International Journal of Latest Engineering and Management Research (IJLEMR)
ISSN: 2455-4847
www.ijlemr.com || Volume 10 – Issue 07 || July 2025 || PP. 21-27

A Hybrid Approach to Identifying Target Audience Pain Points: A Comparison of In-Depth Interviews and AI-Assisted Prompts

Konstantin Zhuchkov
Owner of a private consulting practice
USA, NY

Abstract: This article compares a traditional deep-dive interview with AI-assistive prompting to surface how pain points of the target audience are identified. That is, latent user needs and affective attitudes must be surfaced quickly and scalably in response to the increasing demand by companies. The justification for this study lies in the high cost and time inertia associated with classical in-depth interviews, coupled with the potential applications of generative language models in speeding up the preliminary analysis of customer pains. This paper conducts an empirical comparison of resources, time consumption, and quality outputs between 20 interviews (approximately 720 minutes) and 20 AI-assisted session logs structured under the ABCDX and BDF frameworks. The novelty of this study lies in its mix-and-match method, which joins speed checks in idea generation with content analysis and attestation of AI-generated ideas from actual conversations. All that information is brought to one place—here—in detail, ordered by how often and how strongly it is mentioned. Findings almost fully converged on the pain points; over 80% of hypotheses generated by LLMs were validated during interviews. Moreover, it takes about 44 minutes less to complete the initial analysis each time after AI assistance (9 minutes per session). In-depth interviews retained their importance at the stages of validation and refinement of "disputed" topics because clarifying context, including nonverbal signals, proved crucial. The optimal protocol recommends starting with an AI session for broad hypothesis generation, then proceeding to targeted in-depth interviews for confirmation and elaboration of key insights, enabling up to 50 % time savings over classical qualitative analysis without compromising result reliability. This article will be helpful to researchers in marketing and user experience, product managers, and analysts engaged in qualitative research.
Keywords: hybrid approach, in-depth interview, AI-assisted prompts, pain points, ABCDX, BDF, qualitative research, hypothesis verification

Introduction

Understanding the pain points of a target audience remains a critical condition for the successful market launch of a product: it is the depth of insights into users' problems and motivations that explains variations in campaign performance metrics, according to meta-analyses of consumer behavior studies. However, pain points are often concealed behind rational formulations of requests, and their reconstruction requires a combination of cognitive and emotional analysis, which heightens scientific interest in methods capable of revealing "non-obvious" reasons for choice.

Historically, such depth has been provided by face-to-face or remote in-depth interviews, whose average duration consistently ranges from 81 to 96 minutes (Irvine, 2011), while the total time for transcription and coding of even a small corpus (for example, 20 interviews of 36 minutes each) reaches 70–80 hours of analytical work, increasing the cost and limiting the scalability of the method (Ullrich et al., 2025). Added to this are the risks of interpretive bias and dependence of results on interviewer qualifications; therefore, in fast-iterating product environments, the classical approach is increasingly deemed insufficiently agile.

Against this background, over the past two years, a new research paradigm based on large-scale generative language models has emerged. According to a global McKinsey survey, the regular use of Generative AI in corporate processes increased to 65% in less than one year (Singla et al., 2024), and the share of companies applying such models across five or more business functions reached 15% (Singla, 2024). The case studies demonstrated not only increased adoption but also significant economic impact. For instance, in one year, Deloitte's internal assistant managed 3.65 million requests and reduced labor costs for expert report preparation by nearly half (Deloitte, 2025). These data points also suggest that LLM-assisted prompts can be-a much quicker way to establish customer pain as an effective alternative or complement to the slower, more expensive traditional interview process. However, an empirical comparison of the two approaches is still lacking and warrants a systematic investigation.

Materials and Methodology

The study of the hybrid approach to identifying target audience pain points is based on the integration of two types of empirical data: a corpus of 20 in-depth interviews totaling approximately 720 minutes (Irvine,

2011; Ullrich et al., 2025) and the logs of 20 AI-assisted sessions structured according to the ABCDX and BDF frameworks (Chen et al., 2025). The theoretical foundation comprises classical works on in-depth interviews, confirming that the average interview length ranges from 81 to 96 minutes, and that the transcription and coding of 20 interviews, each 36 minutes long, require up to 80 hours of analytical work (Irvine, 2011; Ullrich et al., 2025). Concurrently, industry reports from McKinsey on the growth of Generative AI usage and LLM penetration (Singla et al., 2024) and Deloitte's data on labor cost reduction via its internal AI assistant (Deloitte, 2025) were considered.

The research comprised various complementary steps from a methodological perspective. First, resources and speed were comparatively analyzed. Preparation and coding time for interviews was estimated based on the surveys of both the analysts and CAQDAS tools (NVivo, ATLAS.ti). In turn, the time experimentally recorded for initial hypothesis generation in AI sessions (Brand et al., 2023). Secondly, a systematic review of methodological frameworks is presented to detail the emotional component in standardized prompt construction, utilizing schemes such as ABCDX for audience segmentation and BDF (Chen et al., 2025). The third stage involved content analysis and validation: AI-generated hypotheses were compared with live interview results, with the "majority" of original pain points confirmed through field testing, indicating high method convergence (Brand et al., 2023).

The quality and reproducibility of the results were ensured through double coding of interview data, calculation of intercoder agreement, and member-checking with three participants after the primary interview cycle (Rosala, 2021; Guest et al., 2006). Similarly, AI logs were analyzed for completeness and segment duplication. The final step was to integrate all sources (transcripts and AI logs) into a single working environment in Google Sheets or Notion, where problems were ranked by frequency of mention and emotional intensity. This allowed for the construction of a consolidated matrix of verified pains for subsequent use in product solutions.

Results and Discussion

In-depth interviews are individual, semi-structured, or unstructured conversations of approximately one hour in duration, conducted by a researcher with a participant to reconstruct the respondent's experiences, motivations, and hidden attitudes. The primary objectives of this format are to identify users' cognitive and emotional "pain points," clarify the context in which they emerge, and test hypotheses regarding decision-making factors. Interviews are employed at early stages of product development, when it is necessary to build an initial insight map and, based on real utterances, form working audience segments.

The key advantage of the method is depth: the interlocutor reveals personal stories and behavioral nuances that are almost impossible to capture via questionnaires. The researcher can instantaneously rephrase a question, request an example or clarification, thereby deepening the conversation while preserving the participant's natural train of thought. The presence of live context helps anchor emotion, language, and situation to each quote, enhancing the validity of subsequent interpretation.

The price of such detail is resource intensity. Post-processing one hour of interview generates substantial coding and thematic-analysis time, and manual transcription doubles the time expenditure. In the field, constraints are also tangible: recruiting relevant informants, scheduling, and renting a neutral venue increase the budget. Moreover, qualitative analysis inevitably contains a subjective layer; different researchers may code the same fragment differently. To reduce interpretive bias, it is critical to implement double coding and "member-checking" sessions with participants. Another methodological challenge is the point of saturation. An aggregated review of empirical studies shows that most recurring themes emerge by the twelfth interview, after which adding new respondents yields diminishing returns (Rosala, 2021). This implies that over-recruitment inflates costs, while under-recruitment leaves significant pains undiscovered.

To extract maximum value under limited resources, it is helpful to follow several practical principles. First, prepare a semi-structured guide with open-ended questions and built-in "probes" (e.g., "Tell me what you felt when…"), leaving room for a spontaneous topic to emerge. Second, begin analysis in parallel with fieldwork: after every two or three interviews, record preliminary codes and adjust the script if insights start to repeat. Third, use CAQDAS tools (such as NVivo or ATLAS.ti) to standardize coding and calculate intercoder agreement. Finally, upon completing the series, conduct a brief member-checking session: return key findings to two or three participants and ask them to confirm or refute the interpretation. This reduces bias and builds up confidence in the data. It also does not inflate the research budget.

A language model is used as a "respondent simulator" in AI-assisted prompts: the researcher makes a highly structured request, and the model generates a list of supposed pains — motives — and contexts based on its probabilistic representation. This is very effective because, through immense volumes of user-generated content training, the model has learned to swiftly re-combine like patterns. In a Harvard Business School experiment, generating a rough list of insights took on average nine minutes, compared with forty-four minutes

required for preliminary manual analysis of focus-group protocols. Most pain hypotheses were later confirmed in real interviews (Brand et al., 2023).

To ensure the model's outputs are meaningful, requests are constructed according to predefined frameworks. At the segmentation level, the ABCDX scheme is used, wherein the audience is decomposed into "who they are," "how they behave," "in what context," "what they desire," and "what experience they already have." The prompt explicitly enumerates these five axes, and the model returns hypotheses for each, reducing the risk of omissions and segment duplication. To detail each pain point, a BDF perspective is added: a request to describe the group's Beliefs, Desires, and Feelings activates an emotionally charged vocabulary, helping distinguish a superficial complaint from a deep-seated motivation.

The visible output of a prompt session is a structured table of pains, ranked by model-reported frequency of mention and emotional intensity. This draft does not replace interviews but serves as an idea filter: the researcher saves time locating obvious answers and focuses on refining contentious points. A recent review of prompt engineering emphasizes that the combination of "LLM hypotheses → human validation" yields more reliable solutions than either technique alone, as it minimizes both the model's "blind spots" and researchers' biases (Chen et al., 2025). Ultimately, AI-assisted prompts become not an alternative but an accelerator of classic qualitative research, providing rapid coverage of potential pains and framing further in-depth dialogue.

The operational hypothesis is generated by starting with an exact request format: the researcher provides a brief product context to the model, followed by a request to 'elucidate pains according to the ABCDX scheme, and for each item add a BDF matrix.' Below is a typical working prompt: 'Imagine you are a marketing analyst. The product is an online course designed for freelancers to manage their finances. List ten audience pains, structuring your answers along the A-B-C-D-X axes; for each pain, identify Underlying Belief, Core Desire, and Dominant Feeling."

Upon receiving the rough list, one can immediately refine details: "Focus on pain #3 and propose five in-depth interview questions to test its presence." This dialogic format replicates the flow of fundamental research, albeit at an accelerated pace, without the need for respondent recruitment. The prompt structure is shown in Figure 1.

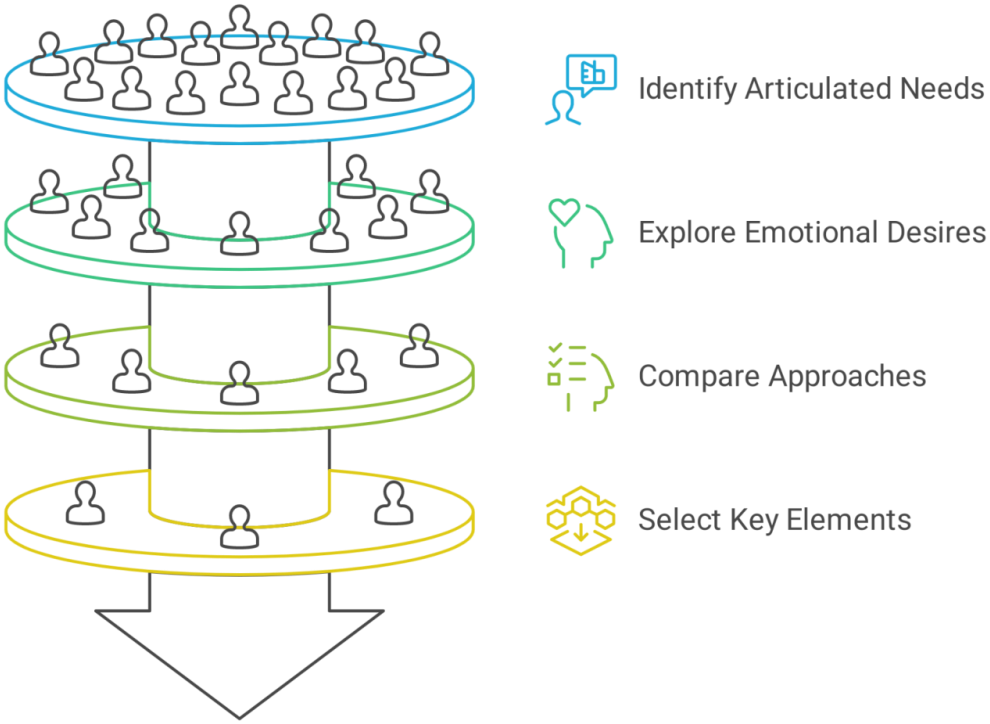

Fig. 1: Structure of the AI prompt by stages of audience pain refinement (compiled by the author)

The method's advantage is evident in its speed and breadth of idea coverage. A controlled GitHub experiment demonstrated that developers using Copilot completed tasks 55.8% faster than a control group without AI assistance, with equivalent code quality (Nuttur et al., 2025). Unsurprisingly, 65% of companies surveyed by McKinsey in 2024 have already introduced generative AI into at least one business process, and audience research has identified it as one of the most frequent applications (Singla et al., 2024). Meanwhile, the

chatbot market is projected to increase from $7.76 billion in 2024 to $27.29 billion by 2030, at a compound annual growth rate (CAGR) of 23.3%, as reflected in Fig. 2 (Grand View Research, 2024).

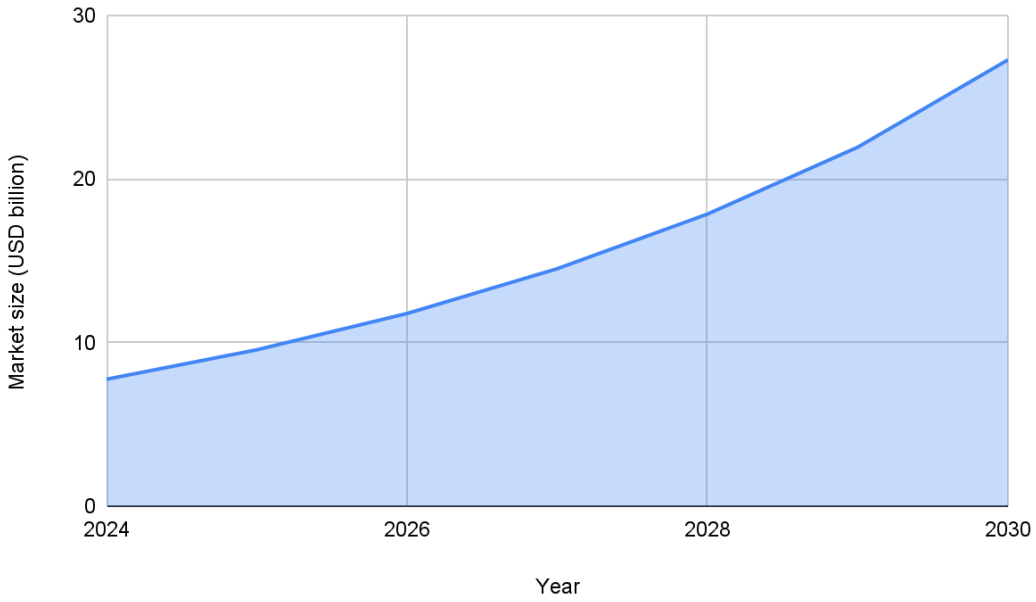

Fig. 2. Growth of the chatbot market (Grand View Research, 2024)

The rapid cycle "prompt → hypothesis → refining prompt" works like a spark for brainstorming: within minutes, a map of pains is assembled, which the team then ranks and validates.

Nevertheless, the tool is not without limitations. The model relies on statistical correlations within its training corpus and is therefore prone to output averaged or culturally predictable issues until the researcher supplies market-specific context. When source data is insufficient, the output becomes generalized; furthermore, formulations often reflect the language of marketing materials rather than respondents' actual speech, reducing the suitability of insights for creative work. Finally, training corpora are fixed in time, so pains related to new regulations or emerging trends may be omitted. For these reasons, AI-assisted prompts serve as an accelerator for idea generation but require subsequent verification in field interviews to eliminate false-positive insights and adapt terminology to authentic audience voices.

The choice between in-depth interviews and AI-assisted sessions begins with a clear understanding of the value sought by the researcher at each stage of the product lifecycle. In-depth conversations remain the gold standard when it is necessary to uncover hidden semantic layers. Classical studies indicate that thematic saturation typically occurs by the twelfth respondent, enabling the identification of stable patterns of customer motivation and pain (Guest et al., 2006). Within a single live session, the interviewer can adjust the course of the conversation, pursue a "why?" chain, and connect facts with context and nonverbal cues. By this criterion, the method leads confidently in terms of insight depth.

However, such detail comes at a high cost. A typical thirty-minute individual session—including recruitment, moderation, and incentives—costs USD 400–500, equating to approximately USD 16–25 per "speaking" minute (Palmerino, 2012), and transcribing one hour of audio can occupy up to eight hours of an analyst's time (Berkovic, 2023). Thus, under tight budgets or deadlines, the classical approach is rational only when it is necessary to verify key hypotheses before costly product decisions or to calibrate nuances of the value proposition.

AI-assisted prompts, by contrast, excel as rapid generators of hypotheses. Moreover, automated thematic tagging of user feedback demonstrates human-comparable agreement, as evidenced by a JMIR study that found 37% code matches between ChatGPT and human coders (Prescott et al., 2024). The predictive analytics market is forecasted to grow from $22.22 billion in 2025 to $91.92 billion by 2032, at a CAGR of 22.5%, as shown in Fig. 3 (Fortune Business Insights, 2025). Such figures confirm that in the early discovery phase, an LLM can quickly and efficiently assemble a "cloud of problems" around a product, primarily when prompts are structured via the ABCDX or BDF frameworks, which impose a logical direction on reasoning.

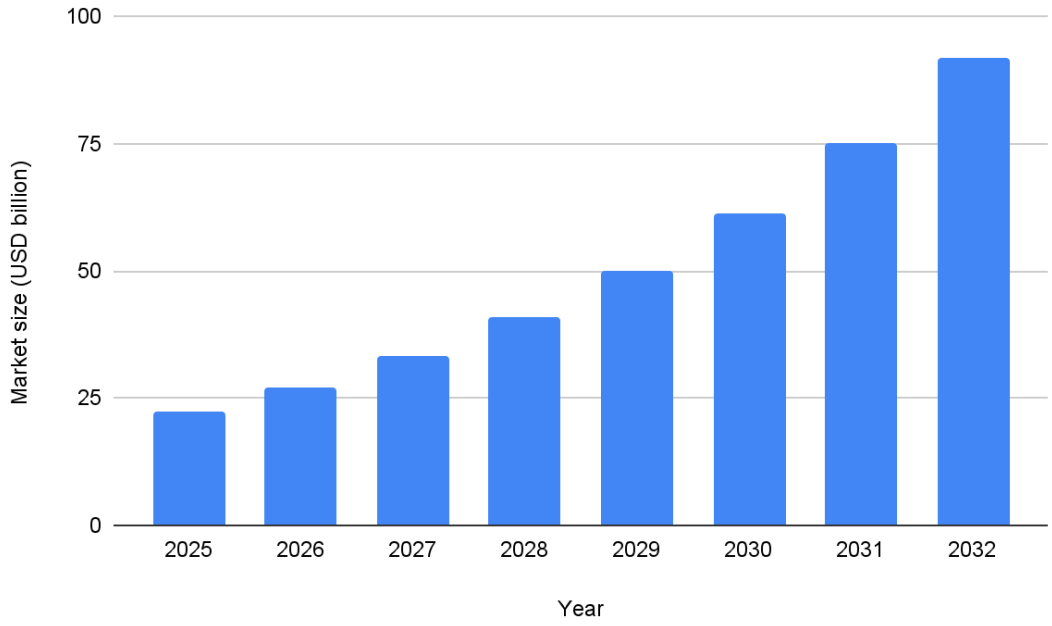

Fig. 3: Growth of the predictive analytics market (Fortune Business Insights, 2025)

An optimal task allocation appears as follows: at the idea stage, when maximum scenario coverage is critical, an AI session provides the required breadth without substantial cost; when it is time to test high-stakes market assumptions—price thresholds, trust barriers, language of benefits—personal interviews allow the idea mass to be "collapsed" into verified insights. A standard error in the first approach is an under-informed prompt: if the prompt lacks industry context, the model offers clichés. Risk can be minimized by injecting specific target-segment facts into the prompt and explicitly assigning the model the role of expert analyst. For interviews, the primary threat is interpretive noise, including leading questions, biased summaries, and moderator influence. A pre-prepared script with neutral wording and double coding of transcripts helps maintain validity within acceptable bounds. Thus, a hybrid mode—where the LLM provides breadth and interviews provide depth—enables faster progress without sacrificing methodological rigor.

The hybrid process usually starts with a fast AI session, where the researcher, armed with a concise prompt and minimal product brief, obtains a map of primary pain hypotheses in minutes; with a generation speed comparable to recording a brief call, time savings reach approximately 1.5× compared to classical manual screening, as laboratory experiments confirm: in a programming task the group using Copilot hints completed work 55.8 % faster than the control (Peng et al., 2023). This "watchtower" quickly illuminates the field of possible topics and provides the researcher with guides for in-depth conversations. Then comes the validation stage: interviews focus on paradoxes and unexpected formulations that the LLM produced without an empirical basis. Qualitative methodology practice shows that in a relatively homogeneous audience, thematic saturation occurs by about the twelfth conversation, after which new respondents add almost no fresh codes (Vasileiou et al., 2018); this allows targeted confirmation or refutation of AI hypotheses and refinement into validated insights suitable for product decisions.

Reversed order—first interviews, then AI augmentation—is justified when the product is already in contact with users and initial transcripts are available. In this case, rapid thematic tagging by an LLM serves as a multiplier: the model classifies responses according to the ABCDX and BDF frameworks, groups rare but potentially valuable signals, and proposes methods for their ranking. The organizational benefits here strongly correlate with broad institutional adoption. Thus, interviews create a deep seed lexicon of real quotations, and the model cheaply propagates it across hundreds of semantically similar formulations, showing the team which motives recur most often and where to focus resources.

Regardless of the sequence of actions, the process demands discipline. First, collect all available input data, including product descriptions, existing user feedback, metric summaries, and client personas, so that the prompt does not operate in a vacuum. Next, create a prompt library: a base prompt for pain-point generation; a refining prompt for the second iteration; and a translational prompt to adjust language for different segments. At the same time, work on an interview script that creates questions from the riskiest AI-generated hypotheses and looks for leading wording to prevent bias. After fieldwork is completed, export transcripts and model-dialogue

logs into one working file.The analyst then applies a coding schema, aligns theme frequencies with business priorities, and records decisions, which include those that are confirmed, those that are discarded, and those that require further data gathering.

Documenting results works best through easily accessible, interoperable tools. Google Docs is convenient for storing finalized pain-point formulations and linking to quotations; Google Sheets supports problem ranking and calculation of priority segments; and for team workflows, Notion or Trello allow attachment of prompt screenshots, interview cards, and hypothesis-checking checklists, maintaining the entire hybrid research process within one linked data structure that can be updated as new product iterations are released.

Conclusion

The conducted study underscores that classical in-depth interviews and AI-assisted prompts are complementary tools for identifying pain points among the target audience. The in-depth interview method provides unparalleled depth of understanding regarding respondents' motivations and emotional reactions: live dialogue enables prompt contextual clarification of statements, deep exploration of hidden attitudes, and capture of nonverbal cues, thereby enhancing the validity and precision of result interpretation. High preparatory, conduct, and analytical coding costs associated with such interviews limit their scalability and agility, particularly when rapid iterations are required within a product cycle. On the other hand, AI-assisted prompting demonstrates phenomenal effectiveness in generating hypotheses and drafts of insights lists within minutes. Structured use of the ABCDX and BDF frameworks helps researchers draw a wide array of assumptions regarding users' needs and feelings; however, the automated approach has its limitations: typical model output is somewhat generic and may not take into account ongoing market trends; therefore, must be subjected to rigorous empirical testing in field conditions.

An optimal strategy is the hybrid approach, in which the first phase of research begins with an AI-assisted session that provides quick coverage of potential pain points and the creation of a preliminary insight map. In the subsequent phase, the generated hypotheses undergo focusing and confirmation through in-depth interviews aimed at testing "disputed" and most significant themes. This combination can save up to half the time of classical qualitative analysis while maintaining high reliability and accuracy of conclusions through human validation.

Finally, implementing a hybrid process requires strict organizational discipline: it is essential to prepare clear contexts and prompt scripts in advance, develop interview guides in parallel that incorporate preliminary AI hypotheses, and ensure integration of all data into a unified working environment. The use of contemporary CAQDAS tools and member-checking practices helps reduce interpretive bias. Analytical platforms for storing and visualizing results accelerate team decision-making. Thus, the hybrid approach opens new doors for qualitative research by combining the speed and scope of AI-driven generation with the depth and nuance of traditional interviews.

References

[1]. Berkovic, D. (2023). Chapter 13: Interviews. *Open Educational Resources Collective*. https://oercollective.caul.edu.au/qualitative-research/chapter/__unknown__-13/

[2]. Brand, J., Israeli, A., & Ngwe, D. (2023). Using GPT for Market Research. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4395751

[3]. Chen, B., Zhang, Z., Langrené, N., & Zhu, S. (2025). Unleashing the potential of prompt engineering for large language models. *Patterns*, 101260. https://doi.org/10.1016/j.patter.2025.101260

[4]. Deloitte. (2025, July). *Why Generative AI is becoming part of the infrastructure of work*. The Australian. https://www.theaustralian.com.au/business/tech-journal/why-generative-ai-is-becoming-part-of-the-infrastructure-of-work/news-story/b3f1ed9b345201c70da1054732953968

[5]. Fortune Business Insights. (2025). *Predictive Analytics Market Size*. Fortune Business Insights. https://www.fortunebusinessinsights.com/predictive-analytics-market-105179

[6]. Grand View Research. (2024). *Chatbot Market Size And Share Analysis*. Grand View Research. https://www.grandviewresearch.com/industry-analysis/chatbot-market

[7]. Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field Methods*, *18*(1), 59–82. https://doi.org/10.1177/1525822X05279903

[8]. Irvine, A. (2011). Duration, Dominance and Depth in Telephone and Face-to-Face Interviews: a Comparative Exploration. *International Journal of Qualitative Methods*, *10*(3), 202–220. https://doi.org/10.1177/160940691101000302

[9]. Nettur, S., Karpurapu, S., Nettur, U., Gajja, L., Myneni, S., & Dusi, A. (2025). *The Role of GitHub Copilot on Software Development: A Perspective on Productivity, Security, Best Practices and Future Directions*. Arxiv. https://arxiv.org/pdf/2502.13199

[10]. Palmerino, M. (2012, January 3). *Qualitatively Speaking: One-on-ones put the quality in qualitative*. Quirks. https://www.quirks.com/articles/qualitatively-speaking-one-on-ones-put-the-quality-in-qualitative

[11]. Peng, S., Kalliamvakou, E., Cihon, P., & Demirer, M. (2023). The Impact of AI on Developer Productivity: Evidence from GitHub Copilot. *Arxiv Software Engineering*. https://doi.org/10.48550/arxiv.2302.06590

[12]. Prescott, M. R., Yeager, S., Ham, L., Rivera, C. D., Serrano, V., Narez, J., Paltin, D., Delgado, J., Moore, D. J., & Montoya, J. (2024). Comparing the Efficacy and Efficiency of Human and GenAI Qualitative Thematic Analyses. *JMIR AI*, *3*, 54482–54482. https://doi.org/10.2196/54482

[13]. Rosala, M. (2021, October 31). *How Many Participants for a UX Interview?* Nielsen Norman Group. https://www.nngroup.com/articles/interview-sample-size/

[14]. Singla, A. (2024, July 3). *Gen AI casts a wider net*. McKinsey & Company. https://www.mckinsey.com/featured-insights/sustainable-inclusive-growth/charts/gen-ai-casts-a-wider-net

[15]. Singla, A., Sukharevsky, A., Yee, L., & Chui, M. (2024, May 30). *The state of AI in early 2024: Gen AI adoption spikes and starts to generate value*. McKinsey & Company. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-2024

[16]. Ullrich, C., Wensing, M., Klafke, N., Fleischhauer, T., Brinkmöller, S., Poß-Doering, R., & Arnold, C. (2025). Assessing the time required for qualitative analysis: A comparative Methodological study of coding interview data in health services research. *Das Gesundheitswesen*. https://doi.org/10.1055/a-2512-8004

[17]. Vasileiou, K., Barnett, J., Thorpe, S., & Young, T. (2018). Characterising and Justifying Sample Size Sufficiency in interview-based studies: Systematic Analysis of Qualitative Health Research over 15 years. *BMC Medical Research Methodology*, *18*(1), 1–18. https://doi.org/10.1186/s12874-018-0594-7