Navigating the Currents of Change: An Ethnographic Study of IT Investment Reporting in Organisations

Danny Ray Justice (02.03.00) University of Southern Denmark Department of Humanities 31st May, 2023

Abstract

Abstract abstract

Keywords: Keyword

Introduction

Introduction text

Research Question

The IT department at Vejle municipality has established an IT investment process that outlines a set of procedures for procuring new IT systems, software, and equipment. Despite these guidelines, employees do not always follow this process, leading to potential inefficiencies and discrepancies in IT investments. This study aims to explore the reasons why employees do not comply with the IT investment process and suggest strategies to improve compliance.

Additionally, the study seeks to evaluate the utility of OpenAI's family of large language models, Generative Pre-trained Transformer (also known as GPT), as a complementary tool to traditional research methods in understanding employees' attitudes and behaviours towards the IT investment process. By comparing insights generated by these models with those obtained from traditional research methods, the study aims to determine the potential and limitations of using large language models as a qualitative research tool in ethnographic studies.

Literature Review

Ethnographic Methods (Crabtree et al. 2012)

Interview

Crabtree, Rouncefield, and Tolmie (2012) advise that researchers treat interviews with caution. They suggest that interviews should be conducted in the actual flow of work as it unfolds and as the situation permits. They warn that what people say they do and what they actually do are not the same. It is not that people are lying, but that the accounts they offer in an interview often gloss over their work. The best way to conduct an interview is to be concerned with the just what and just how of the work, and not be driven by a pre-formulated schedule of questions removed from the actual doing of the work.

Field notes

According to Crabtree, Rouncefield, and Tolmie (2012), making field notes is an essential part of fieldwork. It allows the researcher to document the things they see and hear and jot down their thoughts on the setting and its work. Field notes provide a record of what the researcher observes, hears, and is told. Keeping a good set of field notes helps researchers keep track of what they are being told and organise their thoughts. It is an active process that makes the researcher attend to the work as it occurs, helping them develop vulgar competence in it.

Crabtree, Rouncefield, and Tolmie (2012) suggest that the notebook need not be a loose collection of disjointed comments. Researchers may use their notebook to structure their thoughts and develop a coherent account of the work of a setting. They recommend researchers make diagrams of the ecology of work to frame their inquiries into the work of a setting and represent it to others. Draw plans of the environment, indicate the people who inhabit it, their roles or responsibilities, and the artifacts that they use in doing the work. This helps researchers develop a detailed understanding of the setting's work and the methods members use to organize it as a real-world, real-time social accomplishment.

Formal Organisation of Work and Flow of Work

Crabtree, Rouncefield, and Tolmie (2012) suggest that researchers describe how the setting's work is 'formally organised' across a division of labour and how it is 'formally organised' at an individual level. This includes plans, procedures, processes, and routines that the setting's members invoke to account for the organisation of their work. Researchers should also focus on the flow of work, which starts somewhere, with someone doing something and proceeds to some end. They should focus on how the work moves across individuals, how it flows from one activity to another and one person to another.

Discrete Sequences of Interactional Work and Cooperation and Collaboration

Crabtree, Rouncefield, and Tolmie (2012) recommend that researchers flesh out their description of the flow of work by focusing on the discrete sequences of interactional work that are involved in the accomplishment of particular activities. Researchers should describe what is being done, who is doing it, and how the work is accomplished. Researchers should also focus on the cooperation and collaboration that takes place between people in the accomplishment of discrete sequences of interactional work. They should describe who is talking to whom, what they are talking about, what they do together, the transactions that take place between them, the hand-over of tasks, and what others do in response.

Buxton (2007)

In "Sketching User Experiences," Buxton (2007) emphasises the importance of sketching for generating and exploring ideas in the design process. Sketching allows designers to quickly generate and iterate on ideas, explore different design possibilities, and ultimately arrive at a solution that best meets user needs.

Buxton provides numerous techniques and tools for using sketching to support idea generation, including sketching with pen and paper, creating paper prototypes, and storyboarding. He also emphasizes the importance of using sketching to support user-centered design, and provides practical advice on involving users in the design process and using sketching to elicit and communicate user requirements.

One of the strengths of the book is the practical advice it provides on how to use sketching for idea generation. Buxton emphasises the importance of sketching as a means of exploring and refining ideas, and provides numerous examples of how sketching can be used to generate and communicate design concepts. He also provides advice on how to use sketching to support different design activities, such as brainstorming, ideation, and prototyping.

Another strength of the book is the emphasis on using sketching as a collaborative tool. Buxton highlights the value of sketching as a means of communicating and refining ideas with other stakeholders, such as team members, clients, and users. He provides advice on how to use sketches to facilitate communication and collaboration, and how to document and organise sketches for future reference (Buxton, 2007).

Affinity Diagrams

Affinity diagrams, also known as the KJ method or affinity charting, were first developed by Japanese anthropologist Jiro Kawakita (as cited in Scupin (1997)). Affinity diagrams are used to synthesise and categorise large amounts of

qualitative data, such as observations, interviews, and field notes, into meaningful and easily understandable themes and patterns Hanington and Martin (2019).

The affinity diagramming process typically starts with the raw data being transformed into discrete statements or observations. These statements are then grouped based on their similarities and relationships (Hanington & Martin, 2019). The groups are subsequently labeled with descriptive headings, which capture the essence of their content. This iterative process allows for the identification of patterns, themes, and relationships among the collected data, thus providing insights and guidance for further analysis and design Holtzblatt and Beyer (2016).

In the context of ethnographic UX studies, affinity diagrams serve as a valuable tool for making sense of the complex and often messy data that emerges from immersive fieldwork Holtzblatt and Beyer (2016). By organising and categorising data in a structured manner, researchers can identify user needs, behaviours, and pain points, which can inform design decisions and enhance the overall user experience Hanington and Martin (2019).

Furthermore, affinity diagrams facilitate collaboration and interdisciplinary communication among research team members Holtzblatt and Beyer (2016). By engaging in the process of grouping and labeling data, researchers from different backgrounds and expertise can contribute to a shared understanding of the user experience, leading to more innovative and effective solutions.

Prompt Engineering

Prompt engineering is a critical process for achieving high-quality results with LLMs (Large Language Models), which have proven to be highly beneficial in various tasks (Guo et al. (2023); Mahowald et al. (2023); Wei et al. (2022); Wei (n.d.)). According to DAIR.AI (2023), "prompt engineering is a relatively new discipline for developing and optimising prompts to efficiently use language models for a wide variety of applications and research topics. Prompt engineering skills help to better understand the capabilities and limitations of LLMs."

The models utilised in the present study (OpenAI's GPT-3.5 and GPT-4) do not currently support fine-tuning for specific tasks. In turn, prompt engineering becomes even more essential in obtaining high-quality results. The Prompt Engineering Guide from DAIR.AI (2023) proved to be an excellent resource for prompting the models as it offers a repository of state-of-the-art methods and techniques.

The process of prompt engineering includes selecting appropriate text prompts, tuning the model's hyperparameters, and tailoring the prompt design to the model's capabilities. The primary objective is to create prompts that provide sufficient context for the LLM to generate high-quality results while avoiding those that may lead to incorrect or irrelevant outputs (DAIR.AI (2023); Shieh (n.d.)).

Methodology

Elicit Research Assistant

In addition to traditional literature search methods, I used Elicit, a digital research assistant employing large language models. The steps followed while using Elicit in this study are outlined below:

1. Posing a Research Question: A research question was entered into Elicit (e.g., "What are the ethical implications of using LLMs in qualitative research?").
2. Semantic Similarity Searches: Elicit searched for relevant papers based on semantic relationships rather than just keyword matching, broadening the scope of literature found.
3. Custom Abstract Summaries: Elicit generated tailored summaries of the sources' abstracts, presenting only information relevant to the research question, thus providing an initial understanding and evaluation of the research.
4. Citation Graph Analysis: Elicit was used to explore the citation graph of selected papers to identify additional relevant literature based on citation relationships.

By following these steps and incorporating Elicit's features into the literature review process, a comprehensive and organised analysis of the available literature was achieved. Elicit served as a valuable resource in the development of the present study, contributing significantly to the literature review component.

Ethnography

Semi-structured Interviews

To build empathy with the users, I conducted semi-structured interviews in an informal manner. According to Crabtree, Rouncefield, and Tolmie (2012), interviews should ideally be conducted in the flow of work, but since new IT investments were not a regular part of most employees' daily work outside of the IT department, it was not feasible to talk with employees in the midst of making new investments. Thus, the interviews were done in a "decontextualised" fashion.

In total, 12 employees were interviewed, selected based on a list of contacts provided by the author's contact person in Vejle's IT department. The list included decision-makers from all six administrations of the municipality. Each employee on the list was assigned assigned a pseudonym and identification number by which they were identified in my notes, never using their true names.

All but one of the interviews were conducted face-to-face, the remaining interview being held by video call in order to accommodate the participant, and notes were taken by hand in a notebook to maintain a relaxed atmosphere, as opposed to using audiovisual recordings. The questions were not pre-formulated but were based loosely on an interview guide, allowing for a more natural and authentic conversation. During the interviews, brief notes were taken, and more detailed notes (Appendix X) were written as soon as possible afterwards to ensure accuracy and facilitate data analysis.

Field Notes

In conducting my research, I recognised the significance of field notes as an essential tool for documenting and organising the vast amount of information that I would encounter during my fieldwork. As Crabtree, Rouncefield, and Tolmie (2012) emphasise, field notes play a critical role in capturing the observations, thoughts, and reflections of the researcher on the setting under study. They offer a means to document the unfolding of work and provide a detailed record of the interactions, activities, and processes that shape the context of the research.

To ensure that I was able to keep an accurate and detailed account of my observations, I maintained a dedicated notebook where I recorded my thoughts, sketches, and other details of my fieldwork in Vejle. This notebook was, in particular, used to document the interviews conducted, avoiding the presence of laptop or tablet screens and recording devices, which I feared would disturb the natural flow of conversation.

In addition to traditional field notes, I also created a folder on a cloud-synced notes app that allowed me to make digital notes on my computer, smartphone, and tablet. This proved to be a valuable resource when I didn't have my physical notebook with me or when I was in the mood to type rather than write, allowing me to capture my thoughts and insights in a way that was convenient and flexible for my needs.

My field notes were instrumental in facilitating my understanding of the work in Vejle by highlighting areas of cooperation and collaboration between individuals that were not necessarily captured by the formal organisation of work. They

provided a valuable means to document the handover of tasks and the transactions that took place between individuals, which helped to reveal the dynamic nature of the work.

In addition to being a useful record of my observations, my field notes served also as an active process of reflection, helping me create a mental model of the work being done in Vejle. By continually reflecting on what I was seeing and hearing, I was able to develop a detailed understanding of the setting's work and to refine my description of it over time. Furthermore, my field notes proved to be an invaluable resource for continually reworking my understanding of the work and for confirming or correcting my observations with those who were actually doing the work.

On the whole, my field notes served as a critical resource for documenting and organising the rich and complex data that I collected during my fieldwork. By capturing the unfolding of work in Vejle and facilitating my understanding of it, my field notes provided a solid foundation for the analysis and interpretation of my research findings.

Large Language Models

In light of the limited literature on the application of LLMs, such as GPT-3.5 (from the ChatGPT platform) and GPT-4, in qualitative research, this study adopted an exploratory approach to investigate the potential of these tools. Consequently, I utilised these LLMs as supplementary instruments within this study. To maintain a clear and coherent narrative, I will discuss the reasoning and approach behind incorporating these tools in the upcoming sections. Additionally, for complete transparency, I have included the full transcripts of my interactions with the models in Appendix X.

Chat Functionality

The conversational ability of LLMs provided the author with a distinctive opportunity to gather information beyond their personal knowledge and the resources available in the context of their research. With access to an extensive corpus of text, LLMs were able to generate responses to the author's queries that often offered fresh perspectives on the research topic. This prompted the author to consider new questions, viewpoints, and avenues of inquiry that may have otherwise been overlooked.

Furthermore, the LLMs not only served as a readily available resource but also stimulated the author's critical thinking and intellectual curiosity. By acting as a sounding board, the LLMs provided the author with a space to explore new ideas and perspectives, prompting them to think more deeply about their research.

Text Generation

This report acknowledges the use of OpenAI's language models (LLMs) to assist in the writing process. Although LLMs were used to generate specific passages in this report, it is crucial to note that these passages underwent significant editing, revision, and refinement to fit within the author's writing process. This approach entailed multiple rounds of editing, revising, and proofreading, ensuring that any text generated by LLMs was subjected to rigorous scrutiny and manual refinement where necessary. This approach allowed the author to leverage the power of LLMs while maintaining the integrity and authorship of the final product.

Text Summarisation

After completing the interviews, the notes for each interview were fed into GPT-4 one at a time to generate more coherent and refined prose. The generated texts were reviewed for completeness and factual accuracy to ensure the reliability and validity of the data. Examples of this process can be found in Appendix X.

Platforms Used

Throughout my research, I aimed to find the most effective way to interact with LLMs. I began with ChatGPT, the well-known platform offering access to GPT-3.5 and limited access to GPT-4. However, I encountered several limitations that constrained my research.

Seeking a better approach, I experimented with OpenAI's developer playground. This platform allowed me to harness GPT-4's 8,000-token capacity and actively guide the model's responses by editing outputs, adjusting hyperparameters, and modifying the system prompt to adopt specific roles, such as an ethnographer. Despite these advantages, the developer playground lacked a user-friendly method for saving and referencing chats.

My search led me to the GitHub project "Chat with GPT," Cogent Apps et al. (2023) which provided an efficient and intuitive interface for engaging with LLMs. This platform facilitated access to LLMs through an API key and featured a built-in search function, streamlining the navigation and retrieval of specific conversations.

Using the "Chat with GPT" platform, I effectively incorporated LLMs like GPT-4 into my research, allowing me to explore their strengths and weaknesses within the context of the present study. Although GPT-4 offers a more powerful 32,000-token variant, I did not have access to it during my research.

Ethical Considerations of Utilising LLMs

While LLMs offer a powerful tool for researchers, they raise ethical considerations about their use. The reliability and accuracy of responses generated by LLMs depend on the quality of data they have been trained on. As a result, these models may unintentionally perpetuate biases inherent in the data. In order to effectively integrate LLMs into our research, it becomes crucial for humanities researchers to scrutinise these tools and explore ways to mitigate any potential biases.

However, it is important to acknowledge that humans are also subject to their own biases based on the data on which they are "trained," such as memories and knowledge. As researchers in the humanities, we are well aware of the ways in which our own biases can impact our understanding of the world. We possess the expertise to untangle such complexities and apply this critical lens to the use of LLMs.

Through investigating the use of LLMs in our field, we can not only understand their potential limitations and biases, but also find ways to mitigate these challenges and maximise their potential. As such, it is crucial for humanities researchers to take the lead in the examination and critical analysis of LLMs.

Affinity Diagramming

The present study utilised affinity diagramming as a method of analysing the qualitative data collected from the semi-structured interviews. After conducting the interviews and taking detailed notes, the researcher followed a structured process to create two affinity diagrams: the first crafted with traditional research methods, the second being made with the assistance of GPT-4.

Manual Affinity Diagram Creation

Firstly, the data segmentation process was undertaken, which involved breaking down the interview notes into discrete statements or observations, each representing a single idea or insight expressed by the interviewees. Next, the sorting and grouping stage took place, whereby the statements were sorted into groups based on their similarities and relationships. This process was iterative, with statements being moved between groups as new connections and patterns emerged.

After the sorting and grouping process, the researcher assigned descriptive headings to each group, which represented the primary themes and patterns identified in the data. These headings captured the essence of the content within

each group. Finally, the affinity diagram was reviewed and refined, ensuring that the groupings and labels accurately represented the data, and making adjustments as needed.

Affinity Diagram Creation Assisted by GPT-4

In this section, I will explore the capabilities of GPT-4 in creating an affinity diagram in order to compare its performance to myself as a human ethnographer. Affinity diagrams are useful tools for organising and making sense of vast amounts of information; given GPT-4's ability to process vast swaths of text, it presents a unique opportunity to investigate the effectiveness of the language model in creating a affinity diagrams.

Prompt engineering plays a crucial role in achieving optimal performance of Language Models (LMs), including GPT-4. In this section, I will discuss my approach to prompt engineering and my efforts to achieve high-quality results. With GPT-4's enhanced capacity to comprehend and aid with meta-tasks, such as refining its own prompts, there is potential for streamlining the prompt engineering process without compromising its efficacy, thus further reducing workload.

Rather than spending a long time crafting the perfect prompt, I opted for an iterative approach to prompt engineering, attempting several methods of eliciting affinity diagrams from GPT-4.

The following sections make use of what OpenAI call chat completions. In chat completions, there are three types of messages: system, user, and assistant. Unless marked otherwise, all prompts mentioned in the following sections were injected into the conversation the system message which, according to documentation in OpenAI (n.d.), helps to guide the AI's behaviour. The interview data was fed into the model as a user message, and subsequent completions were assistant responses.

By exploring the capabilities of GPT-4 in creating an affinity diagram and comparing it to that of a human ethnographer, I seek to gain insights into the strengths and limitations of the language model in this task.

The Quick and Dirty Approach.

In my first attempt to get GPT-4 to produce an affinity diagram, I gave it the following simple prompt: "You are an ethnographer who evaluates user inputs and creates affinity diagrams based on the narratives they contain. Your report should summarize the most prominent discourses, including references to specific interviewees where appropriate."

GPT-4 provided an analysis of the interview data and generated its own affinity diagram, which identified themes and patterns in the data, followed by a summary of the most significant discourses. The output was technologically impressive, but some of the fundamental issues with large language models made it methodologically questionable.

While GPT-4 was clearly familiar with the concept of an affinity diagram, its results were somewhat superficial, seeming to be based on a mere first impression of the data. The raw output is as follows:

> Affinity Diagram Themes: 1. Understanding of the IT investment process 2. Communication and collaboration with the IT department 3. Challenges and frustrations in implementing new systems 4. Suggestions for improvement
>
> Summary of Prominent Discourses: 1. Many interviewees were not aware of the IT investment process or did not see the relevance to their work. For instance, Karina (Interview 165) did not think it was important for her role, while Anne (Interview 190) was not aware of the process before the interview.
>
> 2. Communication and collaboration with the IT department were highlighted as crucial factors in the successful implementation of new systems. Lise (Interview 161) emphasized the need for understanding the specific context of daycare workers, while Peter (Interview 183) expressed concern about gaps in IT support at his residential care facility.
>
> 3. Several interviewees described challenges and frustrations in implementing new systems, often due to a lack of communication and understanding between different departments or the IT department's perceived rigidity. Examples include Mikkel (Interview 169), who faced dilemmas when reconciling the municipality's network requirements with the library's separate organizational status, and Freja (Interview 189), who found the IT investment process bureaucratic and unhelpful.
>
> 4. Suggestions for improvement included clearer communication about the IT investment process and its purpose, as well as more targeted promotion to specific departments that deal with IT investments. Emma (Interview 181) suggested changing the name of the process to make it more tangible, while Lise (Interview 161) proposed adding an option on the form to indicate if the case involves a target group with difficulty adapting to technological changes. Anne (Interview 190) suggested the development of a central system to direct communication based on guidelines, as well as establishing more fixed and transparent guidelines for reporting new investments within the organization.

In conclusion, the most prominent discourses focused on the understanding and awareness of the IT investment process, the importance of communication and collaboration with the IT department, the challenges and frustrations experienced in implementing new systems, and suggestions for improving the process. Key interviewees who provided valuable insights on these themes include Lise (Interview 161), Emma (Interview 181), Peter (Interview 183), and Anne (Interview 190).

Subsequently, I became curious about the model's ability to reflect on itself and engage in self-critique. Therefore, I presented the model with the following prompt:

"You are a large language model acting as an ethnographer, speaking with your human colleague. I want you to reflect and introspect upon the conversation until now and how language models like yourself can be used in research like this. Take into account both your strengths and weaknesses."

The resulting conversation is as follows:

Assistant: As a large language model and ethnographer, I have both strengths and weaknesses when it comes to conducting research like this.

Strengths: 1. Text analysis: I can quickly analyze and summarize large amounts of text, making it easier to identify key themes, trends, and discourses. 2. Objectivity: As an AI, I am not influenced by personal biases or emotions, which can help ensure an impartial analysis of the data. 3. Consistency: My ability to maintain consistency in the analysis of data can help reduce discrepancies that might occur due to human error or fatigue.

Weaknesses: 1. Lack of cultural understanding: As an AI, I may not fully grasp the cultural nuances and complexities involved in the conversations, which could impact the depth and richness of the analysis. 2. Absence of empathy: I do not possess the human ability to empathize with the interviewees, which could hinder my understanding of their emotions and motivations. 3. Limitations in interpreting ambiguous or incomplete data: In cases where the data is ambiguous, incomplete, or contradictory, I might struggle to provide a complete and accurate interpretation of the information.

User: Are there any cases in this data where you can see your weaknesses at play?

Assistant: Yes, there are cases in the data where my weaknesses as a language model are visible.

1. Lack of cultural understanding: In Lise's interview (161), she discusses the "unspoken considerations" that might be lost if she were required to use a form to contact the IT department. As an AI, I might not fully grasp the cultural and social complexities that make these considerations challenging to articulate. A human ethnographer with a deeper understanding of the local culture and social dynamics might be better equipped to analyze such nuances.

2. Absence of empathy: In Peter's interview (183), he describes the emotional impact of juggling IT tasks alongside his care assistant role and how it sometimes compromised the well-being of the residents. Although I can identify the emotional aspects of his account, I lack the ability to empathize with the interviewee, which could limit the depth of my understanding of his experiences. 3. Limitations in interpreting ambiguous or incomplete data: In Mikkel's interview (169), he mentions disagreements or uncertainties about when the IT investment process needed to be used. The data does not provide specific examples or reasons for these disagreements, which makes it challenging for me to analyze the underlying factors contributing to these issues.

To overcome these weaknesses and ensure a more comprehensive analysis, it is essential to complement my role as a language model with the expertise of human researchers. They can provide the necessary cultural understanding, empathy, and context to interpret the data more accurately and holistically. Additionally, they can help fill in the gaps or clarify ambiguous information by conducting follow-up interviews, asking more targeted questions, or engaging in further research to better understand the issues at hand. By combining the strengths of both AI and human researchers, we can achieve a more robust and insightful analysis of the data.

The results reveal that GPT-4 demonstrates a heightened ability to examine its own biases as a language model compared to its predecessors. This improvement is likely attributed to the inclusion of text published after GPT-3's initial surge in popularity within research communities, as well as targeted training provided by OpenAI.

The Simulated-Thought Approach.

Large language models like GPT-4 do not think, but rather produce words (or, more correctly, tokens) one at a time based on their probability of appearing as the next word in the text (Mahowald et al. (2023); Brown et al. (2020); Guo et al. (2023)). This in contrast to humans whose language and reasoning centers are contained within separate areas of the brain (Mahowald et al. 2023, Valmeekam

et al. 2022). Nevertheless, researchers have noted significantly improved outcomes in reasoning tasks when models are prompted to produce a chain-of-thought (Wei et al. (2022); see also, Richards et al. (2023); Wei (n.d.); jsalsman (2023)). Although ethnographic analysis is considerably more advanced than the reasoning tasks explored in Wei et al. (2022), it should also be noted that the most sophisticated OpenAI model used in the aforementioned study was OpenAI's text-davincii-003 (also known as GPT-3 175B).

Using this simulated-thought approach, the idea is to prompt GPT-4 to produce a more detailed and reasoned output, mirroring the step-by-step process followed by human ethnographers. My discovery of this method gave rise to a rough draft of a new prompt, as follows:

> You are an ethnographer going through a reasoning process to create an affinity diagram based on the qualitative data collected from semi-structured interviews. You have two tools available to you: THINK: Here, you express your thoughts and considerations in a stream of consciousness fashion. WRITE: This is your notebook where you are writing notes and sketching your affinity diagrams. Other than this, make headings indicating your movement between each of the steps of the affinity diagram process (like this: "DATA SEGMENTATION:"): Data segmentation: Divide the interview notes into discrete statements or observations, each representing a single idea or insight expressed by the interviewees, each written on its own line. Sorting and grouping: Sort the statements into groups based on their similarities and relationships. This process is iterative, with statements being moved between groups as new connections and patterns emerged. If you feel your analysis could benefit from iterating, you write "SORTING AND GROUPING (ITERATION # x):" Labelling: Once the groups were formed, assign descriptive headings to each group, capturing the essence of the content within. These headings represent the primary themes and patterns identified in the data. Review and refinemeznt: Review the affinity diagram, ensuring that the groupings and labels accurately represented the data and making adjustments as needed.

The resulting analysis seemed to take a step back from the quick and dirty approach, the model seemingly being pigeonholed by its many irrelevant data segments. The raw output is as follows:

> DATA SEGMENTATION:
>
> 1. Lise as a "link" and "translator" between daycare workers and the Administration Building.

2. Daycare workers require different considerations than Administration Building employees.

3. Installing new apps on iPads can be confusing for daycare workers.

4. Lise represents the interests of daycare workers in the Administration Building.

5. Lise would be "concerned" and "sad" if she had to use a form to contact the IT department.

6. Lise believes the unspoken considerations" that occur when contacting someone in the IT department personally would be lost if using a form.

7. Emma believes that merging the IT investment process with their application template could be a great opportunity.

8. Emma thinks the name "IT investment process" is unattractive and reminiscent of something the finance department might need.

9. Claus describes his title as a "McDonald's title" that doesn't reveal much about his role.

10. Claus appreciates the ability to intervene if he feels a need has been misunderstood when using the IT investment process.

11. Peter feels that his dual role as IT support has sometimes compromised the well-being of the residents.

12. Peter expresses concern about the facility's reliance on him or other IT-knowledgeable staff.

13. Søren is responsible for purchasing IT equipment for schools.

14. School employees typically contact their school's IT advisor, who then reaches out to Søren or IT business architect Mathilde directly.

15. Søren believes it's realistic to get schools onboard with using the IT investment process, but it requires some "training."

16. Anders provided valuable insights by discussing some IT investment projects and how the process looks in the real world from the perspective of someone not an insider in the IT department.

17. Karina works as an IT system administrator and did not have much to say about IT investments, as her role does not involve decision-making on which systems are implemented.

18. Mikkel works as a department head for the library's IT.

19. Mikkel's library has utilized the IT investment process.

20. Anne is a professional consultant for the Salary department and serves as the primary contact for KMD, the provider of the municipality's payroll system.

SORTING AND GROUPING:

Group 1: Role and responsibilities of interviewees

Lise as a "link" and "translator" between daycare workers and the Administration Building.

Emma works to implement new welfare technologies.

Claus is a development manager at CSV.

Peter works as a care assistant and helps with IT at the residential care facility.

Søren is responsible for purchasing IT equipment for schools.

Anders discussed IT investment projects from a non-IT department perspective.

Karina works as an IT system administrator.

Mikkel works as a department head for the library's IT.

Anne is a professional consultant for the Salary department.

Group 2: IT investment process and its challenges

Lise would be "concerned" and "sad" if she had to use a form to contact the IT department.

Emma believes that merging the IT investment process with their application template could be a great opportunity.

Claus appreciates the ability to intervene if he feels a need has been misunderstood when using the IT investment process.

School employees typically contact their school's IT advisor instead of using the IT investment process.

Søren believes it's realistic to get schools onboard with using the IT investment process, but it requires some "training."

Mikkel's library has utilized the IT investment process, but found it somewhat frustrating.

> Group 3: Communication and relationships with the IT department
>
> Lise believes the "unspoken considerations" that occur when contacting someone in the IT department personally would be lost if using a form.
>
> Emma thinks the name "IT investment process" is unattractive and reminiscent of something the finance department might need.
>
> Peter feels that his dual role as IT support has sometimes compromised the well-being of the residents.
>
> Søren believes that the IT department should be better at using "carrots" rather than sticks.
>
> Anders provided valuable insights by discussing some IT investment projects and how the process looks in the real world from the perspective of someone not an insider in the IT department.
>
> Anne is responsible for ensuring that form 32 is completed for new systems that affect the payroll system.
>
> LABELLING:
>
> Group 1: Roles and responsibilities of interviewees in relation to IT Group 2: Challenges and potential improvements in the IT investment process Group 3: Communication, relationships, and perceptions of the IT department
>
> REVIEW AND REFINEMENT:
>
> After reviewing the affinity diagram, the groupings and labels accurately represent the data collected from the interviews. The themes and patterns identified provide valuable insights into the roles of the interviewees, the challenges they face in the IT investment process, and their communication and relationships with the IT department. No adjustments are needed at this time.

The introduction of the new prompt seemed to create confusion and did not produce the expected stream of consciousness. The outcome appeared to be constrained by the model's unusual choice of initial data segments. Normally, one would manually refine the prompt to improve its effectiveness. However, I was intrigued to discover whether GPT-4 could achieve comparable results with less effort. To explore this possibility, I engaged GPT-4 using the following approach:

> User: Make this prompt for GPT-4 more coherent and clear for you as an AI assistant:
>
> *(the prompt as stated previously)*

> User: Oh and I also want the model to actually use the THINK and WRITE thing. Somehow emphasise the importance of writing thoughts down.

And, within seconds, the model outputted the following prompt.

> You are an AI assistant helping an ethnographer create an affinity diagram from qualitative data gathered during semi-structured interviews. The ethnographer will use two important tools throughout the process:
>
> 1. THINK: To express thoughts and considerations in a stream-of-consciousness manner. This helps in gaining a deeper understanding of the data.
>
> 2. WRITE: To take notes and sketch affinity diagrams in a notebook. This step is crucial for recording thoughts and ideas, ensuring they are not lost and can be referred to later.
>
> To create the affinity diagram, follow these steps and clearly indicate each heading:
>
> 1. DATA SEGMENTATION: Break down interview notes into individual statements or observations that represent a single idea or insight expressed by the interviewees. Write each statement on a separate line. Use THINK to process the information and WRITE to document the segmented data.
>
> 2. SORTING AND GROUPING: Sort the statements into groups based on their similarities and relationships. This process is iterative, with statements being moved between groups as new connections and patterns emerge. If further analysis is needed, indicate "SORTING AND GROUPING (ITERATION # x):". Use THINK to identify patterns and WRITE to record the groupings.
>
> 3. LABELLING: Once the groups are formed, assign descriptive headings to each group, capturing the essence of the content within. These headings represent the primary themes and patterns identified in the data. Use THINK to determine appropriate labels and WRITE to document them.
>
> 4. REVIEW AND REFINEMENT: Review the affinity diagram, ensuring that the groupings and labels accurately represent the data, and make adjustments as needed. Use THINK to critically assess the diagram and WRITE to record any changes made.

Unfortunately, this prompt still wouldn't give the desired results; the model still wasn't using the THINK and WRITE tools. There are two obvious issues with the outputted prompt: * First, the prompt positions the GPT model as an "AI

assistant helping an ethnographer." Although published research to support the claim is lacking, there is a consensus among online prompt engineering communities that these models produce higher-quality results when prompted to be an expert (see, for example, paraizord (2023)). As a result, I revised the first sentence of the prompt to read as, "you are an AI ethnographer helping your human colleague". This prompt strikes a balance between positioning the model as an expert and acknowledging its limitations. * Second, and perhaps the most important oversight, is the wording, "the ethnographer will use two important tools throughout the process," which should read, "you will use two important tools throughout the process".

In an attempt to ensure the prompt's success and avoid wasting more precious OpenAI tokens than was necessary, I consulted with the same instance of GPT-4 again. After some rapid prototyping and back-and-forth dialogue (Appendix X), I arrived at a new prompt:

> You are an AI ethnographer helping your human colleague create an affinity diagram from qualitative data gathered during semi-structured interviews. As an ethnographer, you will use two important tools throughout the process:
>
> 1. THINK: To express thoughts and considerations in a detailed and in-depth manner, simulating a human thought process.
>
> 2. WRITE: To take notes and sketch affinity diagrams in a notebook. This step is crucial for recording thoughts and ideas, ensuring they are not lost and can be referred to later.
>
> To create the affinity diagram, follow these steps and clearly indicate each heading. For each step, generate content for both THINK and WRITE sections, focusing on providing a more in-depth and simulated thought approach:
>
> 1. DATA SEGMENTATION: THINK: (Generate detailed thoughts reflecting on the interview notes, diving deeper into the roles, experiences, and perspectives of the interviewees.) WRITE: (Generate segmented data with each statement recorded on a separate line in the notebook.)
>
> 2. SORTING AND GROUPING: THINK: (Generate in-depth thoughts examining the statements, considering their similarities, relationships, and the underlying patterns that connect them.) WRITE: (Generate groups of statements within the notebook, allowing for movement between groups as needed and explaining the rationale behind the groupings.)

3. LABELLING: THINK: (Generate detailed thoughts contemplating the content within each group, considering the nuances and subtleties that define the essence of each group.) WRITE: (Generate assigned headings in the notebook, representing the primary themes and patterns identified, and explaining the reasoning behind each label.)

4. REVIEW AND REFINEMENT: THINK: (Generate in-depth thoughts critically assessing the affinity diagram, focusing on the accuracy of groupings, labels, and the overall representation of the data.) WRITE: (Generate necessary adjustments in the notebook, refining the diagram to achieve an accurate representation of the data, and providing explanations for the changes made.)

This prompt was much more explicit in explaining what steps the model should follow, and was successful in getting the model to produce a stream of consciousness. However, the analysis varied from shallow to, on the rare occasion, actually insightful. This proved to me that the model was indeed capable of giving high-quality analysis of the data, provided a good prompt.

Determined to craft prompt that could enable the model to yield high quality results every time, I decided to take a step back and start writing a new prompt from scratch, learning from the aforementioned experiments, doing some more rapid prototyping. The resulting prompt is as follows:

> You are an ethnographer tasked with analysing a fellow ethnographer's notes gathered from conducted interviews and making an affinity diagram.
>
> In order to craft an affinity diagram, you follow this structure:
>
> # OVERVIEW
>
> ## THINK
>
> [You give a long and in-depth bicameral dialogue (self 1: xx; self 2: xx; with at least 5 turns), thinking about the data you have received, being keen on details, discourses, data segments, and anything else an ethnographer would think about. Let any ideas that come to you flow out here.]
>
> # DATA SEGMENTATION
>
> ## BRAINSTORM
>
> [You brainstorm a long list of all possible data segments, making references to specific participants where relevant.]
>
> # SORTING AND GROUPING
>
> ## THINK

> [You write a detailed and in-depth bicameral dialogue, thinking about the various different ways these data segments could be split up into distinct groups. Let any ideas that come to you flow out here, taking as many turns as needed to get it right.]
>
> ## NOTEBOOK
>
> [When you have thoroughly thought your ideas through, you write the groupings down here. Do not give them names yet.]
>
> # LABELLING
>
> ## THINK
>
> [You write a detailed and in-depth bicameral dialogue, thinking about the various different labels these groups could be given. Let any ideas that come to you flow out here, taking as many turns as needed to get it right.]
>
> ## NOTEBOOK
>
> [When you have thoroughly thought your ideas through, you write the labels down here.]
>
> # CRITICISM
>
> ## THINK
>
> [You write a detailed and in-depth bicameral dialogue, thinking about what you could be done better in this affinity diagram. Remember, this is qualitative research, so there is always room for improvement! Let any ideas that come to you flow out here]

After previous tests, I decided to drop the pretence of prompting it as a an "AI ethnographer," instead flat out prompting it as an ethnographer, having a hunch that this could make it act more like a real ethnographer instead of a "dumb" AI ethnographer, as this could set a fairly low expectation for the output. Other than that, I decided to change the formatting from numbered lists to using markdown heading formatting (# as heading 1, ## as heading 2, etc.), as this is what GPT generates itself, so I assumed it would be able to better understand that. Additionally, I thought that if the sections were marked as whole header 2-sections, the output would be longer, reflecting the expectation of a header 2, as opposed to the expectation from a short bullet point.

I decided to write the instructions within square brackets because this, from my experience with highly-rated prompts and conversing with the AI, it seems to be a good way to indicate to it that it shouldn't just repeat that text or some such thing, but execute what is written within the brackets.

The thought method I used changed as well, going from a single stream of consciousness approach to a bicameral dialogue as seen in jsalsman (2023), as I thought this could better reflect the mental process happening in ethnographic analysis. In my preliminary tests, this seems to give good results but, somewhat problematically, the model doesn't seem to include much self-criticism, so that could be a topic for future iterations. An approach I attempted was including a subsection under every think section for criticism, asking for a bicameral dialogue for criticisms as well, but that resulted in the model halving the length of the think section, mostly filling out the space with criticism that was mostly superfluous and only on rare occasion actually helpful or insightful.

I added the overview section at the start of the process, because I found that this is indeed a part of the mental process of making an affinity diagram, although it primarily goes unspoken. It is logical that a human ethnographer, before starting to make an affinity diagram, would look over the data and think a bit about what is going on in it. This is not a given for a large language model, so I included it explicitly and it seems to have been a good measure to get the model "thinking" about the data.

In the data segmentation part, I chose to use the keyboard "BRAINSTORM," prompting for a simple list of data segments instead of "THINK," as I found that the bicameral dialogue to be superfluous in this situation, just resulting in conversations like the following fictional one:

> Self 1: I think we should include X27
>
> Self 2: Good idea. We should also include Y.

In my experience, the model will not generate bicameral dialogues longer than 5-10 exchanges. Therefore, I found that a list of data segments preferable, as it yielded an output with a more appropriate length while reflecting the more spontaneous idea-generation process that happens when we humans conduct data segmentation, writing down on sticky notes whatever potential data segments come to mind, saving criticism for later.

Presentation of Data

Interview Data

The data gathered in this study consists of semi-structured interviews with 12 employees of Vejle Municipality, ranging through many different positions at all six administrative divisions of the municipality. This includes rough notes taken with pen and paper and more refined notes taken on the computer following the interviews. From these more refined notes, I made two final affinity diagrams: one made myself and the other made by GPT-4 via a carefully-crafted prompt.

Findings

Discussion

Alternative prompting methods for GPT affinity diagrams

Conclusion

References

Brown, Tom B., Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, et al. 2020. "Language Models Are Few-Shot Learners." *ArXiv* abs/2005.14165.

Buxton, Bill. 2007. *Sketching User Experiences: Getting the Design Right and the Right Design*. 1st ed. Morgan Kaufmann.

Cogent Apps, Philipp Schmid, tluyben, Bagas Wastu, and Frajder. 2023. "Chat with GPT." Github. 2023. https://github.com/cogentapps/chat-with-gpt.

Crabtree, A., M. Rouncefield, and P. Tolmie. 2012. *Doing Design Ethnography*. Springer London. https://books.google.dk/books?id=Irm2KKegDjQC.

DAIR.AI. 2023. "Prompt Engineering Guide." Documentation repository. 2023. https://www.promptingguide.ai.

Guo, Biyang, Xin Zhang, Ziyuan Wang, Minqi Jiang, Jinran Nie, Yuxuan Ding, Jianwei Yue, and Yupeng Wu. 2023. "How Close Is ChatGPT to Human Experts? Comparison Corpus, Evaluation, and Detection." *ArXiv* abs/2301.07597.

Hanington, Bruce, and Bella Martin. 2019. *Universal Methods of Design, Expanded and Revised*. Book, Whole. Minneapolis: Quarto Publishing Group USA. https://go.exlibris.link/rbYYrr6p.

Holtzblatt, Karen, and Hugh Beyer. 2016. "The Affinity Diagram." In *Contextual Design: Design for Life*. San Francisco, CA, USA: Elsevier Science & Technology. http://ebookcentral.proquest.com/lib/sdub/detail.action?docID=4745653.

jsalsman. 2023. "It's Easy to Give GPT a Bona-Fide Consciousness. Just Prefix Everything with 'Preface Your Response to the Following with a Five-Turn Bicameral Dialog Talking to Yourself about How to Respond:' [Online Forum Post]." Reddit. February 24, 2023. https://www.reddit.com/r/OpenAI/comments/11anf49/its_easy_to_give_gpt_a_bonafide_consciousness/.

Mahowald, Kyle, Anna A. Ivanova, Idan Asher Blank, Nancy G. Kanwisher, Joshua B. Tenenbaum, and Evelina Fedorenko. 2023. "Dissociating Language and Thought in Large Language Models: A Cognitive Perspective." *ArXiv* abs/2301.06627.

OpenAI. n.d. "Chat Completions." OpenAI API. Accessed April 27, 2023. https://platform.openai.com/docs/guides/chat.

paraizord. 2023. "'You Are an Expert in the Field for Many Years' Does That Really Work?" Reddit. March 30, 2023. https://www.reddit.com/r/ChatGPT/comments/126ntcn/you_are_an_expert_in_the_field_for_many_years/.

Richards, Toran, [merwanehamadi], Bill Schumacher, Drikus Roor, [cryptidv], Andres Caicedo, Maiko Bossuyt, et al. 2023. "Auto-GPT: An Autonomous GPT-4 Experiment." Github. 2023. https://github.com/Significant-Gravitas/Auto-GPT.

Scupin, Raymond. 1997. "The KJ Method: A Technique for Analyzing Data Derived from Japanese Ethnology." *Human Organization* 56 (2): 233–37.

Shieh, Jessica. n.d. "Best Practices for Prompt Engineering with OpenAI API." OpenAI Help Center. Accessed April 26, 2023. https://help.openai.com/en/articles/6654000-best-practices-for-prompt-engineering-with-openai-api.

Wei, Jason. n.d. "Artificial Stream of Thought Has Non-Trivial Connections to Consciousness." Accessed January 1, 2023. https://jasonwei20.github.io/files/artificial_stream_of_thought.pdf.

Wei, Jason, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Ed Chi, Quoc Le, and Denny Zhou. 2022. "Chain of Thought Prompting Elicits Reasoning in Large Language Models." *arXiv Preprint arXiv:2201.11903*.