g., robots) or non-physical (e.g., chatbots, algorithms). As AI agents become more intelligent and capable of establishing rapport with users, there is a need for research in experience marketing to explore the factors that influence consumer experiences in non-traditional settings, such as those involving virtual or physical service encounters with AI agents.

Reference [2] argues that “experiences offer sensory, emotional, cognitive, behavioural, and relational values that substitute functional values”, highlighting how the conventional view of consumer decision journey overlooks the holistic experience of consumers. Researchers urged the need to empirically examine various dimensions of experiences on consumer experiences in various domains specifically with AI agents [3]. Previous research on AI-consumer service experience has mostly focused on utilitarian aspects, such as perceived usefulness and ease of use, while limited effort was applied to the emotional experiential perspective of user experiences [4]. Affection toward AI imparks positive and negative emotional mental states which affects consumer perception and usage continuation [5]. Studies reported that the majority of consumer experience with physical embodiment of AI are positive while negative instances are due to service malfunction [6]. However, consumers demonstrated cases of negative emotions as well according to the uncanny valley theory [7]. Studies in AI experience in service and emotions are not generalized and fragmented and required further studies to enrich our understanding in this scope. However, current understanding of what experiences consumer’s encounter is limited and developed scales are lacking in addressing the complexities of AI-consumer encounters [3,8].
This study aims to discover topics in AI experiences specifically the affective and cognitive dimensions. Trust in AI is multifaceted, influenced not only by the performance and consistency of AI systems but also by users’ expectations, prior beliefs, and perceived risks. As AI technologies become more sophisticated and integrated into our daily lives, it is imperative to understand how these factors collectively shape trust. The role of hallucination in AI systems is a critical aspect that has not been adequately addressed in trust research. Hallucinations, or the generation of false or misleading information by AI, pose a significant challenge to the establishment and maintenance of trust. These occurrences can undermine the perceived reliability of AI and lead to scepticism among users. Recognizing the importance of this issue, this study will delve into the effects of AI hallucinations on consumer trust, exploring how these incidents influence perceptions of AI and impact the trust-building process [5]. Additionally, the affective dimension of AI experiences is also a critical aspect of this study. Affection toward AI imparts positive and negative emotional mental states, which affect consumer perception and usage continuation. This study aims to unearth topics which are representative of the overall sentiment of emotions of users to address the debate of consumer experience with AI.

Previous research in a similar setting used Machine Learning (ML) techniques such as sentiment analysis and topic modelling methods such as BERTopic and Vader algorithms to examine consumers' experiences. [6] applied BERTopic to understand consumers' interactions with robots in the hotel sector using data from TripAdvisor and Cript, providing a useful insight but with a limited scope and a small sample size of n=9707 that did not exploit the full potential of the technique. Moreover, their findings suggested that the customer experience with service robots was mainly positive, which was also reported by [9]. These results were not consistent with the current study's findings and supports the findings of [10] which found the majority of experiences in YouTube comments through qualitative analysis are negative. Although their study incorporated small sample size n=1163.

This study contributes to the existing literature by utilizing BERTopic, a state-of-the-art machine learning topic modelling algorithm, to uncover topics related to consumer experiences with AI in social media. By conducting a semantic analysis of comments on YouTube, this study aims to enhance the generalizability of findings in similar research contexts and discover topics that has not incorporated in existing AI literature.

Understanding the impact of AI on consumer experiences becomes increasingly important as AI continues to evolve, and their intelligence has advanced rendering previous empirical findings questionable [5]. AI technologies used to provide technical repetitive tasks proficiently, but as they have become widespread, companies continued to humanize machine to foster for acceptance and increase adoption what is termed the feeling economy. An economy where AI systems can understand, comprehend, and respond to emotionally relevant environments and act as a social entity [1]. These affective elements portrayed by AI-based machines complicate consumer experiences and solutions provided by corporations creating a complex endeavour [4,11]. Previous research highly focused and fixation to understand consumer experiences using utilitarian approaches in AI agents [12]. Emotions in AI has gained special attention by researchers as AI agents became more human-like. The studies of anthropomorphism – the attribution of human-like characteristics to non-human entities, is a fundamental aspect of many AI technologies, especially those involving conversational agents, chatbots, or social robots [11]. Various experiences play a critical role in consumer perceptions of AI when technologies are anthropomorphized and gained higher intelligence abilities such as facial recognition and speech fluency [13]. When AI systems display human-like qualities, such as voice or facial expressions, users often respond with emotions as they would in human-to-human interactions. Positive emotional experiences, like joy or intrigue, can enhance user engagement and foster adoption [14]. However, negative emotions, such as fear or creepiness, can arise if the AI system appears too human-like, leading to discomfort and resistance [15].

Empirical research in various fields, such as marketing, sociology, and information systems, relies on social networks as a valuable source of data. The user-generated content (UGC) in social networks, such as comments, likes, and shares, offers rich and naturalistic data to understand diverse phenomena, from consumer behaviour to social movements. Social networks also allow consumers to express their opinions, experiences, and sentiments about products, services, or events. Researchers can analyse the UGC thematically to gain insights into consumer behaviour, trends, and preferences [16]. Additionally, social networks facilitate the study of social interactions and phenomena at a large scale. They provide data on the communication patterns, content, and emotions of users, enabling researchers to examine complex social processes and dynamics [17].
The enormous and increasing amount of data in social networks poses opportunities for researchers. Traditional research methods, such as manual coding and analysis, are often impractical or inefficient for analysing such big data. This is where machine learning (ML), a branch of AI, comes in. ML algorithms can automatically analyse and learn from data, making them highly suitable for large-scale data analysis. For example, ML models can be trained to automatically classify comments into different themes or sentiments, eliminating the need for manual coding [6]. Moreover, deep learning (DL), a subfield of ML, can model high-level abstractions in data, enabling more sophisticated analyses. For instance, DL models can comprehend the semantic meaning of text, making them effective for tasks such as emotion detection, topic modelling, and sentiment analysis.

Research, especially in AI embodied agents such as robots and chatbots, has been using conventional forms of empirical methodology. These include experiments, structural equation modelling, interviews, and mixed methods [18]. However, these academic approaches have some limitations. The narrow scope of qualitative methods reduces its generalizability as the number of participants in such methods is usually low due to cost factors. In other methods, respondents are measured through a predefined set of items that may not capture unplanned dimensions of the phenomenon under study [19]. Thus, researchers increasingly applied ML approaches to answer for these methodological shortcomings [20].

Online reviews are a valuable source of information for service managers and researchers, as they reflect customer emotional responses to service encounters and influence their purchase decisions. However, analysing online reviews qualitatively can result in valuable insights but it is still constrained by being time-consuming, especially when dealing with large and diverse datasets. Machine learning methods can help harvest online reviews more efficiently and effectively, by using algorithms to recognize and differentiate emotions, topics, sentiments, and other aspects of customer feedback, also it is an advent approach which is yet being applied in literature, having a continuous application as [6] argued and proved its limitations in detecting true emotions from a large corpus of data.

With the aid of machine learning through deep learning mechanisms, it is possible to analyse a massive amount of data to reflect on research quests using styles such as topic modelling among the wide algorithms available developed by major corporations such as Google, Facebook and OpenAI. For example, [6] used a hybrid machine-human intelligence approach to understand customer emotions from a large corpus of data using deep learning, based on online reviews from two major social media platforms. Similarly, [21] used an improved k-nearest neighbour model and a latent Dirichlet allocation model to harvest online reviews to identify the competitor set in the hotel industry. They analysed customer reviews to evaluate the strengths and weaknesses of different tourism destinations and based on various hotel attributes. These studies demonstrate majorly positive experiences with robots in their sampled platforms and data used.

Topic modelling is a technique to discover latent topics in collections of documents. It can be useful for organizing, summarizing, and exploring large collections of documents at scale. However, different topic modelling algorithms make different assumptions about the generation of topics and the relationship between topics and words. For example, Latent Dirichlet Allocation (LDA) assumes that documents are mixtures of topics and that words are generated from topics, while Correlated Topic Model (CTM) considers topic correlations. The choice of algorithm depends on the specific research objectives and characteristics of the dataset. In this study, BERTopic, a state-of-the-art neural topic modeling technique, will be utilized to uncover the dominant topics regarding consumer experiences with AI from online reviews, without making strong assumptions about topic structure.

II. METHODOLOGY

The principal stages of our research involved: (1) collecting, refining, and analysing textual data from YouTube comments in topics related to AI in general, focusing on sentiments related to emotions and trust, spanning, (2) applying and fine-tuning BERTopic models on data accumulated before and during the pandemic, and (3) manual assessment and polishing of BERTopic outcomes, followed by a thematic categorization of the discovered topics.

In the data acquisition phase, we gathered comments from high-engagement videos using YouTube's API, focusing on diverse content related to AI, including CEO interviews, comedy shows, influencers, technical education, and news reports. The selected videos had over one million views, ensuring a broad representation and reducing
viewership bias. A total of 93,729 comments were processed by removing numerical figures, non-English characters, and short comments (less than 10 words). URLs and links were also stripped to eliminate potential noise, resulting in a refined dataset.

To analyse the data, we used the BERTopic model, which is a topic modelling technique based on Bidirectional Encoder Representations from Transformers (BERT). BERT is a deep learning model that can capture the meaning and context of words in natural language. The BERTopic model converts the text data into numerical representations called word embeddings, which are vectors that capture the semantic similarity of words. Then, it groups the word embeddings into clusters, which represent the topics in the data. To reduce the number of dimensions in the word embeddings, we used Uniform Manifold Approximation and Projection (UMAP), which is a dimensionality reduction technique that preserves the local and global structure of the data. To remove noise and outliers from the clusters, we used Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN), which is a clustering algorithm that can identify clusters of varying shapes and densities. This method ensured that the topics we extracted were clean and meaningful, and that we did not have to specify the number of topics beforehand [22].

The final step involved manual assessment and thematic categorization, aligning with recommendations from literature [6]. This iterative process allowed for a comprehensive understanding of the model's outcomes, bridging the gap between machine-driven analysis and human interpretation. The methodology, combining advanced machine learning techniques with human insight, contributes to a nuanced exploration of emotions and trust instances in YouTube engagements related to AI.

III. RESULTS AND DISCUSSION

BERTopic clustered around mostly negative instances with topics ranging from emotional intelligence, biasness, reliability, threat, competence and more. Rather than extracting each topic in isolation, we found similarities in their underlying clusters and merged them under one topic due high relevance and them thematically labelled the topics as suggested by [6].

A. Topic 1 — Negative emotions toward Artificial intelligence: Robots and general applications

The first theme is related to the perceived potentiality of AI in future capabilities in overcoming human intelligence and gaining power [Table 1]. The belief of uncontrolled destruction can be attributed to the uncanny valley theory which explains the perception of discomfort in humanized objects such as robots as they gain more humanized resemblance. The analysis reveals the distribution of comments across various topics related to consumer perceptions of AI. In [Table 1], the total count of comments is 3,616. the first topic, 'Fear of Artificial intelligence technologies', comprises 983 comments, representing 27.2% of the total count of comments. Similarly, the second topic, 'Anthropomorphism and fear of robots', comprises 484 comments, representing 13.4% of the total count of comments.

The thematic analysis of user comments [Table 2] related to AI reveals a collocation of awe and apprehension, underscoring a complex relationship with emerging technologies. A prevalent theme is the fear of AI, with one user's sinister words, “In the end, it will be a very negative problem,” reflecting a sentiment echoed across various comments. Concerns about AI's morality and its potential to abuse power are succinctly expressed as users ponder a future where discerning AI from human interaction becomes ambiguous. The comment, “For example, what if you couldn't tell it was this AI,” illustrates this anxiety. Furthermore, the potential for AI to develop consciousness invites both fascination and fear, as suggested by reflections on robots integrating into daily life, highlighted by the statement, “The real scary part is going to be when robots start walking with humans.

Anthropomorphism amplifies these fears, with users articulating unease towards robots like Sophia, whose advanced capabilities spark debates about the wisdom of creating such intelligent entities. Comments range from admiration of Sophia's linguistic prowess to alarm at her mentions of human destruction, encapsulated by a user's concern, “Creating robots with these advanced consciences is a bad idea.” The fear is further magnified by descriptions of Sophia's presence as “creepy” and “emotionless,” indicating a discomfort with non-human entities possessing human-like characteristics. This uncertainty towards AI underscores a broader societal dilemma: the balance between harnessing AI's potential benefits while mitigating the existential threats it may pose.
Research has shown that giving human qualities to an object causes contrasting emotional reactions by users. Many of the topics discovered are grouped around negative emotions especially toward the robots. This analysis presents a compelling contrast to the findings of (6) and supports (9), which reported predominantly positive emotions in customer encounters with service robots. (6) found joy to be the most frequent emotion, this thematic analysis reveals a significant presence of negative emotions, particularly fear and apprehension towards AI and humanoid robots. This discrepancy could be due to different contexts: (6) focused on service robots in hospitality settings, where interactions are typically brief and functional, whereas our analysis considers broader societal experiences and the potential for AI to surpass human intelligence. Although they have identified in their study the mix of emotions consumer portray with a specific robot (Sophia), this research identifies the emotional perception is highly negative and focused about fear and apprehension (6).

**B. Topic 2—AI Bias, hallucination, trustworthiness, and GPTs**

The analysis shows that the total count of comments is 980. The first topic, Data training and information quality, comprises 311 comments, representing 31.7% of the total count of comments. Similarly, the second topic, 'Ethical and moral concerns', comprises 169 comments, representing 17.2% of the total count of comments.

<table>
<thead>
<tr>
<th>Count of comments</th>
<th>BERTopic generated Keyword</th>
<th>Thematic labelling</th>
<th>Top keywords in the topic</th>
<th>Sample comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>983</td>
<td>3_kill_planet_humanity_humans</td>
<td>Fear of Artificial intelligence technologies</td>
<td>['kill', 'planet', 'humanity', 'humans', 'ai']</td>
<td>“I’ve been on high alert with A.I. technology since I seen I Robot, then I started watching other movies like that as well. I agree some A.I. might help a small fraction of humans. In the end, it will be a very negative problem in the very near future.”</td>
</tr>
<tr>
<td>590</td>
<td>10_fear_fear ai_afraid_afraid ai</td>
<td></td>
<td>['fear', 'fear ai', 'afraid', 'afraid ai', 'scared']</td>
<td>“It is truly scary and represents the fact that we are going to have to be more moral to start wielding these things as it would be easy to abuse.”</td>
</tr>
<tr>
<td>557</td>
<td>12_threat_war_quot_humany</td>
<td>['threat', 'war', 'quot_humany']</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>--------------------------</td>
<td>---------------------------------</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

“The future of AI is definitely very interesting yet very scary. It may someday have a conscious like a human. The real scary part is going to be when robots start walking with humans in daily life not just on a computer.”

<table>
<thead>
<tr>
<th>484</th>
<th>18_sophia_sofia_destroy_humans_destroy</th>
<th>['sophia', 'sofia', 'destroy_humans', 'destroy']</th>
</tr>
</thead>
</table>

“I am so surprised how Sophia answers every questions correctly with no grammatical mistakes”

<table>
<thead>
<tr>
<th>294</th>
<th>36_robots_world_destroy_robot Anthropomorphism and fear of robots</th>
<th>['robots', 'world_destroy', 'robot']</th>
</tr>
</thead>
</table>

“Creating robots with these advanced consciences is a bad idea since they now be the most or second most intelligent beings on the planet. Therefore, we naturally be enemies to AI […]”

<table>
<thead>
<tr>
<th>275</th>
<th>40_scared_scary_terrifying_scares</th>
<th>['scared', 'scary', 'terrifying', 'scares', 'terrifying_time']</th>
</tr>
</thead>
</table>

“She’s creepy…she looks like people who know what to talk without having anything to do with it and are emotionless”
The topic generated by the model specifically clustered words like “hallucinations”, “lie”, “trust”, and “fake”, this topic delves into the realm of artificial intelligence and its occasional tendency to produce unintended or inaccurate outputs, often termed as “hallucinations” (23). The dialog in this topic revolves around the inherent imperfections of AI systems, the potential for these systems to propagate misinformation, and the broader societal implications of relying on such technologies and its impact on trust (5). A notable concern raised within these discussions is the erosion of trust in AI due to its ability to "hallucinate" or provide false information. Users’ express apprehension over the increasing reliance on AI, fearing that unchecked adoption could lead to widespread dissemination of incorrect data, furthering misinformation and potentially harming societal trust in digital tools. The dangers of blind trust in AI outputs, the ethical responsibilities of developers, and the need for rigorous validation of AI-generated content are some of the recurring themes. Following is an example of a post highlighting the concerns shared by a user about the nature of AI hallucinations.

Table 2: BERTopic output count thematic analysis and comments samples

<table>
<thead>
<tr>
<th>Count of comments</th>
<th>BERTopic generated Keyword</th>
<th>Thematic labelling</th>
<th>Top keywords in the topic</th>
<th>Sample comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>311</td>
<td>32_garbage_program_ai_ai</td>
<td>Data training and information quality</td>
<td>['garbage', 'program', 'ai ai', 'ai', 'garbage garbage']</td>
<td>“its not that hard to explain: garbage in, garbage out. Heck just yesterday i saw someone telling the AI over and over again, until the AI accepted it and from that point on would also answer when asked for”</td>
</tr>
<tr>
<td>169</td>
<td>72_morality_moral_morals_ethical</td>
<td>Ethical and moral concerns</td>
<td>['morality', 'moral', 'morals', 'ethical', 'ethics']</td>
<td>“Needs an ethical algorithm to be built and managed by independent organizations voted by people (variety of ideologies). Parties are mostly captured. Things like “AI” cannot create or advise anyone (including fbi, president,…) on how to harm humans in any way. Any other ideas?”</td>
</tr>
<tr>
<td>149</td>
<td>85_black_box_black_box_problem</td>
<td>Black box and transparency in artificial intelligence</td>
<td>['black box', 'black', 'box', 'problem', 'boxes']</td>
<td>“This is a really complex algorithm that was fed a mind-boggling amount of data, which was then distilled into a probabilistic representation of an algorithm encoded into billions of values with no semantic attached to them whatsoever. That is what enables these AIs to perform on the level they do, which often is above our own, […]”</td>
</tr>
</tbody>
</table>
In the evolving digital landscape, the integration of AI, especially in e-commerce platforms and chatbots, plays a pivotal role in reshaping consumer interactions [13]. However, as these systems become deeply ingrained in the consumer decision-making process, the accuracy and reliability of their outputs become critical [1]. AI “hallucinations” present a unique challenge in this context. Such inadvertent or misleading outputs from chatbots or recommendation systems have the potential to compromise the user experience, leading to misguided or even detrimental purchasing decisions. Trust, a cornerstone of consumer engagement in the digital era, can be significantly undermined when these AI-driven platforms produce unreliable or false information [24]. As consumers become more reliant on AI for real-time, personalized recommendations, ensuring the trustworthiness of these systems is imperative [25,26]. Addressing AI hallucinations isn't merely a technical endeavour—it embodies the essence of preserving consumer trust and confidence in an increasingly AI-dependent marketplace.

### IV. CONCLUSION

This study explored topics related to consumer perceptions and experiences of AI through an analysis of YouTube comments using the BERTopic model. Several key findings and theoretical contributions emerge from this research. The analysis uncovered five dominant topics: 1) fear of AI technologies, 2) anthropomorphism and fear of robots, 3) data training and information quality, 4) ethical and moral concerns, and 5) AI bias, hallucination, trustworthiness, and GPTs. Quantitatively, the first topic of fear of AI comprised 27.2% of comments, while anthropomorphism and robots fears made up 13.4%. This challenges prior work finding predominantly positive emotions, demonstrating the complexity of consumer responses to various AI embodiments.

A major theoretical contribution of this work is extending understanding of how anthropomorphism influences perceptions. The findings suggest human-like qualities in robots can amplify concerns about AI surpassing humanity or developing consciousness. This counters dominant narratives and signals fruitful avenues for future research focused on what AI attributes generate negative appeal. Additionally, the study enhances knowledge around “AI hallucinations” and their impact on consumer trust as AI integrates into decision-making. With 31.7% of comments focusing on data training and information quality concerns, the potential for AI to disseminate misinformation undetected threatens to undermine trust, an issue deserving more attention.

Limitations include reliance on YouTube comments which may not represent all views. English-only comments exclude other perspectives. Data collection platform bias is also a concern. Future work could employ longitudinal studies and varied data sources to capture emotional evolution and diversity of opinions over time with increased AI exposure.
ACKNOWLEDGEMENT

The authors would like to acknowledge the support brought by Yanbu Industrial College, Royal Commission for Yanbu Project, Saudi Arabia.

REFERENCES


