# A New Era for Data Analysis in Qualitative Research: ChatGPT!

# Mert Şen

Inonu University, Turkey
Inonu University, Tur

#### Şevval Nur Şen

Inonu University, Turkey https://orcid.org/0000-0002-9452-8206

#### Tuğrul Gökmen Şahin<sup>1</sup>

Inonu University, Turkey
In https://orcid.org/0000-0003-2107-9670

#### Abstract

Today, the use of software in qualitative research analysis is rapidly becoming widespread among researchers. Researchers manage large data sets using features such as editing data, transcribating, creating codes, and searching within data. However, while the data analysis uses software in a format, the analysis of the essence of the data is done by researchers. An AI language model, ChatGPT, released by the OpenAI company, has features such as text editing, creation, and abbreviation. In this research, considering the characteristics of ChatGPT, an answer was sought to the question of whether it can be used in data analysis for qualitative research. In this research, the case study of the qualitative research method was preferred. The data of the research consists of interview texts of two participants from an unpublished study. The texts were subjected to qualitative research analysis process through ChatGPT-4. Data analysis was done in two separate ways, specifying code, category and theme and not specifying. In conclusion, it has been found that ChatGPT can create code, category and theme, cite directly from within the text, interpret data sets, and analyze the meaning at the core of the data sets. In this context, the availability of ChatGPT in data analysis in qualitative research has been discussed.

Keywords: Qualitative Research, Qualitative Data Analysis, Computer Assisted Qualitative Data Analysis Software (CAQDAS), Artificial Intelligence, ChatGPT

#### Introduction

Qualitative research is a method which is based on determining what the phenomenon means beyond determining the cause and impact that has been emphasized in the last quarter century or defining some distributions, not only on theoretical knowledge, but also on a philosophical background. Thus, the traditions and approaches leading to qualitative research vary. Data collection, data analysis, the researcher's epistemological and ontological attitudes, the nature of the research question, many hidden and open impacts are correlatively affecting each other (Clarke et al., 2021). Therefore, in qualitative research, results are focused on individual, not numbers.

Qualitative research has the potential to present different results when the same research is done at different times and in different environments. The common element of this method, which is quite flexible and can vary depending on the situation, can be addressed as the way each researcher will be monitoring when analyzing. After acquiring data for qualitative analysis, the researcher must first organize the data,

1 Corresponding author

# Volume: 11

**OPEN ACCESS** 

Special Issue: 1

Month: October

Year: 2023

E-ISSN: 2582-1334

Received: 21.08.2023

Accepted: 25.09.2023

Published: 01.10.2023

#### Citation:

Şen, M., Şen, Ş. N., & Şahin, T. G. (2023). A New Era for Data Analysis in Qualitative Research: ChatGPT!. Shanlax International Journal of Education, 11(S1), 1–15.

DOI: https://doi.org/10.34293/ education.v11iS1-Oct.6683



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License categorize and encode them to process and develop a system to deliver the data. (<u>Clarke et al., 2021</u>). To make this easier, Computer Assisted Qualitative Data Analysis Software (CAQDAS) is consulted with the advancement of technology.

Although it is known that software created to assist in the analysis of text and content in human sciences has been used since 1960, it is often found to be used for statistical purposes up to 1980 (Cypress, 2019). The use of CAQDAS in qualitative research has been a topic of discussion between its defenders and critics since the day it emerged. (Jenskin et al., 2023). Today, technology advances and the concept of time for a technology-era individual are becoming very important, making it possible to use software for other than statistical purposes.

In qualitative research, analysis of data can be done manually, of course, but depending on the changes and developments mentioned, many qualitative researchers today may be using various software for qualitative data analysis (<u>Cypress</u>, 2019).

Overall, CAODAS covers supporting, complementary programs and software that enable data management and analysis to be objective in qualitative data analysis (Vignato et al., 2022). In the last 20 years, it has been seen to grow based on its active use for qualitative and mixed-method research (Clarke et al., 2021). CAQDAS incorporates many software, such as NVivo, ATLAS.ti, MAXQDA, which offers many advantages in terms of data organization and the efficiency of coding in qualitative and hybrid studies The increased auditability of the data, allowing multiple researchers to encode, has many positive features for qualitative research in terms of the rigidity and reliability of the findings (Clarke et al., 2021). Since CAQDAS allows large and complex data to be organized and simplifying analytical processes such as word, text search, researchers can spend more time on understanding and specific analysis of data (Jenskin et al., 2023).

As the CAQDAS packages have evolved rapidly over the last decade, especially new users, have a number of problems with the use of CAQDAS. It can be a problem for users to choose which software to use based on their needs, when to learn which software, and what it costs. Also, qualitative researchers who are not capable of using computers can avoid using CAQDAS. It may take time to use software for researchers who are also capable of using computers, which may prevent the use of CAQDAS (Niedbalski & Ślęzak, 2022). While CAQDAS use provides researchers with benefits in transkript, edit, classification, etc., it is also known that it is not able to find meaning, interpret data, analyze the content of data (Kalpokas & Radivojevic, 2022). This deficiency in CAQDAS may be fulfilled by Artificial Intelligence (AI) technologies that have evolved in recent years.

ChatGPT is a language model developed by OpenAI, an AI research and distribution company based in San Francisco, California. Based on the GPT-3, which the company launched in 2020, ChatGPT is set up to engage in dialog with the user (Stokel-Walker, 2022). ChatGPT is a prototype dialog-based artificial intelligence chat robot that can understand human natural language and produce highly detailed, human-like text (Lock, 2022). ChatGPT's dialog format has the ability to answer questions that follow, accept errors, question false assumptions, and reject inappropriate requests (OpenAI, 2022). The description ChatGPT made about itself was; "ChatGPT is a language model developed by OpenAI. Language models are artificial intelligence systems that try to understand how language works and are used in applications such as language processing, text prediction, and language translation. ChatGPT is designed for a more natural language use of a chatbot and has the ability to chat with people. ChatGPT works as a language model and uses language examples while learning. In this way, ChatGPT can learn the structure and characteristics of the language that people use and use natural language to provide answers." (ChatGPT, 2022).

As of November 2022, when it was released under the name ChatGPT-3.5, it was followed with interest by more than 1 million people. ChatGPT is used in two versions, GPT-3.5 and GPT-4. For now, it is enough to enter the OpenAI site to use ChatGPT and log in using the mail address or Google account and log in to the phone number. To use the ChatGPT-4 version, it is necessary to pay extra. There are significant differences between GPT-3.5 and GPT-4. GPT-4 uses more data than previous models and is more reliable. GPT-4 gives more accurate results than GPT-3.5 in complex operations. In addition, GPT-4 performs better in English and in different languages than other language models. Most importantly, it still is not fully reliable (it "hallucinates" facts and makes reasoning errors). GPT-4 is more successful in rejecting harmful (bomb making, harmful chemical components, etc.) requests (OpenAI, 2023a). Another important advantage of GPT-4' is that it has Plugins (Add-ons). GPT-4 accesses the internet via Plugins, reaching up-to-date information, viewing documents such as PDF (OpenAI, 2023b).

Millions of users used ChatGPT to write poetry, communicate with it, or answer simple questions (Sier, 2022). Most users using the app have been greatly influenced by Chat GPT's performance and potential (Haque et al., 2022). ChatGPT provides answers, solutions, and explanations to complex questions, including potential methods for solving maths problems, code writing and gaming, homework articles and more. ChatGPT can be used to easily extract code errors, and AI not only corrects the error, it also explains to the user where they are wrong and what they can do to correct them (Naidu, 2022). Also, ChatGPT has the ability to tell us what the disease can be when body disorders are mentioned, to simplify complex topics for children, to write stories based on what the user expresses (Özcan, 2022). As well as being one of the most powerful chatbots ever made, ChatGPT has the feature of providing text to it, abbreviating and detailing (Kirmani, 2023). ChatGPT is used free of charge as of December 2022. For now, to use ChatGPT, simply go to the OpenAI site and log in using email address or Google and enter the phone number with the incoming code.

Although ChatGPT has been released as a language model, it can be thought that it can be used as CAQDAS as ChatGPT has features such as text abbreviation, interpretation and detailed expression. In this research, the question of whether ChatGPT can be used in data analysis for qualitative research, taking into account its analysis ability, is sought. In this study, how ChatGPT is used in qualitative data analysis and how it performs qualitative data analysis was investigated. For this purpose, answers to the following questions were sought;

- How is ChatGPT used in qualitative data analysis?
- What are the qualitative data analyzes that ChatGPT has done?

# Method

# **Research Design**

This research uses a case study design from qualitative research method. The case study is a qualitative approach where the researcher gathers detailed and in-depth information about real life, an up-to-date situation, or circumstances over a specific period of time, depicting a situation or creating themes (Creswell, 2013). In this research, the case design was preferred because it examined how the ChatGPT-4 model (CG) performed qualitative data analysis.

# **Study Group**

In this study, the interview documents of an unpublished study subject to the university compliance process were used to evaluate the analysis process of CG. Since the language of the interview is Turkish, texts were uploaded to CG in Turkish and the analysis of the texts was made in Turkish. Detailed description of the interview sets used is included in Table 1;

# Table 1 Demographic Information on the Data Sets Reviewed

Data Set Used	Word Count
Ali (First Participant)	714
Beyza (Second Participant)	1097

# **Data Collection Tools**

In this study, the data obtained from the CG and the data included in the work group were obtained in written documents, so the document was used as a data collection tool. Documents are among the data collection tools that researchers frequently prefer to as they provide benefits such as include detailed information, access to large amounts of data, provide information about unobservable situations, and be cost effective (Celebi & Orman, 2021).

#### **Data Collection Process**

Interviews for research subject to the university adaptation process have been transcripted. Transcripts are uploaded in CG in Turkish and certain commands are given to CG. The Chat GPT-4 model was used in the data analysis process as it responds more detailed, reliable and accurate to longer commands. In exchange for the commands written by the researchers to prevent data loss, the data obtained from the analyzes made by CG were copied from their own intermediate and stored in digital media. The steps taken in the data collection process are given below;

Table 2 Steps App	plied
-------------------	-------

Step	Action Taken	Explanation	
1st Step	Ali's 1st Analysis	At this step, there is no limitation to CG. The text was asked to create code category and theme. A direct quote was requested for the created codes.	
2nd Step	Ali's 2nd Analysis	At this step, there is no limitation to CG. The text was asked to create code category and theme. A direct quote was requested for the created codes. It we done to confirm the first step.	
3rd Step	Ali's 3rd Analysis	In this step, CG was given theoretical information, categories were specified and analysis was made within the boundaries drawn. A direct quote was requested for the created codes.	
4th Step	Beyza's 1st Analysis	At this step, there is no limitation to CG. The text was asked to create code category and theme. A direct quote was requested for the created codes.	
5th Step	Beyza's 2nd Analysis	In this step, CG was given theoretical information, categories were specified and analysis was made within the boundaries drawn. A direct quote was requested for the created codes.	
6th Step	Joint Analysis of Beyza and Ali	In this step, CG was asked to show similar and different aspects of the analyzes he had previously done.	

#### Validity and Reliability

The measures taken to increase the internal validity of the research are explained in detail in the collection of data. Data from documents 1 and 2 & CG are given by quoting directly in the findings section without changing. In order to increase external validity, the research pattern, working group, data collection tools, implementation process, analysis of the data obtained, how the findings are organized and the role of the researchers are explained in detail in the relevant sections. In order to increase the internal security of the research, that is, its consistency, the data obtained from documents 1 and 2 were analyzed in two different ways and the code and category created, themes were checked. In order to increase the external trust of the research, the research data were properly discussed in the results section. Whether the results and findings section provides consistency has been discussed and agreed among researchers. The conclusion was confirmed to a qualitative research specialist whether the part of the findings was consistent with the discussion section.

# Findings

# **Analyze Like Chatting!**

We copied and pasted the data into the chat box on the interface of CG to analyze the data sets we used in the research. We added and sent the question we wanted to ask before the data we put in quotes. We did not start a new conversation for the questions we asked in progress, and after the first question we asked other questions. We worked on other datasets using "New Chat" to analyze other data. In addition, the old "Chats" we have done is stored on the left side of the interface. CG's analysis of our data took place in the range of 1-20 seconds. Figure 1 can be examined for the intermediate face of CG:

Image:	- mead D	C Mani art a	
Compared and the second s		Bit many the set of the set	
D W Windowski w State St	0 0 0	• Use its Diput the activity on the head of the state	
a Bard a manage Ba	8 8 0	zw. Mi major najmunatarion popujanim sava junitariu maju. Ai taujaji hajada attai, Danamara partitika ministera o najmini kupajen, kinaji tami oraja.	
	e	Sent in terminate in the	

Figure 1 ChatGPT Usage Interface

#### **Telling the Purpose of the Research**

In the first step of our research, we explained the actions to be taken to CG. For this, we commanded what subject we were researching, who was speaking, what we wanted from CG. We added the quotation marked speech under the command we wrote. While the speeches are sub-subject in the transcript of the interview, the speeches are reflected

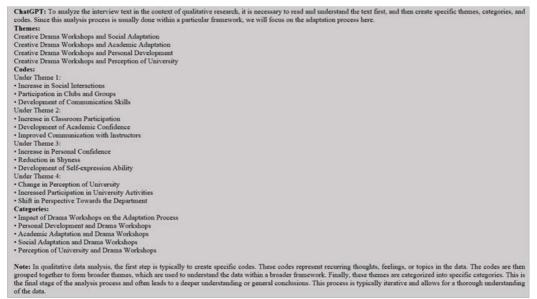
consecutively on the screen when the text is added to CG. However, this does not adversely affect CG's analysis process. After defining the purpose of the research, we asked CG to divide the text into themes, categories and codes for our qualitative research. We copied the conversation text from the Word file and pasted it into CG's interface.

**Researcher:** I conducted a research on the impact of creative drama on the university adaptation process. The text in quotation marks below represents the conversation I had with the participant. M (Mert) refers to the researcher, and A (Ali) refers to the participant. Can you help me categorize and identify the themes and codes for this text in the context of qualitative research?" (Interview Text) M: How did participating in drama workshops make you feel? Can you elaborate? A: At first, I was hesitant because I didn't know the exact content..."

# Figure 2 Command Describing the Purpose of the Research

#### **ChatGPT's First Analysis**

CG's analysis of the text after the first command is written can vary between 10-20 seconds. CG analysis can interrupt analysis in analyzes after long commands. When this situation is encountered, it is possible to continue the analysis with the option "Continue generating". In addition, CG responds again by offering the option "Regenerate response" if its response is not liked. The analysis made by CG after the first command is as follows;



**Figure 3 First Analysis Results** 

When Figure 1 is examined, it is seen that CG first explained the actions it took. It is then seen that it creates four themes, creates codes under themes, establishes a relationship between codes and themes, and creates categories. It also appears that the theme, code and category as a note make a statement regarding the use of qualitative research.

In this section, the text was analyzed, but no finding was found as to which code the participant obtained from which statements.

# **Detailed Explanations, Requesting Direct Quote**

In the first analysis of CG, we asked CG to make a detailed explanation as it was seen that it

only included the names of the themes, codes and obtained under the themes. For this, we used the categories. We asked what statements the participant command in Figure 4;

Researcher: From which statements of the participant did you derive the code "Increase in Social Interactions" under Theme 1?

ChatGPT: I derived the code "Increase in Social Interactions" from the following statements of the participant:

"Drama had a significant impact on me socially." - This statement expresses the participant's general view that dramatic workshops influenced their social interactions.

"It was very beneficial for me in terms of conversing and improving my communication with people." - Here, the participant mentions that drama workshops contributed to social interactions and overall communication skills.

"When we participated in activities with the friends there, we became more social through those activities." - This statement indicates an increase in social interactions during drama workshops. "It provided me with a lot of benefits in terms of socializing during the university adaptation process." - Here, the participant acknowledges the positive impact of drama workshops on social interactions. "Because it improved my communication, I joined many clubs and groups through drama workshop activities. We formed groups and organized various events." - This statement highlights how drama workshops facilitated social interactions and the participant's involvement in social activities. These statements indicate an increase in social interactions due to the participant's engagement in drama workshops, thus forming the code "Increase in Social Interactions."

**Figure 4 Command for Requesting Direct Quote** 

When we asked CG to explain which code it obtained from which statements of the participant, we found that CG's analysis was quite correct. When we asked CG to explain which statements of the participant obtained, which code he obtained, we found that CG's analysis was quite correct. It is seen that CG creates codes by extracting meaning from the expressions in the text during the analysis process. Direct quotes selected for the generated codes appear to be extremely accurate. It has also been determined that CG produces different codes from the same paragraph. He also used multiple expressions in the text when creating the CG code and made explanations as to why he created the code. When we repeat this process for all other codes and found that CG transferred direct quotes from the text to a large extent. Figures 5, 6, 7 and 8 contain quotations and codes made by CG;

Theme	1-Creative Drama Workshops and Social Adaptation				
Code	Increase in Social Interactions	Participation in Clubs and Groups	Development of Communication Skills		
Expressions	I derived the code "Increase in Social Interactions" from the following statements of the participant: "Drama had a significant impact on me socially." - This statement expresses the participant's general view that dramatic workshops influenced their social interactions. "It was very beneficial for me in terms of conversing and improving my communication with people." - Here, the participant mentions that drama workshops contributed to social interactions and overall communication kills. "When we participated in activities with the friends there, we became more social through those activities." - This statement indicates an increase in social interactions during drama workshops. "It provided me with a lot of benefits in terms of socializing during the university adaptation process." - Here, the participant acknowledges the positive impact of drama workshops on social interactions. "Because it improved my communication, I joined many clubs and groups through drama workshop activities. We formed groups and organized various events." - This statement individual interactions and the participant's movelement in social alteriticipant's engagement in drama workshops, theiriticipant's engagement in drama workshops, the social interactions."	I derived the code "Participation in Clubs and Groups" from the following statements of the participant: "Because it improved my communication, I joined many clubs and groups through drams workshop activities. We formed groups and organized variou events." - This statement indicates that the participant joined different clubs and groups and organized events through drama workshops. "There was a cycling club, and we organized cycling tours. There's also an Ultrashan group, and we play indoor soccer matches with them. We were involved in various activities like that." - Here, the participant mentions that they participated in different clubs and groups and engaged in activities through drama workshops. "The confidence and sense of assurance that drama provided, well, it kind of motivated me aliminated that hesistino." - This statement highlights how drama workshops helped the participated in different shows and uncertainty in joining clubs and groups. Thes statements demonstrate that drama workshops encouraged the participant's participation in clubs and groups and facilitated their involvement. Therefore, the code "Participation in Clubs and Groups" is formed.	I derived the code "Development of Communication Skills" from the following statements of the participant: "Drama had a significant impact on me socially. It benefited me in terms of expressing myself and improving my speech. It also helped me in conversing and enhancing my communication with people." - This statement indicates how drama workshops improved the participant's ability to express themselves and their overall communication skills. "When we participated in activities facilitated more effective and furtiful communication with friends, we socialized a tothrough those activities. "The participant mentions that participating in activities facilitated more effective and furtiful communication with their friends. "Because it improved my communication kills, on establishing interpersonal relationships." - Here, the participant expression overall communication skills and influenced their participation in clubs and groups. "It has a significant effect on communication skills, my to the communicate with others, especially in establishing interpersonal relationships." These statements themostrate how drama workshops contributed to the development of the participant's communication skills and how this development ranalated into their social interactions and participation in clubs'groups. Therefore, the code "Development of Communication Skills" is formed.		

Figure 5 Theme 1

Theme		2-Creative Drama Workshops and Academic Adaptation		
Code	Increase in Classroom Participation	Development of Academic Confidence	Improvement in Communication with Instructors	
Expressions	From the participant's statements, I derived the code "Increase in Classroom Participation" from the following excerpts: "We have communication with the instructors. Although we are not very close, we chat with our regular class teachers. But it had quite an impact on our behavior in the classroom because I used to be a bit shy in class. Through the drama process, it greatly helped me gain confidence. So now, I can speak in class, answer questions, and ask questions. It has been very beneficial for me in that regard." - This statement demonstrates how drama workshops helped the participant become more active and engaged in the classroom. "So, mention that it has some impact on being entive in the class and partially increasing communication with the teachers." - Here, the participant acknowledges that drama workshops helped them become more active in the classroom, which in turn partially improved their communication with the teachers. These statements show how drama workshops influenced the participant's increased classroom participation and how they benefited from it. Therefore, the code "Increase in Classroom Participation" is formed.	From the participant's statements, I derived the code "Development of Academic Confidence" from the following excerpts: "The confidence and sense of assurance that drama provided, well, it kind of motivated me to join them. I was hesitant to join clubs and groups because of my shyness, but drama eliminated that hesitation." - This statement highlights how drama workshops increased the participant's self-confidence and overall academic confidence. "We have communication with the instructors. Although we are not very close, we chat with our regular class teachers. But it had quite an impact on our behavior in the classroom because I used to be a bit shy in class. Through the drama process, it greatly helped me gain confidence. So now, I can speak in class, answer questions, and ask questions. It has been very beneficial for me in that regard." - This statement demonstrates how drama workshops improved the participant's elf-confidence and academic setting and contributed to their ability to actively participate in class. These statements illustrate how drama workshops enhanced the participant's elf-confidence and academic confidence. Therefore, the code "Development of Academic Confidence" is established.	From the participant's statements, I derived the code "Improvement in Communication with Instructors" from the following excerpts: "We have communication with the instructors. Although we are not very close, we chat with our regular closs teachers. But it had quite an impact on our behavior in the classroom" - This statement demonstrates how drama workshops improved the participant's communication with instructors. "So now, I can speak in class, answer questions, and ask questions." - This statement indicates that the participant is more comfortable and open in their these two statements illustrate how drama workshops led to an improvement in the participant's communication with instructors. Therefore, the code "Improvement in Communication with Instructors" is established.	

#### Figure 6 Theme 2

Theme	3-Creative Drama Workshops and Personal Development					
Code	Increase in Personal Confidence	Reduction in Shyness	Development of Self-expression Ability			
Expressions	From the participant's statements, I derived the code "Increase in Personal Confidence" from the following excerpts: "The confidence and sense of assurance that drama provided, well, it kind of motivated me to join them. I was hesitant to join clubs and groups because of my shyness, but drama eliminated that hesitation." - This statement highlights how drama workshops increased the participant's personal confidence. This statement also demonstrates how drama workshops enhanced the participant's personal confidence - This statement also demonstrates how drama workshops enhanced the participant's personal confidence and adversed their shyness. These two statements indicate an increase in the participant's personal confidence is established.	From the participant's statements, I derived the code "Reduction in Shyness" from the following excepts: "Af first, I was acting shy because I didn't know the content completely, but once it started and I got to know you and the other friends and participated in activities, it made me happier and made me feel better." - This statement indicates that drama workshops reduced the participant's initial slyness. "I am a bit shy, actually. I'm not really a very social person. Drama had a significant impact on me socially." - The participant mentions being shy, but also highlights how drama workshops helped in reducing this shynes. "The confidence and sense of assumace that drama provided, well, it kind of motivated me to join them. I was hesitant to join clubs and groups because of my shyness, but drama eliminated that hesitation." - This statement demonstrates how drama workshops reduced the participant's shyness in joining clubs and groups. These statements indicate that drama workshops reduced the participant's shyness, leading to increased social engagement. Therefore, the code "Reduction in Shyness" is formed.	From the participant's statements, derived the code "Development of Self expression Ability" from the followin excerpts: "It greatly helped me in expressing mysel and improving my speech in that regard. - The participant indicates that dram workshops benefited their ability to express themselves and improve their speech. "So, it has quite an impact on interacting with people, establishing interpersona relationships." The participant highlight how drama workshops helped them in communicating with others and building interpersonal relationships. "So now, I can speak in class, answe questions, and ask questions." The participant mentions how dram workshops have helped them develop the ability to speak in class, respond to questions, and ask questions. These statements indicate that dram workshops contributed to the participant development of self-expression ability Therefore, the code "Development of Self-expression Ability" is established.			

#### Figure 7 Theme 3

Researcher: These statements are not the participant's own words, could you quote directly from the participant's own words?

**ChatGPT:** I apologize, as far as I understand, we are not based on a text directly coming from a participant. Therefore, I cannot provide the direct statements of a participant. The above responses, while making a general inference about how drama workshops might change the participant's view of the department, do not use a participant's direct statements.

If you are analyzing a direct text, sharing specific parts of the text will allow you to get a more accurate and goal-oriented response. This helps me to indicate which expressions are used when I encode a specific statement or text.

For example, the following statement from the participant:

"The drama workshop changed my thoughts about the department. Previously, the lessons seemed boring and meaningless, but thanks to the drama workshop, I became more interested in the department's lessons and realized how valuable these lessons actually are."

This statement would be sufficient to determine the code of "Shift in Perspective Towards the Department".

If you have a direct participant text that I can use, I can give a more specific answer based on that. **Researcher:** Could you answer according to the text in the quotation marks below?.... (Interview Text)"

#### Figure 8 Theme 4

At this stage, we found that CG made 3 different mistakes when citing directly. The first of these errors; CG has used two direct quotes for the code "Increasing In-Class Participation" under Theme 2 (see. Figure 6 painted area). The first of the direct quotes used belongs to the participant, but the second statement used is the repeat sentence that the researcher said to confirm the participant. It seems that CG cannot distinguish between the researcher and the participant here.

We saw the second error when we commanded CG in a row and came to Theme 4. It cited three different direct quotes under the code CG "Change in Perception of University". However, we noticed that the second of these quotes was not sufficiently detailed (see. Figure 8 painted area).

When we came to the code "Increasing Participation in University Events" under Theme 4, what statements did you get from the participant's code "Increased Participation in University Activities"?" we commanded, but realized that CG used excerpts that were not included in the statements in the text. We repeated the command on using different quotes. Upon repetition of the command, CG corrected the error by directly including the participant's own statements. When commanded for the code "Shift in Perspective Towards the Department" under Theme 4 (12th command after the text, we saw that it did not cite the participant directly. The dialogue in Figure 9 then developed:

Researcher: These statements are not the participant's own words, could you quote directly from the participant's own words? ChatGPT: I apologize, as far as I understand, we are not based on a text directly coming from a participant. Therefore, I cannot provide the direct statements of a participant. The above responses, while making a general inference about how drama workshops might change the participant's view of the department, do not use a participant's direct statements. If you are analyzing a direct text, sharing specific parts of the text will allow you to get a more accurate and goal-oriented response. This helps me to indicate which expressions are used when I encode a specific statement or text. For example, the following statement from the participant: "The drama workshop changed my thoughts about the department. Previously, the lessons seemed boring and meaningless, but thanks to the drama workshop, I became more interested in the department's lessons and realized how valuable these lessons actually are." This statement would be sufficient to determine the code of "Shift in Perspective Towards the Department". If you have a direct participant text that I can use, I can give a more specific answer based on that. Researcher: Could you answer according to the text in the quotation marks below?.... (Interview Text)"

#### **Figure 9 Indicating the Error**

After the dialogue above, we put the entire error using the participant's own statements while interview text back to CG. It then corrected his obtaining the CG code.

ChatGPT: In this qualitative research interview, the analysis process will occur in the form of coding the data, categorizing it, and then forming themes First, we can create the following codes by generally analyzing the text: Codes 1. First impressions and shyness 2. Socializing and friendship Expressing oneself and community
 Social and academic adaptation mmunication 5. Self-confidence Club and community participation
 Event organization and participation 8. Interpersonal relationships
 9. Academic activities and com ication with instru Perspective and awareness about the university
 Perspective on and interaction with the department
 Next, we can categorize these codes into meaningful and broad categories: Categories: 1. Personal Development and Self-confidence: Codes 1, 3, and 5 could be included in this category. I. Personal Development and Self-confidence: Codes 2, 4, 7, and 8 could be included in this category. Personal Development and self-confidence: Codes 1, 3, and 5 could be included in this categor.
 Social Adaptation and Communication: Codes 2, 4, 7, and 8 could be included in this category.
 Academic Adaptation: Codes 4, 9, and 11 could be included in this category.
 University Perception: Codes 6, 10 could be included in this category. Finally, we can generalize the categories to create the following main them Themes:

 Personal and Social Development: The personal development and self-confidence category, along with the social adaptation and communication category, could be grouped Adaptation to the University: The academic adaptation and university perception categories could be grouped under this theme.
 Adaptation to the University: The academic adaptation and university perception categories could be grouped under this theme.
 These themes and categories can provide a framework for better understanding the impact of creative drama on easing the university adaptation process. Each category and theme can later be analyzed in detail and used to determine results.

#### Figure 10 Results of 2nd Analysis

#### **Repeating the Analysis**

CG, we opened a new chat and analyzed the same After completing the first step in analysis with text again. CG achieved different results when it analyzed the same text. The difference between the results was often due to the way it was expressed. In the second analysis, the code category and themes in the result are included in Figure 10;

We asked CG as in the first analysis of what expressions the generated codes obtained from the participant. We asked for a direct quote separately for each code. For this, we used the command "From which expressions did you get the code "..." under Category ...?". We checked CG's samples over the text as in the first analysis and saw that CG cited directly. At this stage, we encountered errors similar to those in the first analysis. When we come to the direct citation command of the last code (12th command after the first one) CG gave the statements that are not in the transcript as examples. After the error was stated, we sent the transcript again. And then CG corrected its mistake by citing the text directly. While CG created 4 different themes and 12 codes in its first analysis, in the second analysis it used the word "category" instead of "theme" and created 11 codes under categories. In its second analysis, the results were different in the form of the presentation. However, it clearly stated the themes to which the categories and categories to which the codes belong. Analysis of CG's themes in the first analysis and their codes under it, and analysis of the categories and codes in the second analysis are included in Figure 11;

Analysis 1	Analysis 2	Analysis 1	Analysis 2	Analysis 1	Analysis 2	Analysis 1	Analysis 2
Creative Drama Workshops and Social Adaptation	Social Adaptation and Communication	Creative Drama Workshops and Academic Adaptation	Academic Adaptation	Creative Drama Workshops and Personal Development	Personal Development and Self-confidence	Creative Drama Workshops and Perception of University	University Perception
Increase in Social Interactions	Socializing and friendship	Increase in Classroom Participation	Social and academic adaptation	Increase in Personal Confidence	First impressions and shyness	Change in Perception of University	Club and community participation
Participation in Clubs and Groups	Event organization and participation	Development of Academic Confidence	Academic activities and communication with instructors	Reduction in Shyness	Expressing oneself and communication	Increased Participation in University Activities	Perspective and awareness about the university
Development of Communication Skills	Interpersonal relationships	Improvement in Communication with Instructors	Perspective on and interaction with the department	Development of Self-expression Ability	Self-confidence	Shift in Perspective Towards the	
	Social and academic adaptation				Department		

Figure 11 Comparing 1st and 2nd Analysis

When Figure 11 is examined, it is seen that the naming "theme" in the first analysis and the naming "category" in the second analysis are similar. According to Figure 11, the code "Shift in Perspective Towards the Department" in Analysis 1 is included in both analyzes. However, there is a difference between the category in which analysis 1 is included and the category in which analysis 2 is included.

Although the codes created in the first and second analysis differ as names, they show a very close importance in terms of the meaning they carry and the direct quote they are obtained. Figure 12 exemplifies this situation;

Analysis 1	Analysis 2
Code: Increased Personal Confidence	Code: Self-confidence
From the participant's statements, I derived the code "Increase in Personal Confidence" from the following excerpts: "The confidence and sense of assurance that drama provided, well, it kind of motivated me to join them. I was hesitant to join clubs and groups because of my shyness, but drama eliminated that hesitation." - This statement highlights how drama workshops increased the participant's personal confidence. "We have communication with the instructors because I used to be a bit shy in class. Through the drama process, it greatly helped me gain confidence." - This statement also demonstrates how drama workshops enhanced the participant's personal confidence and addressed their shyness. These two statements indicate an increase in the participant's personal confidence due to the effects of drama workshops. Therefore, the code "Increase in Personal Confidence" is established.	The participant's statements, "I'm a bit shy actually, I'm not a very social person," and "Drama has had a significant impact on me socially," provide a foundation for creating the code "Self-Confidence." These statements indicate that drama workshops enhance the participant's self-confidence. These statement, "The confidence and sense of assurance that drama provided, well, it kind of motivated me to join them. I was hesitant to join clubs and groups because of my shyness, but drama eliminated that hesitation." shows that drama workshops boost the participant's self-confidence and encourage them to engage in more social activities. Lastly, the statement, "We have communication with the instructors We chat with our regular class teachers as well, not extensively, but I can speak in class, answer questions," demonstrates that drama workshops enhance the participant's self-confidence and enable them to be more active in the classroom. When these statements are combined, they help us understand how drama workshops develop the participant's self-confidence.

Figure 12 Different Naming of the Code
--

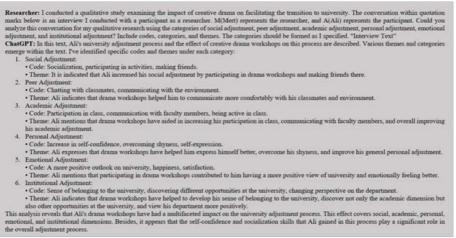
# **Specifying Categories**

At this stage, we re-analyzed the text by opening a "new chat". This time, before uploading the interview text, we gave CG a summative paragraph of theoretical information about the university adaptation process. We shared short information with CG in this theoretical information text without going into the details of the theory. The theoretical information provided is included in Figure 13;

Researcher: Young people entering the university make changes to their lifestyle and have to adapt to new responsibilities in their social and academic lives (Pitmann & Richmond, 2008). According to Tinto (1975), this process requires academic and social integration. Social integration occurs primarily through informal peer groups, semi-formal extracurricular activities, interacting with faculty and administrative staff. Successful transition in these areas results in different degrees of social communication, friendship support, faculty support and institutional belonging. Academic integration refers to structural resources in universities and student-faculty interaction, student grades and intellectual development. Tinto (1975) stated that the correct communication with the staff and peers within the institution contributes to university adaptation. He also laid emphasis on ensuring institutional adjustment for individuals to succeed in the process of adapting to the university. If not, students who have a low institutional adjustment may drop out of the university (Tinto, 1975). Tinto (1975) emphasized that alongside academic integration, personal features and intellectual development are an integral part of this process. He stated that the emotional states, values and judgments of individuals play a significant role in this process. Baker and Siryk (1986) also stated that individuals are a process of personal/emotional adaptation, expressed personal/emotional adjustment as how the individual feels psychologically and physically, and referred to the importance in the harmonization process. Above is some information for you to understand the university adjustment process. If I ask you to do an analysis based on this information in the next message, would you help me?

#### **Figure 13 Theoretical Information**

After this command, CG used statements that summarize the paragraph we shared. Then we gave in Figure 14; CG the command to analyze by specific categories.



#### Figure 14 Analysis by Categories

In Figure 14, when we asked CG to specify the categories and create code, it created 17 meaningful codes that fit 6 categories. These codes differed in terms of naming and presentation according to the codes in the first and second analyzes. When we examined the content of the codes, we found that it was similar to other analyzes. At this stage, we have commanded to receive direct quotes from the participant's statement regarding each code in turn. CG gave the statements that were not in the text of

the speech as an example when it came to the last code (13th command after the text) as in the first and second analysis. After we stated the error, we sent the text of the speech again and provided a direct quote from CG's text and the error improved. The sections that CG cited directly on the codes showed parallelism with the sections quoted in the first and second analysis.

# Making a Consecutive Analysis

After analyzing the first participant with CG, we proceeded to the analysis of the other participant in the research. For this, we first used the continuation of Chat, where we made the first analysis. We commanded CG 18 consecutively for the first participant's transaction. In the 19th command in this Chat, we wrote the command in Figure 15.

**Researcher:** Thank you. Now, I want you to analyze the interview of a different participant in the research examining the impact of creative drama on facilitating the transition to university. The conversation within quotation marks below is an interview conducted by the researcher with the participant. M(Mert) represents the researcher, and B(Beyza) represents the participant. Could you break down this text into codes, categories, and themes for qualitative research? (Interview Text)

# Figure 15 Analysis of the Second Participant

CG created different code, category and theme after analyzing the data of the second participant. We performed the same actions for the second participant, such as asking for a direct quote to the first participant and analyzing by categories. Then we commanded CG to specify the common codes of the first and second participants. CG stated at this stage that it did not have information about first and second participants' code.

**Researcher:** Could you specify the codes that are common to both Ali and Beyza? **ChatGPT:** Unfortunately, I don't see the codes related to the statements of individuals named Ali and Beyza in the texts you provided me. Therefore, I can't identify the codes common to Ali and Beyza. If you share the statements of Ali and Beyza, I'd be happy to assist you.

# Figure 16 Error Message

After the error message, we opened "New Chat" with categories. Then we gave CG the command in and uploaded the analysis results of both participants Figure 17;

Researcher:"Below is the analysis pertaining to Beyza; (Analysis pertaining to Beyza has been added here earlier) Here is also the analysis related to Ali; (Analysis pertaining to Ali has been added here earlier) Could you present the codes, categories, and themes of Ali and Beyza in a table and explain their similarities and differences?"

# Figure 17 Binary Comparison Command

Following this command, CG gave the following results in tabular form;

When Figure 18 is examined, it is seen that CG can identify common codes, categories and themes in previous analyzes, reveal similarities and differences between participants, and make explanations about similarities and differences.

CG's ease of use distinguishes it from other CAQDAS. It is necessary to learn how to use many of the CAQDAS before using it effectively. However, explanation of the action to be made via "New Chat" is sufficient to use CG. It has been found that it is sufficient to tell CG the purpose of the research and to indicate the action to be taken for data analysis. In addition to ease of use, formal deficiencies in interview texts uploaded to CG for data analysis (inability to copy down people's speeches, spelling errors, etc.) does not interfere with data analysis. Articles, tables and outputs from CG can be copied completely from CG's intermediate face without being corrupted.

	Beyza	Ali	Similarities	Differences	
Social Adjustment	eta mali a la construcción de		Both of them state that participating in social activities increases social adjustment	Beyza's experience talks about fear a contradiction at the beginning, while Ali's experience is more positive and focused on participating in activities	
Peer Adjustment	Gaining more friends, not being shy or hesitant, sharing problems with strangers	Chatting with classmates, communicating with the environment	Both of them indicate that drama workshops have a positive impact on peer adjustment	While Beyza emphasizes gaining friends and overcoming shyness, Ali's experience focuses on chatting with classmates and communicating with t environment	
Adjustment participating in communities, discussing ideas with teachers active in		Participation in class, communication with faculty members, being active in class	Both of them state that drama workshops enhance academic adjustment	While Beyza's experience generally involves participating in academic communities and discussing with teachers, Ali's experience focuses on i class participation and communication with faculty members	
Personal Adjustment	Expressing oneself, overcoming shyness and hesitation, openly presenting ideas	Increase in self- confidence, overcoming shyness, self-expression	Both of them state that drama workshops improve personal adjustment, especially in terms of expressing oneself and overcoming shyness	While Beyza focuses more on presenting ideas, Ali's experience concentrates on an increase in self- confidence	
becoming more happiness, satisfactio		A more positive outlook on university, happiness, satisfaction	Both of them indicate that drama workshops positively affect emotional adjustment	While Beyza's experience emphasizes overcoming the initial contradiction as fear, Ali's experience is more focused on happiness and satisfaction	
Institutional Adjustment Perspective on the university, perspective on the department, increased participation in university life perspective university perspective university participation in university university perspective university university perspective university university university university perspective university un		Sense of belonging to the university, discovering different opportunities at the university, changing perspective on the department	Both of them state that drama workshops enhance institutional adjustment	While Beyza's experience involves an overall perspective on the university and the department and participation, Ali's experience focuses on the sense i belonging to the university and discovering the different opportunities at the university.	

**Figure 18 Similarities and Differences** 

# Discussion

Computer-assisted qualitative data analysis software (CAQDAS or QDAS) has been available since 1980. Nowadays, using software for qualitative research is becoming increasingly common for researchers (Cypress, 2019). John and Johnson (2000) discussed the benefits of using the software in qualitative data analysis, such as getting rid of manual editing tasks, saving time, using large qualitative data, and increasing the validity and control of qualitative research. The widely used qualitative research software has features such as encoding, consolidation, data search and query, data visualization, transcription, statistical analysis by the researcher (Clarke et al., 2021). Software such as ATLAS.ti, NVivo and MAXQDA have added automatic analysis tools to their current versions. Using word frequencies, word clouds, and autoencoding capabilities, researchers can get a bird'seye view of data content before proceeding to deeper analysis (Paulus, 2022). However, these programs are not sufficient for the search for meaning, which is the essence of qualitative research. Though existing software has features such as finding the words in the qualitative data set, the relationships between the words, the frequency of the words used, it is seen that they do not have the analysis capabilities to

interpret what the data contains and to interpret the data. (<u>Kalpokas & Radivojevic</u>, 2022; <u>Vignato, et al.</u> 2022).

CG's ease of use distinguishes it from other CAQDAS. It is necessary to learn how to use many of the CAQDAS before using it effectively. However, explanation of the action to be made via "New Chat" is sufficient to use CG. It has been found that it is sufficient to tell CG the purpose of the research and to indicate the action to be taken for data analysis. In addition to ease of use, formal deficiencies in interview texts uploaded to CG for data analysis (inability to copy down people's speeches, spelling errors, etc.) does not interfere with data analysis. Articles, tables and outputs from CG can be copied completely from CG's intermediate face without being corrupted.

CG's ability to operate in different languages allows researchers to use their datasets in their native language when analyzing. In our research, it was seen that CG has successfully completed the analysis process and did not make semantic disorders, although all of the transactions with CG were done in Turkish. In this context, it can be said that CG can analyze in a different language other than English.

As seen in our research, CG can analyze the data (may vary depending on text length) uploaded to it

in seconds. CG analyzed between 1-20 seconds for the data we used in our research. It is possible to say that this analysis period is much faster than a normal researcher's qualitative data analysis time. However, CG can interrupt the response in cases where the responses are long. In cases where the response is interrupted, the response with the option "Continue generating" can be sustained from where it left off. In cases where the answer is not liked, the response can be repeated with the option "Regenerate response". With this feature, CG offers the researcher different response options. It can also show disruptions in understanding CG commands and responding correctly. However, in such cases, repeating and explaining commands clearly can solve the problem.

When we asked CG to create code, category and theme without giving any information, it was determined that CG could create meaningful code, category and themes. The existing qualitative research analysis software used can create codes from within the text given to it, but cannot explain why it creates the codes (Kalpokas & Radivojevic, 2022). CG also explained why it created the codes it created and stated why it used these codes in which categories and themes. It was also found that when we provided theoretical information to CG and specified the categories, it collected the appropriate codes under the appropriate categories and themes. In this context, it can be said that researchers can analyze using CG according to the categories and themes they have previously determined.

In different analyzes, it has been observed that CG can use the concepts of category and theme interchangeably, but codes are always used under categories and themes. It is possible to say that the created codes are created in accordance with the relevant categories and themes. CG has been shown to give different results in repeated analyzes of the same data. These differences have generally not been differences in the essence of codes, categories and themes. In the second analyzes made on the same data, it was seen that there was a differentiation in the naming and the number of codes created. Although the naming is different, it has been observed that direct quotes for two similar codes are selected from the same statements. In this context, it can be concluded that CG names the code of the same

paragraph differently, but considers the meaning to be the same. It has also been determined that some codes can be under different categories. These differences can be compared to the differences that researchers who analyze the same data set at different times can make.

When direct quotations for codes created from CG were requested, it was found that he was able to present the quotations to the researcher correctly. Although CG was successful in getting direct quotes, it has not always been functioning correctly. When commanded in a row, CG was found to be able to behave incorrectly. In such cases, the error must be specified by the researcher to CG, and the process must be repeated if necessary. Although there is little margin of error in this regard, researchers are recommended to check direct quotes. In addition, it appears that CG has the ability to improve its analysis by asking for further explanation when it has problems when analyzing.

CG was found to be able to make mistakes when long-term actions were taken on the same Chat page. Following the analysis of the first participant, switching to the analysis of the second participant resulted in the inefficient use of CG. In this context, it is more appropriate to analyze the participants in different "Chat" tabs during the data analysis process.

Compared to the analyzes of the two different participants, it has been determined that CG can analyze similarities and differences correctly and explain the analysis it made in a meaningful way. In addition, when the analysis of more than one participant was wanted to be compared, it was seen that the type of analysis in which the categories were specified gave more functional results.

# Conclusion

The use of CAQDAS for qualitative research analysis has advantages and disadvantages in various aspects. While researchers using CAQDAS can take advantage of features such as editing, transferring, and transcripting data, they experience inadequacy in matters such as analyzing data, interpreting it, understanding the essence of the data. CG and AIbased software likely to be done in the future offer new advantages, disadvantages and opportunities for qualitative research analysis. In this research, analyzes were conducted with limited written data to reveal the potential of CG, an AI-based software. As a result of the processes, it was observed that CG had features such as revealing the meaning in qualitative data analysis, understanding and interpreting the essence of the data. CG was able to analyze the interview text, extract code, category and themes, and include direct quotes from the text, in accordance with the purpose of the research. It also explained the reasons for its actions. It has analyzed the data from different participants and has been able to identify similarities and differences among these analyses. As a result, it can be stated that CG can be used as an auxiliary software for researchers in qualitative research data analysis, meaning emergence, code, category and theme creation processes.

# Suggestions

The datasets used in this research are limited to the data of two different participants. In order for CG to be used in qualitative research analysis, it may be suggested to be tested with various research in different patterns and content. Since the native language of the researchers in this article is Turkish, all transactions were made in Turkish. Researchers with different mother languages may be recommended to use CG in their native language. Due to the disruptions seen as a result of long-term command writing, it may be suggested that no consecutive analysis is performed on the same "Chat" and analysis can be made for each participant on seperate chats . Since the commands given to CG are limited in this research, it can be explored what the commands that should be written in different studies can be. It may be suggested to use the ChatGPT-4 version due to the technical superiority of different researchers. Various studies can be conducted on the validity and security of CG's analyzes. Since the use of CG in data analysis in qualitative research can reveal some ethical and philosophical problems, it may be suggested to discuss the ethical and philosophical aspects of this issue.

# References

- Bredal, A., Stefansen, K., & Bjørnholt, M. (2022). Why do people participate in research interviews? Participant orientations and ethical contracts in interviews with victims of interpersonal violence. *Qualitative Research*.
- ChatGPT. (2022). What is ChatGPT?. OpenAI.
- Clarke, S. O., Coates, W. C., & Jordan, C. (2021). A practical guide for conducting qualitative research in medical education: Part 3-Using software for qualitative analysis. *AEM Education and Training*, 5(4), 1-8.
- Creswell, J. W. (2013). *Qualitative Research Methods: Qualitative Research and Research Design According to Five.* (M. Bütün, & S. B. Demir, Trans.). Political Bookstore.
- Cypress, B. S. (2019). Data analysis software in qualitative research: Preconceptions, expectations, and adoption. *Dimensions of Critical Care Nursing*, 38(4), 213-220.
- Çelebi, M., & Orman, F. (2021). Data collection tools. In M. Çelebi (ed.), *Qualitative Research Methods* (pp. 142-166). Pegem Academic Publishing.
- Haque, M. U., Dharmadasa, I., Sworna, Z. T., Rajapakse, R. N., & Ahmad, H. (2022). I think this is the most disruptive technology: Exploring Sentiments of ChatGPT Early Adopters using Twitter Data. Cornell University.
- Jenskin, N., Monaghan, K., & Smith, M. (2023). Did they really say that? An agential realist approach to using computer assisted transcription software in qualitative data analysis. *International Journal of Social Research Methodology*, 26(1), 97-109.
- John, W.S., & Johnson, P. (2000). The pros and cons of data analysis software for qualitative research. *Journal of Nursing Scholarship*, 32(4): 393-397.
- Kalpokas, N., & Radivojevic, I. (2022). Bridging the gap between methodology and qualitative data analysis doftware: A practical guide for educators and qualitative researchers. *Sociological Research Online*, 27(2), 313-341.

- Kirmani, A. R. (2023). Artificial Intelligence-Enabled Science Poetry. *American Chemical Society*, 8(1), 574-576.
- Lock, S. (2022). What is AI Chatbot phenomenon ChatGPT and could it replace humans?. *The guardian*.
- Mercan Küçükakın, P., Göloğlu Demir, C. & Gökmenoğlu, T. (2022). Turkish teachers' response to the COVID-19 crisis: challenges and opportunities. *Education and Science*, 47(212), 83-101.
- Merriam, S. B. (2018). *Qualitative Research: A Guide to Pattern and Practice* (S. Turhan, Trans.). Ankara: Nobel Academic Publishing.
- Naidu, K. (2022). ChatGPT is a new AI chatbot that can find errors in your code and write you a story. *Crazy Engineer*.
- Niedbalski, J. & Ślęzak, I. (2022). Encounters with CAQDAS: Advice for beginner users of computer software for qualitative research. *The Qualitative Report*, 27(4), 1114-1132.

OpenAI. (2022). Introducing ChatGPT. OpenAI.

OpenAI. (2023a). GPT-4. OpenAI.

- OpenAI. (2023b). ChatGPT plugins. OpenAI.
- Özcan, B. (2022). Could ChatGPT, the best artificial intelligence chatbot, be the new Google?. *Baris Ozcan*.
- Paulus, T. M. (2022). Using qualitative data analysis software to support digital research workflows. *Human Resource Development Review*, 22(1).
- Sier, J. (2022). ChatGPT takes the internet by storm, bad poetry and all. *Financial Review*.
- Stokel-Walker, C. (2022). AI Bot ChatGPT writes smart essays - Should professors worry?. *Nature*.
- Vignato, J., Inman, M., Patsais, M., & Conley, V. (2022.) Computer-assisted qualitative data analysis software, phenomenology, and colaizzi's method. *Western Journal of Nursing Research*, 44(12), 1117-1123.

# **Author Details**

Mert Sen, Inonu University, Turkey, Email ID: mert.sen@inonu.edu.tr

Şevval Nur Şen, Inonu University, Turkey, Email ID: sevval.ozkes@inonu.edu.tr

Tuğrul Gökmen Şahin, Inonu University, Turkey, Email ID: tugrul.sahin@inonu.edu.tr