An Exploratory Study of Generative AI Service Quality using LDA Topic Modeling and Comparison with Existing Dimensions*

YaeEun Ahn** · Jungsuk Oh***

Content

Abstract

Artificial Intelligence (AI), especially in the domain of text-generative services, has witnessed a significant surge, with forecasts indicating the AI-as-a-Service (AIaaS) market reaching a valuation of $55.0 Billion by 2028. This research set out to explore the quality dimensions characterizing synthetic text media software, with a focus on four key players in the industry: ChatGPT, Writesonic, Jasper, and Anyword. Drawing from a comprehensive dataset of over 4,000 reviews sourced from a software evaluation platform, the study employed the Latent Dirichlet Allocation (LDA) topic modeling technique using the Gensim library. This process resulted the data into 11 distinct topics. Subsequent analysis involved comparing these topics against established AI service quality dimensions, specifically AICSQ and AISAQUAL. Notably, the reviews predominantly emphasized dimensions like availability and efficiency, while others, such as anthropomorphism, which have been underscored in prior literature, were absent. This observation is attributed to the inherent nature of the reviews of AI services examined, which lean more towards semantic understanding rather than direct user interaction. The study acknowledges inherent limitations, mainly potential biases stemming from the singular review source and the specific nature of the reviewer demographic. Possible future research includes gauging the real-world implications of these quality dimensions on user satisfaction and to discuss deeper into how individual dimensions might impact overall ratings.

Keywords : Quality Dimension, Generative AI, Service Quality, LDA Topic Modeling, Review Analysis

접수일 (2023년 12월 04일), 수정일 (2023년 12월 11일), 게재확정일(2023년 12월 13일)
* This study was supported by the Institute of Management Research at Seoul National University.
** First Author, Business School, Seoul National University, ayu7045097@snu.ac.kr
*** Corresponding Author, Business School, Seoul National University, joh@snu.ac.kr

1. Introduction

Artificial intelligence (AI) is defined as technology, such as machine learning, big data, natural language processing and understanding, that enables software agents to 'act intelligently' (Poole & Mackworth, 2010). AI has offered business and service providers the potential to boost revenue and reduce operational costs (Davenport et al., 2020). According to the research, providing AI-as-a-Service (AIaaS) market is evaluated as $9.3 Billion and it is expected to worth $55.0 Billion by 2028 (MarketandMarkets, 2023). The market value of chatbots and virtual assistants, the two common types of AI agents used by business to provide service to consumers, is expected to increase at the compound annual growth rate of 33% between 2020 and 2025 (AMR, 2020). Additionally, it is found that 35% of the businesses worldwide used AI in 2022 and 15% of all customer service interactions globally were expected to have been fully powered by AI in 2021 (IBM, 2022; Gartner, 2019). Adding on, 54% of the organizations have reported cost savings and efficiencies as a result of AI implementation (IBM, 2022).

Despite the usefulness of AIaaS, there also has been issues on trustfulness of the results and so on. About 78% of the businesses say it is important for them to be able to trust AI's analysis results and recommendations (Thomas, 2020). This concern has been further enhanced with introduction of generative AI. The generative AI is the latest AI technology that produce media with given input such as voice or likeness or prompt from the users. The media created by the generative AI is called the synthetic media and it is estimated to be accounted for 10% of all the data produced by 2025. Due to this reason, some generative AI software are called synthetic media software as well. Considering the fact that less than 1% of the data were synthetic media in 2021, it is a huge growth, and this area is expected to grow rapidly (Gartner, 2021).

A popular generative AI service would be ChatGPT, a conversational AI developed by OpenAI, that can chat with the users, answer follow-up questions, admits its mistakes, challenge incorrect premises and reject inappropriate requests (OpenAI, 2023). Despite its' usefulness, several concerns were raised. It is said when embracing generative AI into a corporate culture, several issues should be considered such as distribution of harmful content, copyright and legal exposure, data privacy violations, sensitive information disclosure, amplification of existing bias, workforce roles and morale, data provenance and lack of explainability and interpretability (Lawton, 2023).

There are several studies that have researched the quality dimensions of the AI service agents such as chatbots and virtual assistants. These studies have each came up with a service quality dimension of AISAQUAL and AICSQ. Although they share similar components, some dimensions differ as the focus of each study differs. Adding on, these studies are based on the conversational agent, which is considered to be previous stage of generative AI service. Therefore, applicability of existing quality dimension is questionable.

This study aims to extract the quality dimension from the online reviews using LDA topic modelling and compare it with existing AI service quality dimensions which only focused on the conversational AI agents. This paper would explore whether the existing quality dimension is applicable to evaluate the service of generative AI.

2. Theoretical Background

2.1 Evolution of AI Chatbots

The first generation of chatbot was ELIZA which began in 1966. ELIZA was created by a MIT professor, Joseph Weizenbaum, and it used pattern matching and substitution methodology to simulate conversation and was intended to mimic human conversation (Weizenbaum, 1966). Along with ELIZA, the first generation of chatbots, what we call as 'Basic Chatbots' used decision trees and simple keyword-recognition capabilities to generate scripted responses (Koury & Murphy, 2023).

The next generation of AI chatbots, what so called, 'Conversational Agents', what we are more familiar with. It includes chatbot systems such as IBM Watson and virtual assistants like Siri and Alexa. These conversational agents use advanced natural language processing and machine learning to understand complex human language, process voice commands and learn from past interactions (Koury & Murphy, 2023). Unlike the basic chatbots, they can answer more complex customer questions beyond what was scripted by the developer.

Lastly, the current state of AI chatbots are called generative AI chatbots. They are advancements of conversational agents as it includes machine learning tools such as transformers and this has let developers to train machine learning models on massive data sets to create generative AI chatbots (Koury & Murphy, 2023). The generative AI that we are familiar with would be ChatGPT or Jasper AI. They are both capable of generating new text with the provided input and the intention or the purpose of the generated text. This type of chatbot is beyond what was taught, they are now capable of learning the new information through what was given to them and correct them once there are mistakes. However, there are still ethical and legal concerns with the use of this generative AI chatbots in business settings such as inversion attacks. It is studied in recent papers that generation AI models are vulnerable to inversion attacks, providing the input text as output text (Hacker et al., 2023). This goes against the data protection regulation which can lead to confidential information leakage. Therefore, a careful evaluation is needed when adopting generating AI model into business settings.

2.2 AI-Service Quality

Since the proposition of SERVQUAL model in 1988, copious literatures have developed service quality model to fit into their research areas especially in the online context. Especially many of the studies focused on service quality of online shopping settings such as the, E-SERVQUAL (Yang &, Jun, 2002), eTailQ (Wolfinbarger & Gilly, 2003), WEBQUAL 1.0 to 4.0 (Barnes & Vidgen, 2001; Barnes & Vidgen, 2002) and so on. However, as AI service is a new kind of service that differs from traditional online services, existing service quality dimension cannot be applied directly. To overcome this conflict, AI service Quality model has been proposed by couple researchers.

Noticeable research on AI service quality is "Developing a service quality scale for artificial intelligence service agents" (Noor, 2022). This paper constructs, refines and validates multidimensional AISAQUAL scale through a series of pilot and validation studies. AISAQUAL scale is based on extant service quality research and established scale development techniques to contain 26 items across six dimensions. Proposed dimensions are efficiency, security, availability, enjoyment, contact and anthropomorphism. These dimensions were tested using seven-point Likert scale survey and it was found that these dimensions have significant effects on customer satisfaction, perceived value, and customer loyalty. Nonetheless, as this quality dimension is based on particular AI service agents (AISA), chatbot and virtual assistants, it needs to be validated with other AISA types.

Another recognizable paper would be "Classifying and Measuring the Service Quality of a AI Chatbot in Frontline Service" (Chen et al., 2022). This paper also has proposed a dimensions of AI chatbot service quality (AICSQ) to address the gap between existing dimensions and scales of service quality and new AI environment. This paper specifically focuses on the online retail AI chatbot services, which differs with previous research paper mentioned. This paper includes 7 second-order and 18 first-order constructs. The seven dimensions include, semantic understanding, close human-AI collaboration, human-like, continuous improvement, personalization, cultural adaption, and efficiency. They also have conducted nomological test to show that AICSQ dimensions positively influences consumer's perceived value and satisfaction of AI chatbot which effects intention of continuous use. This paper also has identified limitations in the scope of studies in terms of types of chatbot and industries.

Tab. 2-1 Existing AI Chatbot Service Quality Dimensions

AISAQUAL	AICSQ
Efficiency	Efficiency
Contact	Close-Human AI Collaboration
Anthropomorphism	Human-Like
Security	Continuous Improvement
Availability	Personalization
Enjoyment	Cultural Adaption
	Semantic Understanding

Overall, these two papers both have proposed dimensions to measure AI chatbot service level which is summarized in <Tab. 2-1>. Regardless of differing research area and methodology, they have come up with similar results. Both service quality dimensions include efficiency, close-human AI collaboration, which was referred to as 'Contact' in AISAQUAL model, and anthropomorphism, which was referred to as 'Human-Like' in AICSQ model. Both papers emphasized on effects of anthropomorphism on customer satisfaction as AI chatbot replaces the human work. Although anthropomorphism is considered to affect positively on overall customer satisfaction and perceived value, there are controversial studies related to it, which requires additional verification on the effects. The difference would be that AISAQUAL focused on the technological aspects of the AI service agents whereas AICSQ focused on the consumer contact aspects of AI chatbot. Lastly both models focused on conversational AI agents, such as chatbots and virtual assistants AI chatbots, which differs with the generative AI. Therefore, this paper aims to use the existing AI service quality dimensions and compare it with most frequently mentioned topics of generative AI service to verify if existing dimensions are qualified to be used to evaluate generative AI services.

3. Methodology

This paper aims to verify if existing dimensions of AI chatbots are qualified to be used to evaluate generative AI services, especially in the business settings. To do this, following steps shown in <Fig. 3-1 > will be used.

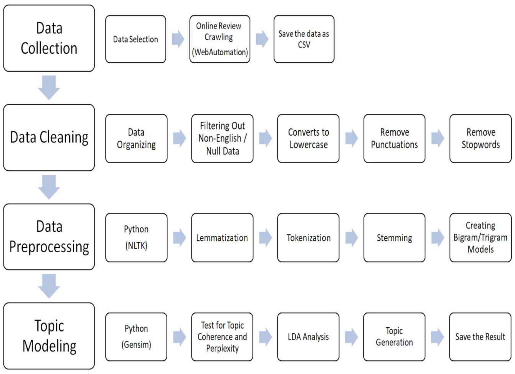

Fig. 3-1 The Method Roadmap

3.1 Data Collection

The reviews on four text-generative AI were obtained from a popular software evaluating platform G2.com, a peer-to-peer review site, where users can sign in with a Linkedin account and review the software products they use for business operations. Reviews are manually screened and voted on by the community for quality management.

The chosen software products are 'ChatGPT', 'Writesonic,' Jasper' 'Anyword'. These companies were chosen as they were categorized into synthetic media software companies that generates text. Only text-generating AI services were looked at as there weren't significant amounts of reviews written for other types of media such as video or picture. These companies had most reviews among the text-synthetic media software companies. Therefore, these companies were used in the data analysis.

The online data was collected using WebAutomation, an online crawling product. All reviews of chosen companies were extracted. There were 216 reviews for ChatGPT, 1,804 reviews for Writesonic, 1,211 for Jasper and 1,175 for Anyword. There were total of 4,406 reviews, and they were analyzed regardless of the product. The reviews were written from March 3rd, 2021, to May 28th 2023. The collected reviews were then saved into a csv file with columns including date, name, rating, header, review which is divided in to likes, dislikes and benefits earned. Among these columns, only likes and dislikes were used for analysis.

3.2 Data Cleaning

Gathered data were then organized into a single file. When merging, 'reviewer_liked' and 'reviewer_disliked' columns were merged into a single column called 'reviews'. After merging, all non-english and null data were removed, leaving 4,171 reviews left. All reviews were then divided into sentences, based on the period mark, for analysis to be done easily. It yielded 13,505 sentences to work with. All sentences were first converted to lower cases then cleaned unwanted noises such as punctuations, web URLs, tags, hashtags, numbers, special characters and others. The numbers were removed as they compose no meaningful information when generating topics. Then the stop-word dictionary was made and included the name of the companies to conduct analysis in non-product specific manner. More words were added in this dictionary during topic modelling section later.

3.3 Data Preprocessing

All extracted data needs preprocessing. All sentences were tokenized to each word and the words were lemmatized so words in third-person changes to first-person form. Lemmatized words were then reduced to their root form in stemming process. For example, all 'are', 'is', 'am', 'was', 'were', and 'being' were all reduced to its' root form 'be'. Only nouns were used to generate topic to determine the quality aspect, which are mostly represented using nouns. Lastly as some words create different meanings when used together, bigram and trigram dictionaries were made to provide better understanding. For example, in the sample, there was a word game-changer. If this word was looked separately, game and changer, it would mean different thing.

Before preceding to actual topic modeling, frequently mentioned noun-phrases were also examined to have better understanding of the sample. The results are shown in <Fig. 3-2>.

3.4 Topic Modeling

To conduct Latent Dirichlet Allocation (LDA) topic modeling, Gensim, an open-source library for unsupervised topic modeling will be used. In topic modeling, documents are represented with a mixture of topics and topics are represented with a probability

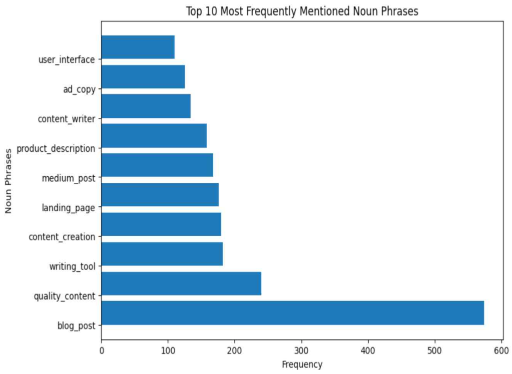

Fig. 3-2 Most Frequently Mentioned Noun Phrase

distribution over words and the documents are represented by a probability distribution topic (Steyvers and Griffiths, 2007). LDA is a popular topic modelling techniques to extract topics from given corpus. Then it assigns these topics to the document present within the same corpus. LDA has a benefit of being able to provide a full generative model and can handle long-length documents (Lee et al., 2010). However, as LDA topic modelling is an unsupervised learning, it differs every time running the code. Therefore, perplexity score was calculated and the result with the highest perplexity score will be used.

Before extracting the topics, the number of topics to be extracted should be determined. This could be done using the coherence test. The result of coherence test is shown as a graph in <Fig. 3-3> and <Fig. 3-4>.

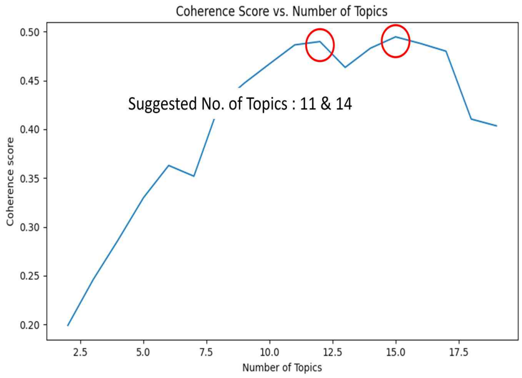

Fig. 3-3 Coherence Score Graph

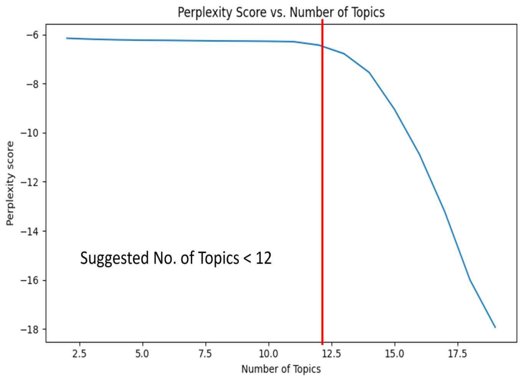

Fig. 3-4 Perplexity Score Graph

We could determine the number of the topics based on the graph, where it has highest and least fluctuation. Based on this, it can be seen that 11 and 14 topics are adequate number of topics to be generated. To confirm this, least perplexity score was driven for each number of topics, which is also included in Figure 3. It is stated that it is best to use the number of topics before the graph decreases drastically. It can be interpreted as to use the number of topics less than 12 as the graph quickly falls from 12.5 topics. Therefore, the final number of topics used was 11 topics.

4. Results & Analysis

4.1 Results

The results of LDA topic modelling is shown in <Tab. 4-1>. When processing, additional stop words were added such as 'thing, 'part', tool, 'product, 'lot, 'way, 'good', 'word' and other words that have too low or too high frequency were removed. The reason behind is most reviews that contained these words just started "The products are great", "The AI Chatbots are excellent product to use" and so on.

Tab. 4-1 Word Proposition of Each Topic Generated

Topic_Num	Word_Prop									
1	platform	idea	writer	service	job	block	generation	template	dislike	user
2	credit	software	quality	writer	option	email	plan	user	team	version
3	business	idea	conversion	ease	copywriting	dislike	need	issue	quality	work
4	service	marketing	template	platform	quality	limit	trial	fact	credit	text
5	idea	credit	user	sentence	option	trial	description	text	work	task
6	idea	work	price	user	option	app	quality	writer	process	website
7	day	quality	text	hour	result	marketing	output	assistant	month	section
8	writer	topic	information	credit	option	medium	response	process	output	software
9	credit	user	system	writer	ad	work	quality	app	service	interface
10	idea	text	credit	output	result	ability	fact	pricing	system	platform
11	website	user	option	platform	idea	ad	business	writer	information	outline

Nonetheless, just by solely observing the words, the topic was hard to identify as the some words were duplicated. Therefore, several steps were done to generate the topic.

First, using pyLDAvis, a visualization of the topics was viewed. This shows the inter-topic distance map via multidimensional scaling. To analyze the results, for All the topics were far detached from each other. The lambda, which shows the relevance metric can be adjusted. When λ=1.0, it means to sort words by their frequency within the specific topic and when λ=0.0, then it sorts words by their "lift", which is a term to represent how much a word's frequency sticks out in a topic above the baseline of its overall frequency in the model. According to the research, it is optimal to set λ=0.6, to get correct identification (Sievert & Shirley, 2014). <Fig. 4-1> is the visualization for topic 3. One thing to note is that when we select the word from this list, it shows which topic has highest component of it.

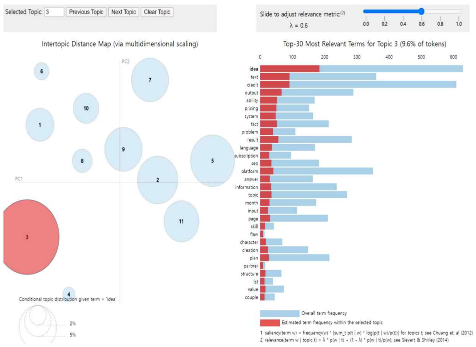

Fig. 4-1 pyLDAvis Visualization of Topic 3

Hence, for topic 3, it has highest component of 'idea' among other topics that included 'idea'. Some combinations of words were hard to be reduced into single topic by just looking at it. Therefore, we could interpret the result of topic 3 to be topic related to idea. Despite the usage of various methods to name the topics, some of the topics were hard to understand with the above methods so revision of original review was required to understand why some words were categorized into single topic. Therefore, frequently mentioned noun phrases extracted from <Fig. 3-2> were used to understand the word combinations. Despite the usage of various methods to name the topics, some of the topics were hard to understand with the above methods so revision of original review was required to understand why some words were categorized into single topic.

Similarly, other topics were named and detailed explanations were added to explain the meaning of each topic in <Tab. 4-2>.

Tab. 4-2 Topic Name and Detailed Explanation of Each Topic

Topic_Num	Topic Name	Detailed Explaination
1	Overall_Quality	Overall Quality of the Generative AI Chatbot Service
2	Website	UI/UX Design
3	Idea	Generating Idea/Creativity/Helping Writer's Block
4	Customer Service	Responding to Inquiries (Human)
5	Price	Pricing Option
6	Service Provided	Various Options Embedded in Software
7	Ease of Use	Easy to Use the Software
8	Datum	Amount of Datum in Software to Generate Answer
9	Output	Generated Ouput Quality _ According to One's Need
10	Availability	Being Able to Use 24/7
11	Speed	Time Required to Generate Answer

Lastly, each topic is categorized into existing service quality dimensions of AI service Chatbot and frequency is mentioned as shown in <Tab. 4-3>.

4.2 Discussion & Analysis

Based on the result and categorization shown in <Tab. 4-3>, topics are mostly categorized into efficiency. Some features such

Tab. 4-3 Existing Service Quality Dimension Categorization and Frequency of Each Topic

Topic_Num	Topic Name	Topic Frequency	AICSQ/AISAQUAL Dimensions
1	Overall_Quality	2872	-
2	Website	1329	Efficiency
3	Idea	979	Efficiency
4	Customer Service	915	Close-Human AI Collaboration
5	Price	920	Availability
6	Service Provided	1028	Personalization
7	Ease of Use	875	Efficiency
8	Datum	1104	Semantic Understanding
9	Output	896	Personalization
10	Availability	1116	Availability
11	Speed	1113	Efficiency

as close-human AI collaboration and semantic understanding, personalization and availability does impact user's choice of usage. On the other hand, anthropomorphism, cultural adaption, security, and enjoyment is not considered in the reviews as to impact user's choice of usage. Unlike previous studies which focused heavily on anthropomorphism, this study resulted in opposite ways. This can be explained in two ways.

First, this study is based on the online review platform where the reviewers are mostly business owners or employees. When AICSQ or AISAQUAL were studied, they targeted the end-users instead of the service providers. Therefore, the focus of the evaluation will differ from the business operating point of view.

Second reasoning would be due to the different types of AI service agents. Existing AI service quality dimensions are based on the AI chatbots and virtual assistants which can be viewed as the replacement of human force in the service industries. However, generative AI is considered to be another type of service providers where 'tool' aspect is more preferred. Especially for the text generating AI, their core job is to understand the input and intention and generate a text accordingly. For instance, when we ask a professional to create an ad-copy, we would not consider the names, looks, and personality when choosing the professionals. We would consider more of technological aspects such as experience, skills and so on. Therefore, the measuring criteria would differ as generative AI's task it not related to the work where friendliness is required.

5. Conclusion

This paper has analyzed the review data of four selected synthetic text media software; ChatGPT, Writesonic, Jasper, and Anyword. The reviews were written by the business owners or employees and total of more than 4000 reviews were used to analyze. LDA topic modeling using Gensim library was done to extract the topics from the reviews. In result 11 topics were made with 10 keywords each. Nonetheless, unlike other papers regarding topic modelling, where highest weight keywords are mostly chosen as the topic name, this analysis could not follow same structure as some words were duplicated and could not be removed. Therefore, visualization program pyLDAvis and analyzing most frequently mentioned noun phrases were done to name all the topics extracted. However, some of the topics were difficult to understand with the above methods so revision of original review was required to understand the categorzation of particular words. After the topic extraction, each topic was categorized into related AICSQ and AISAQUAL dimensions to make direct the comparison. Despite it being able to fit all the topics into existing dimensions from AICSQ or AISAQUAL, it only included parts of the existing dimensions, mainly availability and efficiency. Certain dimensions such as anthropomorphism, cultural adaption, security, and enjoyment were not considered in the reviews, which contrasts with the previous studies which highlighted the importance of anthropomorphism. One possible reasoning could be the type of AI service agents that the study targeted at. AICSQ and AISAQUAL dimensions were based on the chatbot or virtual assistant AI services where close interaction with end-users are fundamental. However, the text-generating AI does not require close interaction with the users. Instead, they require semantic understanding to catch the user's intention.

By comparing generative AI chatbots' quality factors with existing quality dimension associated with AI chatbot services, this study extends the existing literature in the field. It challenges the conventional understanding of AI chatbot service quality by demonstrating that the dimensions of quality for generative AI are not entirely aligned with those of interactive AI agents, such as virtual assistants. This finding suggests for new models or dimensions that might be more appropriate for understanding and evaluating the unique characteristics of generative AI chatbots depending on different users.

From a practical perspective, this study offers crucial insights for both developers and businesses considering the adoption of generative AI chatbots. It highlights the importance of focusing on quality dimensions like efficiency and availability, which are more critical for generative AI tools than for traditional interactive AI especially when adopting into business environment. This knowledge is particularly valuable for developers aiming corporate users, guiding them to prioritize features that enhance business processes and decision-making, rather than user engagement. For businesses, these findings underline the need to carefully evaluate how these chatbots align with their operational goals and contribute to overall productivity, rather than relying solely on conventional user satisfaction metrics. This tailored approach can lead to a more strategic and effective integration of generative AI in business environments.

There are several limitations on this study. First, the research is based on single website reviews. Although several products were chosen to minimize the bias caused by single sample, the website reviewers are mainly English-speaking business employees which does not quite capture the general quality of the generative AI. Moreover, as they are using generative AI service to generate profit, efficiency is considered significantly in the reviews. This result may change with different user segments. Moreover, considering that reviews are mostly written by the people who are satisfied with the products, especially in the software where free trial is possible, the results could be rated higher than actual user satisfaction. Usage of multiple review websites or survey could be done to reduce this bias in the future.

Further research could be done by following. For it to be considered as quality dimension, following surveys should be conducted to see if it actually affects the end user's customer satisfaction, perceived value, and intention of continuous use. Additionally, this study has only identified the quality dimensions through the online reviews. Further study on each dimension's impact on overall rating could be done to analyze whether certain dimension affects the rating positively or negatively.

[References]

[1] AMR(2020), Intelligent Virtual Assistant (IVA) Market to grow at 33% CAGR during forecast period (2020-2025) - Insights on Growth Drivers, Size and Share Analysis, Key Trends, Leading Players, and Business Opportunities: Adroit Market Research. www.globenewswire.com/news-release/2020/02/24/1988963/0/en/Intelligent-Virtual-Assistant-IVA-Market-to-grow-at-33-CAGR-during-forecast-period-2020-2025-Insights-on-Growth-Drivers-Size-and-Share-Analysis-Key-Trends-Leading-Players-andBusin.html (accessed 06 June 2023).

[2] Barnes, S.J. and Vidgen, R.(2001). An

evaluation of cyber-bookshops: the WebQual method. *International Journal of Electronic Commerce*, 11-30.

[3] Barnes, S.J. and Vidgen, R.T.(2002). An integrative approach to the assessment of e-commerce quality. J. Electron. Commer. Res., 3(3), 114-127.

[4] Chen, Q., Gong, Y., Lu, Y., and Tang, J. (2022). Classifying and measuring the service quality of AI chatbot in frontline service. Journal of Business Research, 145, 552-568.

[5] Davenport, T., Guha, A., Grewal, D. and Bressgott, T.(2020), How artificial intelligence will change the future of marketing, Journal of the Academy of Marketing Science, 48(1), 24-42.

[6] Gartner(2019). How to manage customer service technology innovation. Gartner. https://www.gartner.com/smarterwithgartner/27297-2

[7] Gartner(2021, October 18). Gartner identifies the top strategic technology trends for 2022. Gartner. https://www.gartner.com/en/newsroom/press-releases/2021-10-18-gartner-identifies-the-top-strategic-technology-trends-for-2022

[8] Hacker, P., Engel, A., and Mauer, M.(2023). Regulating ChatGPT and other large generative AI models. arXiv preprint arXiv:2302.02337.

[9] IBM(2022, May). IBM Global AI Adoption Index 2022. IBM. https://www.ibm.com/downloads/cas/GVAGA3JP

[10] Koury, L. and Murphy, T.(2023). The evolution of chatbots from ELIZA to Bard. TechTarget. TechTarget. Retrieved June 12, 2023, from https://www.techtarget.com/searchcustomerexperience/infographic/The-evolution-of-chatbots-and-generative-AI.

[11] Lawton, G.(2023, April 18). Generative AI Ethics: 8 biggest concerns. Enterprise AI. https://www.techtarget.com/searchenterpriseai/tip/Generative-AI-ethics-8-biggest-concerns

[12] Lee, S., Baker, J., Song, J., and Wetherbe, J.C.(2010). An Empirical Comparison of Four Text Mining Methods. Proceedings of the 43rd Hawaii International Conference on System Sciences, 1-10.

[13] MarketsandMarkets. (2023, May 4). AI as a service market size & analysis, trends, growth, Revenue Forecast - 2028 & Opportunities: MarketsandMarketsTM. MarketsandMarkets. https://www.marketsandmarkets.com/Market-Reports/artificial-intelligence-ai-as-a-service-market-121842268.html

[14] Noor, N., Rao Hill, S., and Troshani, I. (2022). Developing a service quality scale for artificial intelligence service agents. E*uropean Journal of Marketing*, 56(5), 1301-1336.

[15] OpenAI. (2023). Introducing chatgpt. Introducing ChatGPT. https://openai.com/blog/chatgpt

[16] Poole, D.L. and Mackworth, A.K.(2010), Artificial Intelligence: Foundations of Computational Agents, Cambridge University Press, Cambridge.

[17] Sievert, C. and Shirley, K.E.(2014). LDAvis: A method for visualizing and interpreting topics.

[18] Steyvers, M. and Griffiths, T.(2007). Probabilistic Topic Models. In T. Landauer, D. McNamara, S. Dennis, & W. Kintsch (Eds.), Latent Semantic Analysis: A Road to Meaning. Hillsdale, NJ: Laurence Erlbaum.

[19] Thomas, R.(2020, January 8). Ai in 2020: From experimentation to adoption. THINK Blog. https://www.ibm.com/blogs/think/2020/01/ai-in-2020-from-experimentation-to-adoption/

[20] Thormudsson(2023), Rate of generative AI adoption in the workplace in the United States 2023, by industry, Statista, https://www.statista.com/statistics/1361251/generative-ai-adoption-rate-at-work-by-industry-us/ (accessed 06 June 2023).

[21] Weizenbaum, J.(1966). Eliza-a computer program for the study of natural language communication between man and Machine. Communications of the ACM, 9(1), 36-45.

[22] Wolfinbarger, M. and Gilly, M.C.(2003). eTailQ: Dimensionalizing, measuring and predicting retail quality. Journal of Retailing, 79(3), 183-198.

[23] Yang, Z. and Jun, M.(2002). Consumer perception of e-service quality: from internet purchaser and non-purchaser perspectives. Journal of Business Strategies, 19(1), 19-42.

Ahn, YaeEun (ayu7045097@snu.ac.kr)

Ahn, YaeEun is a currently a MS student at Business School, Seoul National University. She is interested in studies of Quality Management and Service Industry Operations Management.

Oh, Jungsuk Oh (joh@snu.ac.kr)

Oh, Jungsuk is a Professor of Operations Management at Seoul National University. He received S.B. in Management Science from M.I.T. and M.S. in Operations Research as well as Ph.D. in Management Science & Engineering from Stanford University. Prior to joining Seoul National University, he worked as a Professor at Korea Advanced Institute of Science and Technology. His research has focused on development and applications of platform economics and ecosystem strategy in ICT, mobile and new economies related to the 4th industrial revolution. The importance of value appropriation in the context of ecosystem formation has been emphasised throughout. Additional research interests are the role of business analytics and evolution of quality management practice in the 21st century business environment.

LDA토픽 모델링을 활용한 생성형 AI 챗봇의 탐색적 연구 : 기존 AI 챗봇 서비스 품질 요인과의 비교

안예은* · 오정석**

요약

인공 지능 (AI), 특히 텍스트 생성 서비스 분야에서의 발전은 두드러지게 나타나고 있으며, AI-as-a-Service (AIaaS) 시장은 2028년까지 550억 달러에 달할 것으로 예상된다. 본 연구는 합성 텍스트 미디어 소프트웨어의 품질 요소를 탐구하였으며, 이를 위해 ChatGPT, Writesonic, Jasper, 그리고 Anyword와 같은 산업의 주요 서비스에 주목하였다. 소프트웨어 평가 플랫폼에서 수집된 4,000개 이상의 리뷰를 바탕으로, Gensim 라이브러리를 활용한 잠재 디리클레 할당 (LDA) 주제 모델링 기법을 적용하였다. 이 분석을 통해 11개의 주제가 도출되었다. 이후 이 주제들을 AICSQ 및 AISAQUAL과 같은 기존 논문에서 다루었던 AI 서비스 품질 차원과 비교 분석하였다. 리뷰에서는 가용성 및 효율성과 같은 차원이 주로 강조되었으며, 이전 연구에서 중요하게 여겨졌던 사람다움과 같은 요소는 본 연구에서 강조되지 않았다. 이러한 결과는 AI 서비스의 본질적 특성, 즉 사용자와의 직접적인 상호작용보다 의미론적 이해에 더 중점을 둔다는 특성 때문으로 해석된다. 본 연구는 단일 리뷰 원천 및 평가자들의 인구 통계의 특정성과 같은 잠재적 편향을 인정하며, 향후 연구 방향으로는 이러한 품질 차원이 사용자 만족도에 어떻게 영향을 미치는지, 그리고 개별 차원이 전체 평점에 어떻게 영향을 미치는지에 대한 깊은 분석을 제안한다.

Keywords : *품질 차원, 생성형 AI, 서비스품질, LDA 토픽모델링, 리뷰분석*

* 제1저자, 서울대학교 경영대학 대학원 학생, ayu7045097@snu.ac.kr
** 교신저자, 서울대학교 경영대학 교수, joh@snu.ac.kr