"I Need to Find That One Chart": How Data Workers Navigate, Summarize and Communicate Analytical Conversations

Ken Gu
University of Washington
Seattle, Washington, USA
kenqgu@cs.washington.edu

Srishti Palani
Tableau Research
Palo Alto, California, USA
srishti.palani@salesforce.com

Vidya Setlur
Tableau Research
Palo Alto, California, USA
vsetlur@tableau.com

Abstract

Conversational interfaces are increasingly used for data analysis, enabling data workers to express complex analytical intents in natural language. Yet, these interactions unfold as long, linear transcripts that are misaligned with the iterative, nonlinear nature of real-world analyses. Revisiting and summarizing conversations for different contexts is therefore challenging. This paper investigates how data workers navigate, make sense of, and communicate prior analytical conversations. To study behaviors beyond those supported by standard interfaces (i.e., scrolling and keyword search), we develop a design probe that supplements analytical conversations with structured elements and affordances (e.g., filtering, multi-level navigation and detail-on-demand). In a user study (n = 10), participants used the probe to navigate and communicate past analyses, fulfilling information needs (recall, reorient, prioritize) through navigation strategies (visual recall, sequential and abstractive) and summarization practices (adding process details and context). Based on these findings, we discuss design implications to support re-visitation and communication of analytical conversations.

CCS Concepts

• **Human-centered computing** → **Visualization systems and tools**; *Interactive systems and tools.*

Keywords

Conversational analysis, insight generation, human-AI interaction, adaptive summarization.

ACM Reference Format:
Ken Gu, Srishti Palani, and Vidya Setlur. 2026. "I Need to Find That One Chart": How Data Workers Navigate, Summarize and Communicate Analytical Conversations. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 33 pages. https://doi.org/10.1145/3772318.3790802

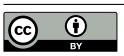
1 Introduction

Conversational interfaces powered by Large Language Models (LLMs) are now widely deployed as tools for data analysis [18, 59, 62, 91]. These systems enable users to express diverse, often open-ended analytical goals and to iteratively explore data through natural language, without the need for programming. This significantly lowers the barriers to entry, broadening access for *data workers* (i.e., individuals who access, analyze, manipulate, or communicate data to support decision-making, reporting, or operational tasks) [12, 40, 47, 61].

While prior work has examined how data workers steer LLMs [41, 80], execute analyses [31], and verify system outputs [26, 107], there is limited empirical understanding of how data workers interact with past analytical conversations to revisit, navigate, make sense of, and communicate their findings. These interactions are critical in real-world data workflows, such as sharing findings with business stakeholders or handing off analyses to collaborators [56, 65, 68, 75], but are poorly supported by the linear transcripts produced by current conversational interfaces.

Beyond the known limitations of LLM-based conversational interfaces (e.g., a lack of structure, information overload [36, 87]), analytical conversations exacerbate common challenges of data analysis in computational notebooks. These include overlapping code variants, interleaved heterogeneous outputs (i.e., visualizations, data tables, and code execution outputs), and fragmented analysis trails [43, 44]. These issues are further compounded by the abstraction of code in analytical conversations. As data workers no longer author code directly, they are less engaged in analytical decisions and less likely to retain a clear mental model of their analyses [16, 25, 26, 105]. This reduced engagement heightens the need for tools that support revisitation, navigation, and sensemaking.

Revisitation and communication are fundamentally intertwined in analytical practice. Drawing on Pirolli and Card's sensemaking framework [70], we recognize that foraging (i.e., navigating and finding information) and synthesis (i.e., constructing meaning) form an iterative loop rather than sequential stages. Communication acts as a synthesis activity that reveals gaps in understanding, prompting renewed navigation. This integration is well-established in computational notebook research. Kery et al. [45] showed that analysts develop narratives during exploration, not just afterward, the narrative in the notebook emerges through use over time. Rule et al. [76] demonstrated that exploration and explanation are interleaved rather than distinct phases. Similarly, Wang et al.'s Callisto [99] revealed how conversational reasoning and computational narratives co-evolve. For analytical conversations specifically, this coupling is even more critical. Unlike notebooks where code cells provide natural breaking points, conversational transcripts are continuous and linear. The same structures (threads, insights, artifacts) that support personal reorientation also enable external communication. Analysts often revisit conversations with communicative intent to prepare presentations, brief colleagues, or hand off analyses. Separating these activities would miss how analysts actually work. Thus, our study treats navigation and communication as inter-related activities sharing infrastructure (structured elements,



CHI '26, Barcelona, Spain

ACM ISBN 979-8-4007-2278-3/26/04
https://doi.org/10.1145/3772318.3790802

filtering, abstraction) while serving both personal sensemaking and collaborative goals. This design choice reflects the reality that understanding one's analysis and explaining it to others are mutually reinforcing processes.

While LLMs could be prompted to help navigate and communicate analytical conversations, their opaque generative processes offer little visibility into the provenance of their outputs, the rationale for the inclusion or exclusion of summarized content, or how that content aligns with communicative intent [32, 52]. Thus, there is a critical need to understand how data workers currently navigate, make sense of, and communicate analytical conversations, and what design affordances can support these activities while preserving their agency.

In this work, we seek to address this gap through a user study that investigates the information needs and workflow strategies when *navigating* and *communicating* analytical conversations (§4). We also explore the broader context of these behaviors, including how data workers engage with LLM-driven conversations and the role of AI in supporting navigation and communication.

To observe a wider range of strategies than standard chat interfaces afford, we developed a high-fidelity design probe, SyncSense for surfacing diverse behaviors and reflections. SyncSense augments conversational interfaces with structured elements and affordances such as filtering and multi-level navigation (§5), offering participants rich scaffolding to explore how they might revisit, structure, and communicate prior analyses. Rather than evaluating a fixed solution, our goal was to utilize the tool as a lens into emerging practices. Using this probe, we conducted a user study with 10 data workers across two sessions (§6). Participants revisited their analytical conversation conducted at least one week prior and summarized it for two distinct audiences (a non-technical executive and a technical analyst) across two formats: a chat message and a data report.

Our findings (§7) reveal recurring information needs, including *re-orienting* within a past conversation, *prioritizing* relevant information, and *recalling* specific details. Participants employed strategies such as sequential navigation, using visual cues as memory triggers, moving between levels of abstraction, and applying filters. When composing summaries, participants made manual edits such as adding audience context, reordering content, and formatting, preferring to retain control over core message curation. While participants valued AI assistance for refining tone, technicality, and length, they resisted full automation, favoring AI as a scaffolding and stylistic aid. We conclude with design implications for future conversational interfaces supporting analytical workflows (§8). Through this work, we make the following research contributions:

(1) Empirical observations from a two-session user study with 10 participants. The study addresses a critical gap in understanding how data workers revisit, make sense of, and communicate findings from analytical conversations, an increasingly common interface for data analysis. To support future research, we release our dataset of participants' LLM-mediated analytical conversations.
(2) An open-source high-fidelity design probe that supplements conversational interfaces with structured elements of analytical conversations and affordances to support observations of user needs and behaviors beyond standard chat interfaces.
(3) Design implications for future tools, emphasizing structured elements for navigation and communication, communication as a core and co-evolving analytical activity, and controllable, contextual AI assistance.

2 Related Work

Our work sits at the intersection of three key research themes: conversational interfaces for data analysis, communication challenges in collaborative data work, and sensemaking tools that support structure and dynamic abstraction.

2.1 Conversational Interfaces for Data Analysis

Conversational interfaces increasingly support data exploration by enabling natural language interactions that lower barriers to analysis. Early systems focused on intent parsing and visualization generation using rule-based or template-driven approaches [1, 2]. These tools gradually improved to support ambiguity resolution [22, 78], follow-up queries [86], and more sophisticated dialogue structures [17, 37, 79, 83]. More recently, the integration of LLMs has significantly expanded conversational capabilities in analytical tools, enabling more fluid and expressive interactions across spreadsheets [59], computational notebooks [91, 103], and analytical conversational interfaces [31, 107]. These systems streamline how data workers formulate analysis plans, refine workflows, and derive insights, relying on natural language as the primary interaction medium.

However, these capabilities come with new challenges. Navigation challenges emerge as conversations grow long and complex, with data workers struggling to locate specific analyses, revisit earlier findings, or understand the progression of their work [41, 107]. The flexibility of natural language makes it difficult to precisely express analytical intent, hindering data workers' ability to effectively steer model behavior and leading to misaligned or incomplete analyses [41, 107].

Sensemaking challenges amplify when AI executes analysis steps directly from high-level prompts, producing outputs like code, tables, and visualizations without user intervention [25]. This adds cognitive load, as data workers must interpret both the AI's responses and the resulting artifacts. While conversational transcripts preserve entire interaction histories, their linear structure obscures the true complexity of analytical work, including branching threads, shifting topics, and evolving goals that characterize real-world data analyses [28, 45, 76]. Post-hoc review becomes difficult, particularly when conversations involve sequential utterances that incrementally modify queries, analyses, or visual encodings [31, 41, 80, 105]. These difficulties also create barriers when conversations must be repurposed for communication, a challenge we discuss in detail next.

2.2 Communication and Collaboration in Data Analysis

Communicating data analyses represents a central but challenging aspect of data work. Team members often differ in domain expertise, tooling familiarity, and documentation habits, making it difficult to

maintain shared context [8, 12, 61, 98, 110]. Breakdowns frequently arise when analytical decisions, assumptions, and uncertainties are not clearly communicated, leaving both technical and non-technical stakeholders struggling to align [12, 110]. Misaligned expectations around analysis depth, presentation style, and tolerance for uncertainty further complicate collaboration across roles [97].

Since analyses reflect the data worker's unique mix of statistical, domain, and coding expertise, they need to be organized and reframed for different audiences and media [46, 69, 97, 100]. Traditional analytical mediums like computational notebooks and spreadsheets present well-documented challenges. Documentation is often incomplete, inconsistent, or disconnected from code and results [97, 98]. Creating comprehensive documentation is time-consuming and often deprioritized [65], which leads to difficulties when sharing work with colleagues or revisiting analyses later [84, 110]. In response, researchers have developed tools for enriching computational notebooks with better documentation [98], generating slides from notebook content [102, 115], and supporting audience-aware communication [6, 50, 101]. While computational notebooks remain the dominant medium for organizing and communicating analysis, our focus is on analytical conversations that increasingly precede or even replace notebooks as the primary locus of work when LLMs are involved.

Conversational interfaces exacerbate many of these communication challenges. Unlike notebooks or spreadsheets, which modularize work into cells or sheets, conversational analyses persist as long, static, linear transcripts where code, visualizations, natural language explanations, and iterative refinements are entangled. The exploratory nature of data analysis also means that false starts and dead ends persist in the transcript, adding noise that complicates downstream communication. To this end, conversational interfaces create significant challenges for navigation, sensemaking, and communication that current tools inadequately support. To explore how these challenges might be addressed, we use a design probe to study how data workers revisit and summarize analytical conversations for different audiences and media.

2.2.1 Analytical Conversations vs. Computational Notebooks. Computational notebooks and analytical conversations share various high-level similarities; both interleave code, narrative text, and visual outputs, and serve as records of exploratory analysis [28, 76]. However, their structure, context, and interaction patterns differ in ways that affect how users navigate the analytical process and communicate findings.

First, notebooks are organized around cells that the analyst can re-order, delete, or rewrite. This feature enables notebooks to be a *curated* representation of work: analysts often clean up or linearize their exploratory process before sharing [45, 76]. In contrast, analytical conversations are organized around *turns* in a dialogue, with strict temporal turn-taking between the user and the AI chat interface. False starts, dead ends, and incremental refinements persist in the transcript and cannot be easily reordered without losing conversational coherence. As a result, while analytical conversations tend to more faithfully preserve the exploratory nature of analysis, their fidelity can obscure the underlying reasoning structure, complicating efforts to summarize or communicate findings for external stakeholders.

Second, computational notebooks typically assume a single human author who is responsible for both writing code and providing explanatory text, even when large language models (LLMs) are used to generate code snippets that are subsequently pasted into cells [103]. In contrast, analytical conversations introduce a qualitatively different form of co-authorship, in which the AI not only generates code but also executes it, suggests lines of analysis, and provides narrative interpretations inline. This blurs the boundary between user intent, system action, and explanation, and means that key analytical decisions are distributed across many turns, rather than encapsulated in a small number of curated cells. Existing notebook tools for documentation, slide generation, and audience-aware communication [98, 101, 102] operate on the assumption that the analytic record is already structured into cells; they do not help users recover structure from raw conversational transcripts where that assumption does not hold.

Third, interaction patterns differ. Notebooks support *direct manipulation* of code and artifacts: users execute cells, inspect outputs, and annotate results in place. By contrast, analytical conversations rely on *natural language* turn-taking to steer analysis; users describe what they want, and the AI responds with code, tables, and visualizations [25, 31]. This indirection makes it harder to specify precise analytical operations [41, 107] and, when revisiting a conversation, harder to see how particular artifacts relate to evolving goals. Our findings around orienting, recalling, and prioritizing (§7.2) highlight that users face distinct information needs when revisiting conversational logs that are not addressed by notebook execution histories or provenance tools alone [95, 106].

While prior work improves how analysts document, refactor, and communicate notebooks once the analysis has been curated into cells [97, 98, 102], our work studies how data workers revisit and summarize *analytical conversations before* any such curation may exist: when the primary record of work is a long, entangled chat with an LLM. Our design probe and empirical study focus on this conversational substrate, examining how structure and dynamic abstraction can help users reconstruct, navigate, and communicate analyses directly from conversational histories, rather than assuming a notebook-style representation is already in place.

2.3 Structure and Dynamic Abstraction Support Navigation and Sensemaking

Sensemaking involves not only accessing information, but also the ability to organize and abstract it to reveal meaning and support decision-making [66]. Research across domains in HCI has consistently shown that introducing structure, whether system-defined or user-defined, significantly enhances users' ability to make sense of complex content [3, 19, 67, 88]. In analytical workflows, this need is well-documented in computational notebooks, where nonlinear exploration and fragmented reasoning can obscure analytical intent [28, 43, 44, 76, 95, 106].

Critical to sensemaking is the ability to dynamically abstract information at multiple levels of granularity [87, 109, 111]. Rule et al. [76] demonstrated how visual hierarchy and structural cues support both comprehension and collaborative review. Recent work has begun exploring structure in analytical conversations specifically.

InsightLens [105] introduced structured representations around insights extracted from analytical conversations, though it remained focused on isolated insight units. DS-Agent [27] demonstrated how structuring analytical workflows can scaffold LLM performance, suggesting benefits for both human comprehension and automated support. Our work builds on this research by examining how users engage with the broader conversational structure, including speech acts, conversation turns, and analytical threads, and how such structure can support user-driven sensemaking and communication goals.

3 Structured Elements of Analytical Conversations

To ground our investigation, we examine the structure of analytical conversations to identify where users may benefit from support. Drawing from discourse analysis, conversational systems, and data analysis workflows, we consolidate elements that reveal the layered, heterogeneous nature of these conversations. Specifically, we adapted components such as speech acts (e.g., inform, request, justify) from discourse and pragmatic analysis to capture the functional intent of utterances; topic threads and turn-taking structures from conversation analysis to represent shifts in focus or sub-goals; and insights and artifacts (e.g., charts, tables, filters) from prior work on analytic reasoning and notebook use to highlight the products and intermediate results of analysis [80, 96]. We then reviewed these taxonomies and merged overlapping constructs into a coherent schema focused on revisitation and communication. Rather than aiming for exhaustiveness, our goal was to provide a practical scaffold that helps participants interpret, organize, and act on prior conversational content.

Turns. A *turn* is the basic unit of an analytical conversation, i.e., a user query and a conversational interface response [39]. An analytical conversation is a sequence of *turn-taking* between the user and the conversational interface, often stretching across dozens of back-and-forth exchanges.

Threads. Across these turns, *analysis threads* group related turns around a shared analytical goal or topic. These can nest within one another and often reflect the iterative, non-linear nature of data exploration [24, 45, 48, 49, 70].

Speech Acts. User utterances take different functional roles, or *speech acts* (i.e., utterances that perform an action by the user [4, 77]). In analytical conversations, there can be a wide range of analysis-specific *speech acts* from "fact-finding" to seeking "specific visualizations" to "debugging" [80].

Analysis Artifacts. Many responses include structured outputs (i.e., visualizations, data tables, and code) separate from natural language text that the analytical conversational interface generates. These *artifacts* embody critical steps in the analysis, but are intermixed with long passages of text and can be difficult to locate later.

Analysis Insights. Finally, conversations surface *data insights* that represent critical outcomes of an analysis [11, 15, 42, 105]. In analytical conversations, prior work underscored the challenge of uncovering insights, as they are often buried within back-and-forth exchanges, lengthy text, and embedded artifacts [105].

4 Study Goals and Design Rationale

Our study explores four research questions:

- **RQ1** What are the characteristics of analytical conversations?
- **RQ2** How data workers' information needs when revisiting analytical conversations shape how they re-enter and make sense of prior analyses
- **RQ3** How data workers navigate and restructure analytical conversations to satisfy these needs, and what tensions and opportunities these workflows expose for tooling
- **RQ4** How data workers perceive the role of AI in structuring and communicating analytical conversations, and what forms of human–AI collaboration they prefer

These research questions intentionally couple navigation (i.e., revisiting past analyses) with communication (i.e., summarizing for audiences) because of their overlapping demands, situated relevance, and theoretical interdependence. Both tasks require similar affordances (e.g., filtering, abstraction, content selection), suggesting shared underlying infrastructure. From a practical perspective, revisitation and communication frequently co-occur in real-world settings; analysts often return to prior conversations in order to prepare presentations or brief stakeholders, making it artificial to study navigation in isolation [56, 65, 68]. Moreover, theoretical frameworks such as sensemaking models [70] and prior research on computational notebooks [45, 76, 99] emphasize that understanding and explanation are not discrete steps, but are rather iterative and mutually reinforcing processes. By examining these activities together, we can better understand how the structured elements support both personal sensemaking and external communication, while also revealing where user needs diverge.

Standard conversational interfaces, limited to linear scrolling and keyword search, constrain users' ability to revisit, structure, or communicate analytical conversations. To investigate how data workers engage with these conversations, we developed a high-fidelity design probe that augments a conventional interface with structured navigation and summary authoring features. Rather than serving as a solution, this probe functions as a research instrument that structures elements of an analytical conversation (**RQ1**) and elicits diverse usage behaviors. Specifically, the probe enables us to investigate data workers' information needs (**RQ2**), workflow strategies (**RQ3**), and preferred roles of AI (**RQ4**) in analytical conversations.

4.1 Design Considerations for the Probe

Our research probe serves as shared infrastructure for both navigation and communication tasks. When designing SyncSense, we incorporate affordances drawn from established analytical tools like Jupyter notebooks, spreadsheets, and code repositories, but these features remain unexplored in the context of analytical conversations. The following design considerations prioritize flexible interaction methods that expand beyond scrolling and keyword search without prescribing specific usage patterns, enabling data workers to naturally adopt their preferred approaches for both personal reorientation and external communication. This approach allows us to observe where navigation and communication share needs (RQ2, RQ3) and where they diverge.

DC1: Layer structured elements over raw conversations. Prior work demonstrates that structured representations of unstructured discussions (e.g., forums, clinical notes) and LLM outputs (e.g., hierarchical views, structured fields) reveal different user behaviors and navigation patterns compared to raw logs or flat summaries [36, 87, 89, 111]. Our probe should layer structured representations of analytical conversation elements (§3) over the original conversation transcript, while preserving full access to the raw conversation. This design allows data workers to toggle between structured and unstructured views, enabling us to observe how they engage with different forms of representation for navigation, sensemaking, and communication.

DC2: Support overview and detail on-demand. To manage long, heterogeneous conversations, data workers need flexible control over information granularity. Prior work has demonstrated that overview structures help users re-orient to the conversation's scope [82, 87], filters surface relevant content and reduce cognitive load [5, 29], and progressive disclosure reduces overload by revealing details incrementally [20]. Our probe should integrate these affordances to let data workers fluidly shift between high-level overviews and deeper exploration, allowing us to observe diverse strategies for managing conversational complexity.

DC3: Support modular composition from conversation elements. Modularity and reuse support flexible re-composition for varied goals [53, 57, 81, 114]. The probe should enable participants to select and recombine structured conversation elements into new compositions. This interaction will enable us to observe what users prioritize, how they filter and organize content, and which strategies they use to communicate insights to varying audiences.

5 Structured Interface for Revisiting Conversations

We introduce SyncSense, a design probe that augments analytical conversations with structured elements to support revisitation and summarization. Rather than offering a fully featured or optimized product, SyncSense intentionally exposes multiple alternative representations of an analytical conversation, i.e., threads, artifacts, insights, and authored summaries within a four-panel interface. By surfacing content at different levels of abstraction and placing revisitation and communication side-by-side, the probe is designed to provoke reflection and elicit diverse strategies for orienting, navigating, and shaping analytical narratives. Its purpose is not to prescribe an ideal workflow, but to create conditions in which participants reveal their information needs, breakdowns, and preferred patterns of interaction that would otherwise remain hidden in conventional chat interfaces or more polished prototypes.

5.1 Adding Structure to an Analytical Conversation

To surface analytical conversation elements across the interface (**DC1**), SyncSense extracts components identified in §3 and displays them using icons and thematically colored tags. These visual cues help foreground latent structures in the conversation to elicit a range of user behaviors and reflections.

Threads. Threads and nested sub-threads (where a thread can be nested within a larger thread) offer natural levels of abstraction to manage growing conversation complexity [31]. As analytical conversations unfold, they often involve topic shifts, branching inquiries, exploratory detours, and iterative refinements; properties that often challenge linear models of interaction. To surface this non-linearity, SyncSense automatically groups related (potentially non-consecutive) turns into threads and organizes them into higher-level parent threads. This feature was introduced to probe how participants perceive, interpret, and traverse complex conversation structures.

Speech Acts. To contextualize and organize conversations, SyncSense categorizes user prompts into analysis-specific speech acts. This tagging helps surface functional distinctions between turns, such as data requests, follow-ups, visualizations, or refinement prompts, making it easier for users to locate relevant parts of the conversation. Building on Setlur et al.'s taxonomy [80], we extend the categorization to better capture the broader range of low-level and iterative interactions that commonly arise in LLM-assisted workflows [26, 41, 74]. Table 1 shows the speech act categories classified in SyncSense. By revealing the functional intent of utterances through speech act tags, the probe invites participants to interpret not just the content but the purpose behind each turn. This feature helps elicit insights into how users mentally organize their analytical process, whether they favor goal-driven exploration, iterative refinement, or opportunistic branching.

Speech Act Type	Definition and Example
Fact Finding	Expecting a single response like a number, yes/no etc. (e.g., *"Did any passengers survive?"*)
Deep Insights	Seeking deeper insights (e.g., *"How much more likely was survival in first class?"*)
Specific Visualization	Requesting a specific type of visualization (e.g., *"Plot a bar chart of passengers by class."*)
Data Transformations	Asking for data manipulations such as filtering, sorting, or aggregating (e.g., *"Can you filter to only show first class passengers?"*)
Refinement	Refining a previous query or asking for more detail (e.g., *"Show survival rates by age group."*)
Recommendation	Requesting recommendations or predictions based on the data (e.g., *"What is the best course of action for the airline?"*)
Domain Knowledge	Asking about domain knowledge not directly present in the dataset (e.g., *"What does first class mean in this context?"*)
Debugging	Debugging code or visuals (e.g., *"Why are there very few counts for first class passengers."*)

Table 1: Speech act types, definitions, and examples.

Analysis Artifacts. SyncSense distinguishes between analysis artifacts (i.e., Visualizations , Data Tables , and Code). *Visualizations* and *data tables* serve as essential elements in the presentation of data analyses [102] while *code and execution outputs* support provenance and verification of how visualizations and tables were derived [26, 107]. Within the design probe, making these

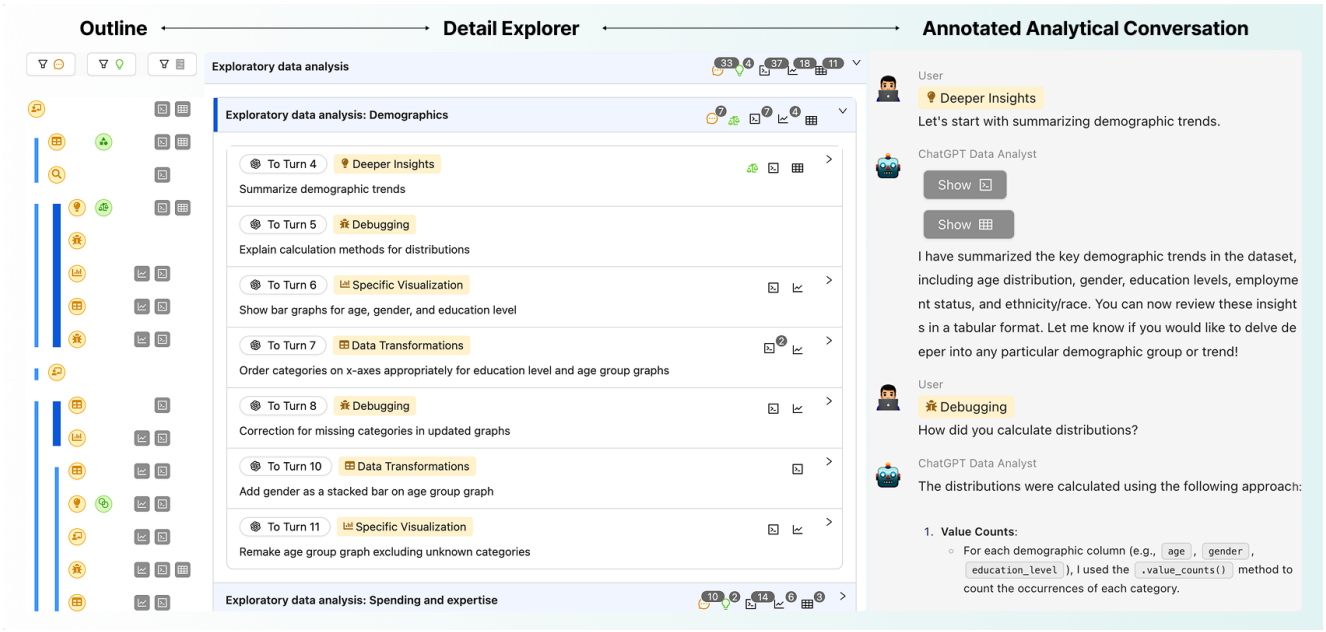


Figure 1: SyncSense extracts and presents elements of an analytical conversation at varying levels of abstraction. The Outline panel provides a high-level overview of the conversation elements, the Detail Explorer panel displays additional details about each element, and the Annotated Conversation panel shows the raw conversation annotated with each component. These three synchronized panels are connected visually and interactively through the structured elements of an analytical conversation (§5.1).

artifacts explicitly visible and filterable allows us to investigate how users revisit and prioritize different outputs depending on their goals (e.g., reporting vs. verification). This design choice lets us observe whether users gravitate toward concrete outputs when orienting themselves in long conversations, and how these anchors shape summarization workflows.

Analysis Insights. SyncSense captures analytical insights as structured elements, linking them with one or more conversation turns. Insights are categorized using the taxonomy proposed by Wang et al. [104] (Table 2), which was derived from analyzing 793 insights across 245 fact sheets focused on tabular data. Structuring insights in this way invites participants to reflect on the nature and diversity of their findings and helps supports abstraction at the right level of granularity.

5.2 Structured Conversation Content Panels

To support overviews and detail-on-demand (**DC2**), SyncSense organizes conversation content across three synchronized panels, arranged from left to right (Fig. 1). These panels are linked through structured analytical conversation elements (§5.1), using consistent iconography and color coding for visual coherence. The Outline panel offers a high-level overview of the conversation elements (§5.2.1), the Detail Explorer panel displays mid-level detail for selected elements (§5.2.2), and the Annotated Conversation panel shows the raw conversation, annotated with speech acts and artifacts for each turn (§5.2.3).

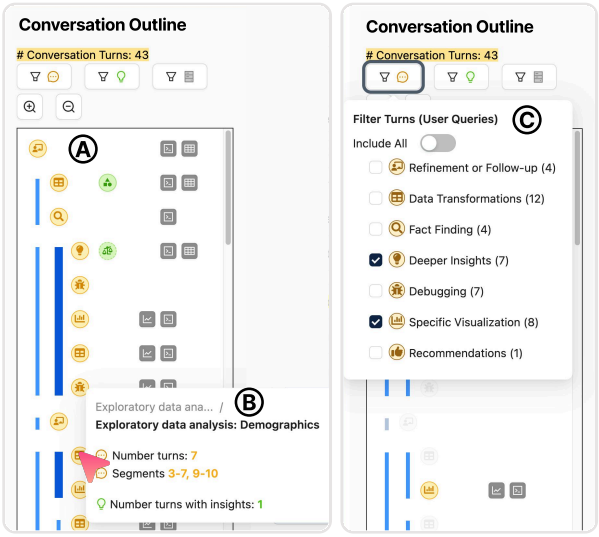


Figure 2: The Outline panel offers a visual overview of the conversation by displaying components (i.e., threads, speech acts, insights, and artifacts) as vertically arranged icons following the dialogue turns (A). Hovering over icons reveals additional details (B). Users can filter by speech acts, insights, and artifacts which update the outline view and the contents in the Detail Explorer (C).

Insight Type	Definition and Example
Value	Retrieve the exact value of data attribute(s) under specific criteria (e.g., *"46 horses have won two out of three Triple Crown races."*)
Proportion	Measure the proportion of selected data attribute(s) within a set (e.g., *"Protein takes 66% in the diet."*)
Difference	Compare two or more data attributes (e.g., *"There are more blocked beds in the Royal London Hospital compared with the UK average."*)
Distribution	Show how values are shared across or broken down by data attributes (e.g., *"The distribution of the unicorn companies is approximately normal over their age."*)
Trend	Present a general tendency over a time segment (e.g., *"The Border Patrol budget rose from 1990 to 2013."*)
Rank	Sort data attributes by their values and show rankings (e.g., *"The top reason for consumers to engage in showrooming is 'the price is better online'."*)
Aggregation	Calculate descriptive statistics (e.g., average, sum, count) (e.g., *"The average price for gas is $4.06."*)
Association	Identify relationships between data attributes (e.g., *"There is a negative correlation between the number of food vendors and their distance from the market."*)
Extreme	Identify extreme data cases within a certain range or along attributes (e.g., *"The character with the most epigrams is Oscar Wilde himself, with 12."*)
Categorization	Select data attribute(s) that satisfy specific conditions (e.g., *"Josh and Sam are two popular names in 2004."*)
Outlier	Explore unexpected or statistically deviant data points (e.g., *"'Lucy' has the highest word count."*)

Table 2: Insight types, their definitions, and examples.

5.2.1 Outline Panel. The Outline panel offers a visual scaffolding of the entire conversation by aligning structured elements along a vertical sequence that mirrors the original flow (Fig. 2A). Each conversation turn appears as a row, with associated elements (i.e., speech acts, artifacts, and insights) presented horizontally in the row corresponding to that turn. Threads are indicated in blue as vertical bars spanning relevant turns, with nested subthreads indented to the right of their parent thread. Hovering over the elements reveals contextual details (Fig. 2B) such as turn information or thumbnails of data tables and visualizations. Clicking an icon displays the corresponding element in the Detail Explorer panel, allowing users to navigate directly to the relevant content for more details and trace back to the original conversation. Artifact icons also open a popover showing artifacts of that type for the selected turn. The popover includes buttons for quick navigation to the related content in both the Detail Explorer and Annotated Conversation panels. Users can filter to specific elements, i.e., the speech acts, insights, and artifacts, using the controls at the top of the panel (Fig. 2C). Filters apply across both the Outline and Detail Explorer panels (§5.2.2), updating the visible icons accordingly. Observing users' interactions with this structured overview helps elicit design-relevant behaviors, such as the value of summarizing via structure, the role of threads in organizing workflows, or how compact representations affect information prioritization.

5.2.2 Detail Explorer Panel. The Detail Explorer panel provides a mid-level abstraction of the conversation; more detailed than the Outline panel but more concise than the raw chat conversation (Fig. 3A). Turns and their corresponding components are grouped under threads (Fig. 3A1), with nested elements expandable via the dropdown (">") button. To help users get a sense of the contents nested within a content block, icons representing the content (i.e., turns, artifacts, or insights) with the counts of that content are presented (e.g., in Fig. 3A2). To support orientation, hovering over the contents in the Detail Explorer panel triggers an increase in the size of the corresponding icon in the Outline. For detail on demand, users can click the To Turn button, which opens the Annotated Conversation and scrolls to the corresponding turn (Fig. 3A3). This panel was designed to elicit user strategies for managing analytical complexity, e.g., how they interpret thread groupings, which content types they prioritize, and how they trace reasoning across nested structures.

5.2.3 Annotated Conversation Panel. To preserve provenance and provide access to full conversational detail, SyncSense includes the original raw analytical conversation in the Annotated Conversation panel. Each turn is annotated with speech act tags and artifact icons to aid browsing (Fig. 3B) **(DC1)**.

5.3 Summary Authoring Panels

To support flexible composition from conversation elements (**DC3**), SyncSense includes an Authoring Panel (Fig. 4). The panel allows users to drag and drop elements into a workspace, organize them into summaries, and optionally, refine the text with LLM assistance (**DC4**). This interaction supports reflection on what content users value, how they organize it, and how much control they desire.

5.3.1 Composing Summary Content. Users can drag any chat element (e.g., thread, turn, insight, or artifact) from the Detail Explorer panel or Annotated Conversation into the composition container (Fig. 4A). Dropped elements retain their nested structure, with each appearing as a block and sub-elements indented accordingly. Users can reorder blocks, edit their text directly, or insert custom notes using freeform blocks. Once users are satisfied with the content, they can click the *"Add Contents to Editor"* button (Fig. 4B). This action serializes the chat contents into markdown format. Each block becomes a bullet point, with nested elements rendered as indented bullets. For some chat elements, metadata tags indicate the type of content (e.g., an insight will be prefixed with "[Insight]").

5.3.2 Editing and Refining Summaries. SyncSense provides multiple levels of AI assistance for summary composition (**DC4**). Users can directly edit markdown text in the editor (Fig. 4C), with code and tables linked as raw files and visualizations embedded inline. To refine the summary, users adjust sliders for length, technical detail, and formality. These settings reflect both structural (length) and contextual (audience and purpose) needs [38, 93]. When the *"Generate Summary"* button is clicked, a refined summary (generated by an LLM based on these settings) is written into the editor.

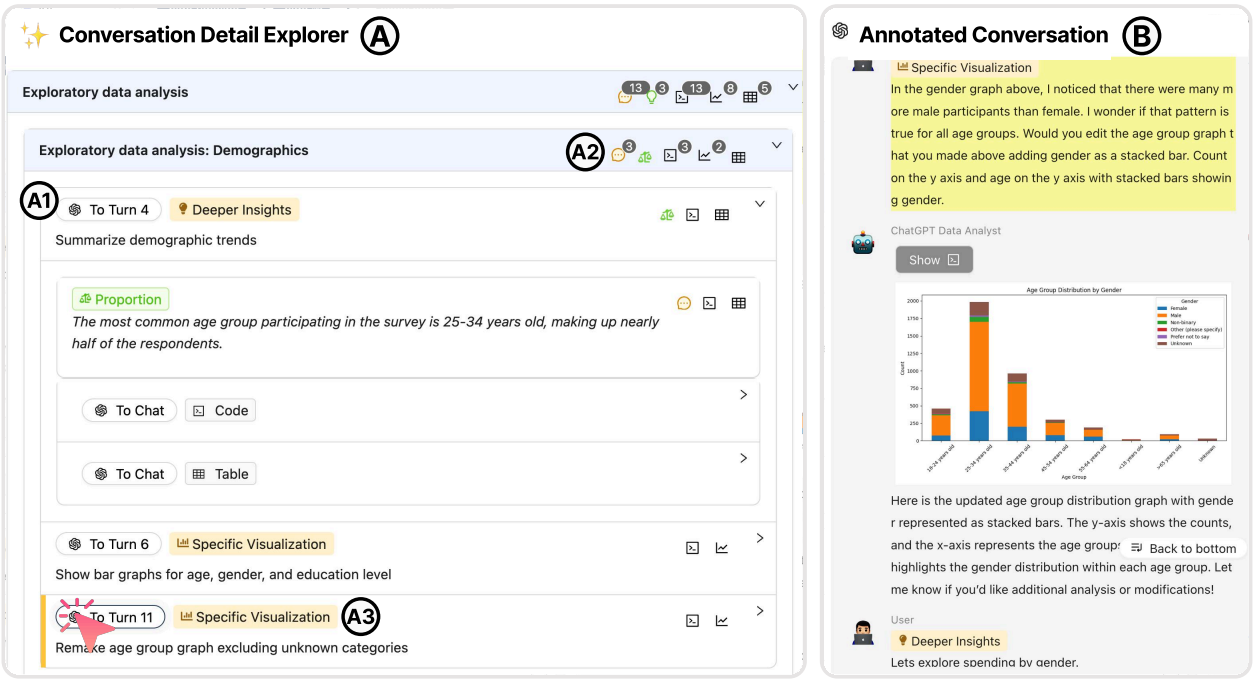


Figure 3: Detail Explorer and Annotated Conversation panels. The Detail Explorer panel presents a structured, mid-level view of the conversation (A). Threads can be expanded to reveal individual turns and elements (A1). Icons with content counts summarize nested elements (A2). Clicking the To Turn button navigates to the corresponding turn in the Annotated Conversation panel (A3). The Annotated Conversation panel displays the full raw conversation, annotated with speech act tags and artifact icons to support provenance and traceability (B).

This probe feature elicits preferences around AI-assisted communication, helping us understand where users welcome automation and where they seek fine-grained control.

5.4 Implementation Details

The SyncSense design probe is implemented in the React [85] framework based on Typescript using the Ant Design library [92] for UI components. The backend is implemented in Python using FastAPI [73], which is designed to extract the chat contents from ChatGPT's conversation JSON object, execute the code within the conversation, and generate summaries. To ensure consistent artifact rendering, code execution occurs in a sandboxed Python environment that mirrors OpenAI's code interpreter setup, using the same libraries inferred via prompting ChatGPT[1]. For all LLM-based functionalities, we use OpenAI's GPT-4o model [63]. Specifically, these tasks include (1) extracting threads from the conversation, (2) extracting the speech act from each user query, (3) extracting and categorizing insights from the conversation, and (4) generating the updated summary given the markdown input. Prompts were iteratively refined and tested using two hand-labeled conversation transcripts and grounded in definitions from prior work (i.e., speech acts, insight types). We include all prompts in our supplementary materials.

[1]OpenAI's shared chat currently does not include generated visualizations.

6 User Study

We conducted a user study that combined the use of SyncSense as a design probe [7] with task-based, semi-structured interviews. This approach enabled us to observe realistic user behavior while eliciting participant reflections on our research questions (§4).

6.1 Participants

We recruited 10 participants (six male, four female) via mailing lists at a data analytics software company. This sample size aligns with HCI norms for design probe studies [10, 30, 72, 108] . Participants reported an average of 11 years ($\sigma = 6.4$) of data analysis experience. All participants engaged in data analysis at least monthly, with six doing so daily, typically using data visualization tools (e.g., Tableau, Power BI) in their workflows. With the exception of one participant, all reported using conversational AI tools (e.g., ChatGPT, Claude, Perplexity) at least a few times per month.

6.2 Procedure & Task

To balance realistic usage scenarios where data workers may need to revisit and summarize past analyses, the study was conducted in two sessions: (1) an open-ended analysis session with ChatGPT Data Analyst, and (2) a summary authoring session held more than a week later, where participants revisited their prior conversation and used SyncSense to create summaries for different audiences.

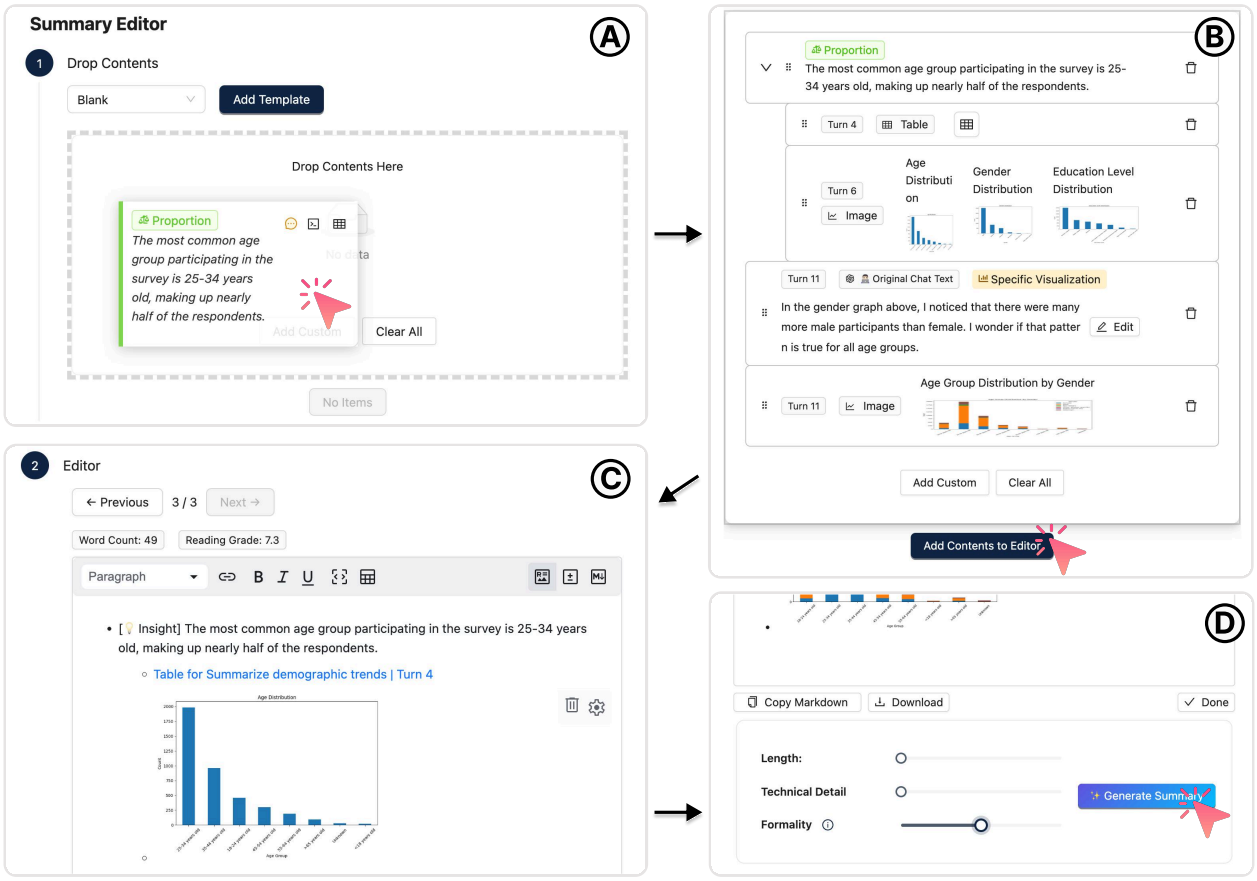


Figure 4: Summary Authoring Process in the Authoring Panel. Users can drag items from the Detail Explorer or Annotated Conversation panels into the drop container within the Authoring panel (A). Within the drop container, the items can be rearranged by dragging to adjust the order or nesting structure (B). The contents are serialized in the post-drop editor, where users can utilize an LLM to reformat the summary based on specified parameters such as *length*, *technical detail*, and *formality* (C). Finally, the LLM-generated output is displayed in the post-LLM editor so users can make any final edits.

This delayed two-part design enabled us to observe behaviors relevant to both long-term revisitation and communicative reframing, such as delayed summarization and tailoring outputs for different stakeholders.

6.2.1 Part 1: Data Analysis Session. Participants analyzed the *Great American Coffee Taste Test* dataset from TidyTuesday [35], a survey of 4,042 respondents' coffee preferences across 57 columns. The dataset combines categorical, numerical, and free-text variables (e.g., brewing methods, taste ratings, spending habits, demographics). The dataset presents realistic challenges, such as over 10 columns with >80% missing values, primarily from optional free responses. We selected this dataset because it is broadly relatable while still complex enough to support meaningful analysis. Participants were provided the following scenario: *"You and your business partner, both trained data analysts, are planning to open a coffee shop in your city. To make informed decisions, you'll analyze data to guide key strategies, such as which coffee to source, pricing, and target customer demographics. Your goal is to uncover actionable insights to shape your business plan. To assist with your analysis, you'll use ChatGPT Data Analyst as your analytical tool."*

To balance ecological validity with experimental control, all participants worked with the same dataset. They were instructed to conduct a well-rounded analysis by uncovering multiple insights, engaging with varied data types (ordinal, categorical, continuous, and unstructured text), and producing different artifacts (e.g., visualizations and tables). The task was designed to take approximately 45 minutes. Rather than priming participants with a detailed prompt, we framed the activity as a business decision-making task, mirroring the goal-oriented style of Jeopardy-style evaluations [22].

6.2.2 Part 2: Revisiting the Analytical Conversation. The second session was conducted remotely, with participants sharing their screen over video conferencing (Google Meet) on their own computers. This 90-minute session included: (i) a tutorial walkthrough of SyncSense (~20 minutes), (ii) two recall and summarization tasks (~50 minutes), and (iii) a semi-structured interview (~20 minutes). Tutorial materials are included in the supplementary material.

Participants completed two summarization tasks, based on their prior analytical conversation. For each task, we asked participants to craft a summary tailored to a specific audience and output medium. Audience options were either an *investor*, who is non-technical, seeking a persuasive business case, or an *analyst business partner*, who is technical and potentially continuing the project. The output format was either a detailed *data report* or a *direct message*, with both audience and medium counterbalanced across participants. During the tasks, participants were encouraged to think aloud and explain their interactions with the probe. After each task, they reflected on their content selection and emphasis decisions. A concluding semi-structured interview explored their information needs, summarization strategies, and perspectives on the role of AI in supporting these activities.

6.3 Analysis

To study the characteristics of an analytical conversation (**RQ1**), we examined the raw conversation data exported from ChatGPT [64] for all participants and recorded the speech acts and insights extracted by our backend LLM modules. For summary-related patterns, we examined elements added to the Authoring panel and analyzed the final authored summaries. To identify recurring themes, we reviewed the transcripts and video of the participants' screens while conducting the tasks. We performed iterative coding on participants' actions, the intent behind their actions, think-aloud comments, and interview responses guided by (**RQ2-4**). Three researchers independently open-coded one participant session, discussed findings, and created an initial code set. They repeated this for a second session to refine consistency and finalize the codebook. Two researchers then independently applied the final codebook to all sessions. The codebook is available in the supplementary material. Findings in §7 are organized by themes derived from this coding process.

7 Findings

We present our study findings organized around the four research questions: characteristics of analytical conversations (**RQ1**), participants' information needs (**RQ2**), their navigation and summarization workflows (**RQ3**), and their preferences on AI's role in this process (**RQ4**). For RQ2 and RQ3, we frame our findings using a codebook, with direct references to thematic codes shown **in bold**.

7.1 RQ1: What are the characteristics of analytical conversations?

While participants pursued the same analytical goal using the same dataset, their interactions with the analytical conversational interfaces were diverse, lengthy, and artifact-rich. On average, participants spent 69 minutes engaging in these analytical sessions (σ = 40 minutes). The conversations were notably long, averaging 29.5 turns per session (σ = 10.4), and heavily skewed toward assistant responses. Specifically, user queries averaged 24.4 tokens (σ = 14.2), while assistant replies averaged 319.7 tokens (σ = 177.7), excluding the artifacts. These conversations were substantially longer and more detailed than typical LLM conversations in the wild. For comparison, conversations in two large-scale real-world LLM datasets, LMSYS-Chat-1M [116] and WildChat [113], average only 2.02 and 2.54 turns, respectively.

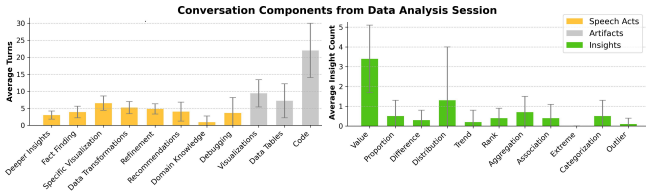


Figure 5: Analytical Conversation Elements from Session 1 (Sec. 6.2.1). Across participant conversations, all speech acts and artifacts types were observed. Except for Extreme, all types of insights appeared. The top chart shows the average number of turns each element was involved in, while the bottom chart shows the average count of unique insights per conversation.

Conversations featured a variety of artifacts: on average, 22 turns involved code (σ = 8.0), with 29.4 unique code (σ = 7.5). 7.2 turns included one or more data tables (σ = 5.0), resulting in 7.7 unique tables per conversation (σ = 7.5). Visualizations appeared in an average of 9.4 turns (σ = 4.0), with 16.3 unique visualizations generated (σ = 9.1). The most common speech acts classified by our probe involved requests for Specific Visualization, Data Transformations, Refinement, Fact Finding, and Deep Insights. Except for Extreme, all insight types also appeared in the conversations. Figure 5 summarizes the occurrences of each artifact, speech act, and insight types.

Code was distributed evenly throughout the conversation, suggesting participants consistently engaged in code execution across all stages. In contrast, data tables and insights appeared more often within the first 10% of turns. Visualizations were more frequent in the following 30% (Fig. 6).

Case Study of P7's Analytical Conversation. To help contextualize these statistics, let us consider P7's conversation[2] which contained 30 turns, 29 of which involved code, 13 with tables, and 10 with visualizations. The conversation begins with a dataset exploration (generating a data table), including a demographic overview that generated two visualizations. The participant then shifts focus to identifying which `coffee type` to buy next year, leading to a deeper investigation of `preferences` across `age` groups. This generated a visualization of `coffee preferences` by `age`. The participant spent several turns refining this visualization, before exploring additional dimensions such as `gender`, `urban` vs. `rural residence`, and `political affiliation`. Throughout, they go back to seeking high-level summaries of the dataset, such as column names and counts. This example reflects a broader pattern we observed across participants: analytical conversations are iterative, shift in focus over time, and involve a mix of exploratory and deeper engagement behaviors. We include the URLs and statistics of all conversations in our supplementary material.

[2]P7's Analytical Conversation: https://chatgpt.com/share/675c77f9-570c-800a-8d2e-a24bb7eb2df8

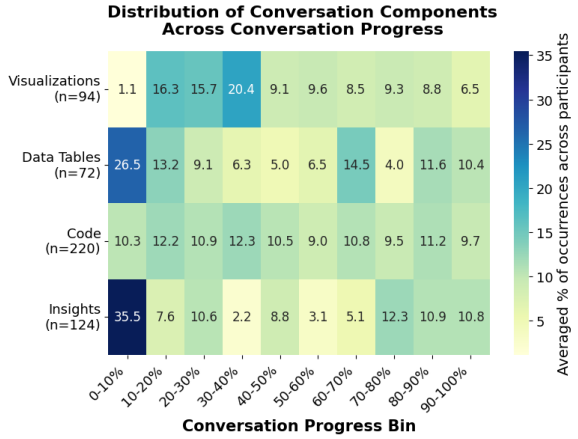


Figure 6: Data tables and insights typically appeared within the first 10% of conversation turns, while visualizations were more frequent in the following 30%. Code occurrences were distributed evenly throughout. Frequencies are normalized by turn decile within each element type (i.e., rows sum to 100%).

7.2 RQ2: How data workers' information needs when revisiting analytical conversations shape how they re-enter and make sense of prior analyses

When reviewing and summarizing their conversations, participants' information needs centered around three key objectives: *orienting* themselves to understand the structure and content of the conversation, *recalling* specific conversation elements, and *prioritizing* the information most relevant for review and summarization.

Orienting. *Orienting* refers to when a participant needs to situate themselves within the broader context of the conversation. All participants (10/10) demonstrated this need, especially when revisiting the conversation after at least a week. The purposes of orienting involved familiarizing themselves with the conversation, jogging their memory of certain elements, and identifying quickly the important main pieces of the conversation. For example, while using the Detail Explorer panel, P2 expressed needing further detail to facilitate their understanding of the conversation: *"All right, I did data cleaning. Let's see here. Coffee consumption. I think I need a little more detail to remember what these things were."* Similarly, P10, while noting that they *"remember a few of the key takeaways or insights,"* described wanting an overview that could *"help recap how things are going on."* These cases illustrate both the variability in participants' memory of prior conversations and the differing levels of detail or overview they sought to fulfill their orienting needs.

Recalling. In contrast to orienting, where participants sought a general understanding of the conversation, *recalling* refers to retrieving specific information from memory. Several participants (4/10) expressed this need [P1, P2, P3, P10], often aiming to find specific visualizations, insights, or sub-threads of analysis. For example, while searching for a specific chart they had generated, P1 stated, *"There was one with the coffee spend. That's the one I'm trying to find where I zoomed in...I need to find that one chart."* P2 was looking for a very specific conversation turn analyzing neighborhood recommendations for opening a coffee shop: *"Oh, there was my neighborhoods. There we are, the very last one."*

Prioritization. *Prioritization* refers to the need to determine which parts of the analytical conversation to focus on. This need was especially salient when summarizing conversations and working with long, complex transcripts. All participants (10/10) emphasized the importance of selectively identifying key content; both to decide what to include or exclude in their summary communication, and to support their own re-engagement with the analysis. For example, P1 chose to prioritize analysis findings over initial data-cleaning turns: *"I need to scroll down a little bit and go to where I remember actually getting something that would be appropriate for this analysis."* Similarly, P6 underscored the need to gather key insights and artifacts first, explaining: *"First important part is to [find] whatever the insights were from the conversation and the data itself, what are important key artifacts I need to look over."* Notably, prioritization often worked in tandem with orienting, helping participants better situate themselves within the conversation by narrowing their focus to what mattered most to them.

Overall, the findings of RQ2 show that revisiting analytical conversations is not a single "summarization" task, but a progression from re-establishing context, to locating specific artifacts, to curating what should be surfaced or communicated. Each stage foregrounds different cognitive work and suggests different forms of support.

7.3 RQ3: How data workers navigate and restructure analytical conversations to satisfy these needs, and what tensions and opportunities these workflows expose for tooling

7.3.1 Navigation Strategies. To satisfy these information needs, participants employed a range of navigation strategies which served the foraging phase of the sensemaking process [70].

Sequential navigation. *Sequential navigation* refers to participants scanning through the analytical conversation chronologically to re-familiarize themselves with the analysis. All participants (10/10) employed this strategy, but at varying levels of abstraction. Some (6/10) participants relied on the high-level overview and quick entry points from the Outline panel [P1, P2, P3, P4, P7, P10]. Others (8/10) navigated using the intermediate Detail Explorer panel [P1, P2, P3, P4, P5, P6, P10], and several (7/10) scrolled directly through the detailed Annotated Conversation [P2, P3, P4, P6, P8, P9, P10]. For example, P6 adopted a fine-grained approach, reading through the entire raw chat: *"I don't remember the conversation that I had so I actually went to the whole conversation and then I was looking at, okay, what was the points or what was the queries I made and then what was the result."* In contrast, P4 opted for a more abstract strategy, scanning the Outline by *"looking at the left-most set of blue lines"* representing the threads.

Abstractive navigation. *Abstractive navigation* refers to participants' strategy of moving between different levels of abstraction

(e.g., moving from the Outline to specific conversation turns' details in the Detail Explorer, to then zoom in or out for detail or context). This horizontal form of navigation often occurred within the broader flow of sequential navigation, enabling participants to zoom in or out as needed for additional context or detail. Participants commonly employed abstractive navigation to drill-down into specific segments or retrieve supporting context [P1, P2, P3, P4, P7, P10]. For example, P1 found a relevant table in the Detail Explorer panel and wanted greater detail in the Annotated Conversation: *"So let me look at this table and code. When I scroll to turn 10, I want to find it on the side so I can go right to the table...I can just click on this [to turn] box."* Similarly, P7 described using the Outline as an entry point to access additional detail to support their orienting: *"I really like that [Outline] view of the different steps that I took and being able to drill-in and see, like, exactly what I did."* Abstractive navigation was used to support all three types of information needs depending on the participant's intent.

Visual recall. *Visual recall* refers to the use of visual elements, such as the element tags, icons, and visualizations as cognitive cues to retrieve specific information. For orienting needs, visual recall was often combined with sequential navigation to make sense of the conversation [P1, P2, P3, P4, P5, P8]. This strategy was supported both by finding existing visualizations in the conversation, and SyncSense's structured tagging and iconography. For example, P3 emphasized the value of these features in helping them make sense of the conversation: *"I'm having to use it to navigate to remind me what we're doing now [and] if that's the case, I really like these icons."* Similarly, P8 expressed a preference for visual recall over browsing text: *"I have a very visual memory, and so, like, pictures and things can trigger more than just, like, reading text."*

Filtering. *Filtering* refers to the strategy of including or excluding content based on specific criteria. Several participants (5/10) used filters primarily to support prioritization [P2, P3, P4, P6, P7]. For example, P3 methodically filtered the conversation to only a few relevant speech acts: *"I'm not interested in recommendations. I'm not interested in fact finding...I don't need comparisons. I don't want my visualizations... Okay, I believe I've now filtered down to just what I need."* Likewise, prioritization to support communicative goals motivated P4's decision to filter: *"I'd want to show him like what stands out in terms of coffee preferences and how I want to market to those people... based on that, I can just filter the conversation to insights or recommendations."* Meanwhile, P2 filtered the conversation to just `insights` before proceeding with sequential navigation.

Overall, RQ3 reveals that users rely on a combination of structural overviews, semantic cues, and visual landmarks to forage within long analytical conversations. A key tension is between maintaining the narrative continuity of the original chat and efficiently jumping to semantically relevant segments.

7.3.2 Strategies for Communicating Analytical Conversations. To tailor their summaries for different audiences and media, participants made or expressed a desire to make a range of edits using both manual changes and LLM assistance. These edits reflected distinctive communicative goals and preferences, often involving multiple iterations of refinement. Here, we include the most prominent strategies participants employed.

Add process details. *Adding process details* involves supplementing the summary with background context or outlining specific steps taken during the analysis. Participants achieved this either by manually editing the summary (Fig. 4B and C) or by using the "*Generate Summary*" button and summary sliders (Fig.4D). This strategy was employed by 7 out of 10 participants [P1, P2, P3, P4, P5, P9, P10]. For instance, P1 clarified how missing data was handled during cleaning: *"Note on steps: The original cleaning added mean values to replace the missing values. I asked it to remove those average values. The new CSV should not have those values. Suggest double-checking the cleaned data file."*

Add action-oriented context. Beyond procedural or analytical details, participants emphasized the importance of framing summaries to support interpretation and downstream use. *Adding action-oriented context* positions the analysis in terms of its goals, relevance, and audience, helping others understand not just what the analysis shows, but why it matters and what to do with it. When aspects of this context existed in their original conversation, participants (4/10) directly selected it during composition [P2, P3, P4, P6]. For example, P4 explained a decision about including non-binary gender responses: *"A small sample size for responses by gender. So maybe that's something that should be considered as to whether we group the people who didn't put male or female."* However, such cues were rarely sufficient. Because context was typically absent from element serializations, participants usually wrote additional notes themselves [P2, P3, P4, P5, P6, P7, P9]. P4 explained that beyond the gender reasoning, they needed to *"put some of my own thoughts in there"*. Similarly, P7 revised their summary's introduction to foreground stakeholder goals: *"[For] handing this off ... here's the questions my stakeholders were asking, here's why I was working on it."* They added that such framing rarely emerges from AI interactions alone: *"That's something that I don't think would necessarily come from my chats with the AI. That would probably just have to be some human color that I put into it."*

Adjust overall length. Participants (6/10) also *adjusted the length of their summaries* either manually or by using the slider to suit their audience [P4, P5, P6, P7, P8, P10]. Participant-authored summaries averaged 245 words (σ = 172) and 134 unique terms (σ = 76), but varied widely across the 20 summaries. P4, for example, noted the need to provide a fuller picture when writing for an investor: *"You gotta hook them up front, right? But you gotta give them some level of detail to make them feel comfortable. When I regenerated the summary, [it was] very short and concise; just feels like I'm not giving them enough information."* These adjustments reflect a desire for audience-sensitive communication, where participants weighed informativeness against conciseness to align with reader expectations.

Reorder content and modify formatting. Participants sometimes *reordered summary content or adjusted its formatting* to improve logical flow, readability, or visual clarity [P2, P3, P6, P7, P10]. These structural adjustments helped draw attention to key insights or made the structure more digestible for their intended audience. For example, P6 noted a desire to emphasize important ideas through formatting: *"These are important points to me, and I would like to focus and just make it bold, make it in bullets. That type of thing."* This behavior was prevalent across participants, with 15

of the 20 authored summaries including either section headings or lists, and 11 including both, demonstrating a strong preference for well-structured and visually scannable summaries.

Together, these strategies show that summaries of analytical conversations are not merely compressions of chat logs. Instead, participants treat them as authored narratives that must balance process transparency, interpretive framing, and audience expectations.

7.4 Characteristics of Authored Summaries

The 20 authored summaries varied in length and structure but shared several features that speak to their perceived quality and usefulness (see Appendix A.2 Figure 8). On average, summaries were 245 words long (σ= 172) and contained 134 unique terms (σ= 76), indicating that participants did more than copy the raw transcript; they condensed and rephrased the original conversation into a more compact representation. Most summaries were structurally organized, where 15 of 20 included either section headings or bullet lists, and 11 included both, reflecting a strong preference for scannable, presentation-ready formats. In terms of content, the majority of summaries combined insights and artifacts: visualizations and insights appeared in over 75% of the compositions in the Authoring Panel just prior to serialization, and many participants also included conversational text to carry interpretive context. Rather than simply enumerating all artifacts, participants prioritized a small set of anchor visualizations or tables and surrounded them with interpretive text and audience-specific framing. These patterns suggest participants produced summaries that were concise, structured, and insight-centric, aligning with prior notions of useful analytic documentation and handoff materials.

7.4.1 Relationship between Navigation Strategies and Summary Outputs. While our study was not designed to isolate causal effects, we observed descriptive patterns suggesting that how participants navigated the conversation shaped the kinds of summaries they produced. Participants who engaged in abstractive navigation (i.e., moving between overview and detailed views across panels) were also those who most frequently enriched their summaries with process details and action-oriented context. Five of the six participants coded with abstractive navigation also added process notes (e.g., data cleaning decisions, caveats about missing data) to their summaries, and four added audience-oriented framing about stakeholder goals or next steps. In contrast, participants who remained mostly in the raw conversation view tended to produce more chronological overview summaries that closely followed the turn order of the chat, with less re-framing of the analysis as a coherent story for a specific audience. Similarly, summary reordering and formatting (e.g., restructuring content into sections and bullet lists) was strongly associated with more selective navigation: four of the five participants who reordered or reformatted their summaries also relied on filtering to narrow the conversation to relevant speech acts or artifacts. This pattern suggests that participants who treated navigation as an opportunity to curate the conversation, by filtering, zooming between abstraction levels, and using visual cues, also treated summarization as an opportunity to author a new, audience-appropriate narrative rather than simply condensing the transcript. These observations empirically ground our framing of navigation and communication as mutually reinforcing activities rather than separate stages.

7.5 RQ4: How data workers perceive the role of AI in structuring and communicating analytical conversations, and what forms of human–AI collaboration they prefer

Given participants' experience with SyncSense's LLM-enabled features for supporting navigation and summarization, we examine their preferences regarding the role of AI in structuring and summarizing analytical conversations. With the exception of P8, who preferred to communicate by directly sharing the original analytical conversation rather than composing a separate summary, all participants expressed a desire to remain involved in the summarization process. Here, we unpack the nuances of these preferences, focusing on how participants balanced automation with control, and where they saw value in human versus AI contributions.

Retaining editorial control in high-stakes communication. Participants (4/10) expressed the importance of maintaining control over summary content, particularly in high-stakes contexts such as presentations or external reports [P1, P2, P5, P7]. Authorship and narrative framing were closely tied to their sense of professionalism and responsibility. For example, P7 emphasized the importance of having *"tight control"* over the final wording, explaining that they would not feel comfortable sharing output that did not reflect their personal communication style. Likewise, P2 noted that the level of AI involvement should align with the stakes of the task: *"If I'm going to do this big, important presentation, it's more important that I can stand behind what I've done."*

Supporting selective and structured authoring. Participants expressed a desire to retain control over which elements were included in summaries, while still benefiting from AI-generated structure [P2, P3, P4, P6, P10]. For example, P6 emphasized the need to selectively include details that might otherwise be overlooked: *"I want to be very selective and concise… having this control will be helpful."* P10 described a workflow where AI provides a structured starting point, paired with tools to support organization and memory: *"I would want to give it everything, let it do a first pass, and then work through it... I would almost want a bookmark feature within the interface, because looking through the entire thing is a lot of work. I kind of lose track."* These reflections highlight participants' preference for a collaborative workflow, where AI accelerates organization and drafting, but users remain the ultimate curators of what gets communicated.

Using AI for acceleration, not automation. Some participants (4/10) viewed AI as useful for drafting or surfacing insights quickly, especially under time constraints [P2, P5, P6, P10]. For instance, P5 appreciated the time-saving benefits when responding to stakeholders: *"It would save me so much time… it's a really valuable tool."* Despite the need for control, participants welcomed AI as a productivity partner, augmenting rather than replacing their work per se, supporting efficiency without taking over key decision-making tasks during data analysis workflows [25, 58].

Scaffolding iterative and ongoing analyses. While participants primarily focused on summarization, several also highlighted the

need for better support during the analysis process itself. These reflections reveal a broader desire for AI to facilitate not just output generation, but also real-time structuring, prioritization, and iterative communication throughout the analytical workflow [P2, P3, P10]. For instance, P2 described wanting lightweight structuring features embedded directly in the ChatGPT interface to help manage and compartmentalize their workflow: *"It would have been nice if I could say, okay, just forget this, we're not getting anywhere. Or, pin this for later but I don't want to do it right now. I want to compartmentalize my experience as I'm doing it, rather than retrofitting the organization after the fact."* P10 similarly emphasized the need for tooling that supports the iterative *"back and forth between analysis and presentation"*, allowing the user to switch between communication and analysis *"modes."*

A consistent theme is that participants want AI that is opinionated enough to structure and accelerate their work, but deferential enough to preserve human authorship, especially when communicating outward.

8 Discussion

We situate our findings within the broader literature, highlighting design implications for future tools (§8.1), and reflecting on limitations of our study (§8.2).

RQ1: Structure of analytical conversations. RQ1 examines how analytical conversations with LLMs unfold and what their structure reveals in comparison to typical LLM chat interactions. As detailed in §7.1, these conversations are substantially longer than those observed in large-scale chat corpora [113, 116], and are densely interwoven with code, tables, visualizations, and insight statements. We observe distinct temporal patterns: early stages emphasize data tables and high-level insights, while later phases shift toward visualization-driven exploration, with code consistently present throughout. These observations address RQ1 by positioning analytical conversations as artifact-rich analytic sessions rather than brief Q&A exchanges. This complexity aligns with prior studies on the inherently iterative and fragmented nature of computational notebooks [28, 43, 76], and reifies the need for structured representations of LLM-driven analysis [87, 105]. These insights motivate a shift in perspective; chat histories should be treated as evolving analytic documents requiring dedicated navigation and structuring support, rather than as flat, linear transcripts.

Implication: Reframing analytical conversations as evolving documents. Analytical conversations with LLMs are substantially longer, more artifact-rich, and more structurally varied than typical chat interactions. Tools should therefore treat these sessions less as ephemeral chats and more as complex analytical documents that unfold over time. In particular, interfaces need to (i) surface and organize heterogeneous artifacts (code, tables, visualizations, insights) as first-class objects, (ii) recognize that different phases of the conversation emphasize different artifacts, and (iii) provide navigation and summarization support that respects this temporal structure rather than assuming a flat, turn-by-turn log.

RQ2: Information needs when revisiting conversations. RQ2 investigates how data workers' information needs shape their approach to revisiting and making sense of prior analytical conversations. While real-time LLM use often supports sensemaking through processes such as interpretation [25], validation [26, 107], and synthesis [105], our findings in §7.2 reveal that revisitation introduces a distinct set of challenges. Participants consistently cycled through three recurring needs: *orienting* themselves to the overall structure and progress of the conversation, *recalling* specific artifacts or analytical threads (e.g., a particular chart or hypothesis), and *prioritizing* segments most relevant to their current goals or audience. These needs often emerged after temporal gaps or when encountering unfamiliar conversations. This characterizes revisitation not as a one-off summarization activity, but as an evolving sensemaking process involving multiple granularities. These findings suggest concrete design directions for tools that support structured overviews, artifact-based retrieval, and audience-specific content selection.

Implication: Supporting multi-level navigation and sensemaking in analytical conversations. Participants employed a range of navigation strategies when revisiting prior analyses, from top-down scanning of overviews to targeted retrieval based on memory. This suggests that tools should support coordinated multi-level navigation, where timelines, structured contents, and raw chat views stay in sync and allow fluid zooming in and out. Visual tags and icons can serve as powerful memory cues and should be designed as intentional, semantically meaningful wayfinding devices. Finally, filters based on speech acts, artifact types, or insight categories can help users carve out task-specific slices of long conversations, supporting both analytical foraging and downstream communication.

RQ3: Workflows, navigation strategies, and selection. RQ3 examines how data workers navigate and restructure analytical conversations to address these information needs, and what tensions and opportunities their workflows reveal for tool design. As detailed in §7.3.1 and §7.3.2, participants engaged in a repertoire of navigation strategies: *sequential navigation* through the transcript, *abstractive navigation* across multiple levels of detail, *visual recall* using icons and visualizations as anchors, and *filtering* based on speech acts or artifact types. In authoring summaries, they selectively assembled structured elements, injected process- and action-oriented context, adjusted length and granularity, and reorganized content to fit audience expectations. These behaviors expose a central tension between maintaining the narrative continuity of the conversation and efficiently extracting task-specific slices for communication. RQ3 is addressed by identifying concrete workflows, both for navigation and for summary construction, that future tools should explicitly support through synchronized overview+detail views, semantically meaningful visual/structural cues, and interaction mechanisms that treat structured elements as first-class units for selection and recomposition. More broadly, these findings extend the principle of overview-and-detail [36, 82, 87] to analytical conversations, highlighting opportunities for conversation-specific affordances that strengthen visual recall and streamline sequential navigation to better support orienting, recalling, and prioritizing.

Implication: Scaffolding human-AI co-authorship in summary composition. Participants used LLM-generated summaries as starting points but invested significant effort in curating, restructuring, and annotating them. Tools should therefore support co-authored summaries that (i) surface candidate content from the conversation, (ii) make it easy to inject process notes and domain context, (iii) provide controls over length and structure, and (iv) support lightweight formatting and reordering. Rather than aiming for fully automatic reporting, tools can create better value by scaffolding human narrative work while preserving traceability back to the underlying analytical conversation.

RQ4: Role of AI in structuring and communicating analyses. RQ4 explores how data workers perceive the role of AI in structuring and communicating analytical conversations, and what forms of human–AI collaboration are preferred. As detailed in §7.5 and §7.3.2, participants consistently valued AI for its ability to scaffold and accelerate their workflows, such as by generating initial summaries, suggesting structure, and proposing language. However, they expressed a strong preference for maintaining editorial control, particularly in high-stakes contexts. Participants viewed AI as a collaborator that could offer suggestions and surface relevant content, but not as an autonomous agent responsible for shaping tone, emphasis, or final framing. These findings address RQ4 by showing that the preferred model of AI integration is one of mixed-initiative collaboration, where AI is proactive enough to support and guide, yet remains responsive to human direction, domain expertise, and communicative intent. The observations echo broader concerns in co-writing and generative systems literature about preserving authorial agency and accountability [32, 34], and reflect the limitations of LLMs in adapting to domain-specific or audience-specific rhetorical norms [25]. Unlike program synthesis tasks, where correctness is paramount [94], communication in data analysis involves interpretive judgment, audience framing, and rhetorical nuance [33, 56, 68]. Thus, fully automatic summary generation may be ill-suited for such tasks; rather, tools should emphasize controllable, contextual AI assistance that supports human authorship.

Implication: Reframing the role of AI toward assistive and accountable collaboration. Participants preferred AI as a collaborator that structures, suggests, and accelerates, rather than an autonomous agent that replaces their judgment. Tool designers should prioritize human-in-the-loop controls: mechanisms for pinning, bookmarking, and selectively including content; transparency about where summary statements come from; and workflows that support iterative refinement. In high-stakes settings, systems should foreground human authorship and accountability, positioning AI outputs as drafts or alternatives rather than final products.

8.1 Design Implications for Future Conversational Interfaces for Analyses

8.1.1 Tools should support structured navigation and selection. Our findings show that data workers relied on structured views to move between overviews and details and to prioritize elements such as charts, insights, or analysis steps (§7.3.1). Future tools should make this navigation and selection central, treating the structured elements as first-class units for interaction. Incorporating such structure can also enhance AI assistance by grounding generative outputs in relevant context, a capability increasingly adopted in commercial tools [14, 23]. Establishing a shared semantic vocabulary over this structure, e.g., "the chart comparing preferences by age," could bridge natural language queries with underlying abstractions [55]. In parallel, developing a visual language for structured elements could support visual recall and information scent [71].

8.1.2 Tools should treat communication as an ongoing, co-evolving activity. While the focus of the study was post-hoc navigation and communication, participants emphasized the need to structure, navigate, and communicate their findings during analysis, not just after the process (§7.5). This need for in-the-moment structuring and articulation of insights mirrors prior observations from computation notebook use, where data workers interleave code, results, and narrative to externalize insights as they emerge [45, 99]. To better support this evolving process, future tools should treat communication as an *integrated*, iterative part of analysis, not simply a final output stage. Real-time structuring, tagging, and pruning could help data workers surface key insights as they explore their data. Systems like *InsightLens* [105] begin to address this by enabling real-time insight extraction and navigation, but broader affordances are needed. Tools could support features such as tagging key turns, flagging important artifacts, pruning less relevant discussion, and offering real-time sharing. These affordances might include integrations with communication and presentation tools (e.g., Slack, PowerPoint) and adjacent editors or side-pane workspaces (akin to SyncSense) for drafting summaries during analysis.

8.1.3 Tools should provide controllable and contextual AI assistance. In our probe, AI was scoped to specific tasks, such as structuring the conversation or rewriting user-authored summaries. However, participants' reflections on AI involvement point to a design opportunity: enabling users to adjust *how much* and in *what ways* AI contributes to communication tasks. This need aligns with prior findings on users' varied communication goals [65, 110] and levels of comfort and trust in AI systems [25, 26, 51, 58, 94]. As agentic systems increasingly automate analytical workflows [13, 18, 60], it remains critical to ensure that users can stay meaningfully involved, especially in communicating data analyses where domain knowledge and rhetorical nuance matter. To better support user agency and accommodate diverse communication needs, future tools should offer *controllable and contextual AI assistance* in data analysis workflows. For example, systems could allow users to toggle between different levels of AI involvement, such as selecting relevant summary content, generating initial drafts and data stories [90, 103], or refining language [54]. These systems could also learn and adapt [21, 112] to individual preferences and communication styles over time.

8.1.4 The Value of Integrated Navigation and Communication. Studying navigation and communication together proved both productive and revealing. Our findings revealed substantial overlap in the underlying structures and affordances that support both tasks. Features such as filtering, multi-level abstraction, and visual recall facilitated both personal reorientation and summary authoring.

This convergence validates our theoretical premise that navigation and communication share infrastructural foundations. Our findings surfaced a directional interplay between the two: communicative intent frequently drove navigational behavior. Several participants (e.g., P2, P10) deliberately structured their conversational interactions with future communication in mind by organizing threads, tagging insights, or calling out key artifacts to facilitate later summarization. These practices reflect Pirolli and Card's model of the sensemaking loop [70], wherein synthesis (e.g., preparing a summary) reveals gaps in understanding, triggering renewed foraging and investigation. These observations suggest that navigation and communication are not sequential stages but interdependent activities. Tools should therefore support this bidirectional sensemaking loop, helping users fluidly move between exploration and explanation, using shared structural cues to guide both inquiry and expression.

8.2 Limitations and Future Work

Our study has several limitations that inform directions for future research. We conducted the study using a single dataset and a specific analysis scenario. While this controlled setup enabled consistent comparisons across participants and helped surface diverse strategies and feedback [9], the study design inherently narrows the domain of inquiry and may limit the generalizability of our findings to other datasets, analytical tasks, or organizational settings. Our scenario centered on consumer preference data and a single high-level business question; analysts in scientific, civic, or operational domains may face different constraints, stakes, and collaboration patterns. To help mitigate this limitation, we selected a real-world dataset comprising heterogeneous attribute types and data irregularities (e.g., missing values). Nonetheless, future research should extend this work across a wider range of datasets, task types, and domains to validate, challenge, and refine the findings presented here.

We employed LLMs to extract conversational elements such as insights, visualizations, and speech acts; techniques that are susceptible to known limitations of stochasticity and hallucinations. In addition, our user study primarily focused on using our design probe to elicit observations of information needs, strategies, and grounded discussion, leaving the validation of the extraction quality unaddressed. While we did not observe any adverse effects on participant comprehension or interaction with the tool, the reliability of these LLM-generated structures remains an open question. Future research should systematically evaluate the accuracy and robustness of LLM-based extraction techniques, especially for high-stakes analytical contexts.

Our study involved a relatively small sample drawn exclusively from a single organization, which may constrain the diversity of perspectives and limit the broader applicability of our findings. The participants' shared institutional context could reflect organizational norms around tools, documentation practices, and collaboration styles that are not representative of other settings. Furthermore, our participants were experienced data workers, which likely influenced their expectations around provenance, editorial control, and communication fidelity. These characteristics may not generalize to novice analysts, cross-functional stakeholders, or teams operating under different constraints or technical infrastructures. That said, the study design incorporated varied communication tasks (e.g., summarizing for an investor vs. an analyst business partner, via direct message or formal data report), allowing us to gather observations across a range of contexts. To validate and refine our design implications, future work should recruit a larger and more heterogeneous participant pool spanning diverse professional roles, experience levels, accessibility needs, and organizational environments.

The structured elements we introduced were evaluated only within the scope of our study dataset and scenario. Participants engaged with nearly all categories (excluding one insight type), but it remains unclear how comprehensive these definitions are. Additional elements, such as key decisions or action-oriented context, may also be relevant to capture. Future research should refine and validate this schema across more complex or domain-specific environments.

Finally, although our analysis identifies descriptive patterns linking navigation strategies (e.g., filtering, abstractive navigation, visual recall) with the structure of participants' summaries, the study was not designed to support causal inferences regarding the impact of navigation on summary quality. Establishing such causal relationships would require a more controlled experimental design. A natural next step is to conduct a larger-scale, systematically varied study in which navigation supports are manipulated, such as the availability of filtering tools or visual cues, and downstream outcomes are evaluated. These outcomes could include summary usefulness, perceived quality, efficiency of composition, and communicative effectiveness across different stakeholder audiences.

9 Conclusion

In this paper, we present an exploratory study on how data workers revisit, summarize, and communicate their analytical conversations with LLMs. To elicit these behaviors, we deployed a design probe that introduced affordances for on-demand overview and detail, alongside AI assistance features designed to preserve human agency during summarization. In a user study, participants used the probe to summarize their conversations for different audiences and communication media. We observed recurring information needs (i.e., orienting, recalling, and prioritizing) alongside the strategies data workers employed to navigate, distill, and communicate effectively. Our findings highlight opportunities for structured navigation and selective content composition, supporting communication as an ongoing, co-evolving activity. These observations also suggest directions for AI assistance that is controllable, communicatively aware, and better aligned with users' goals.

References

[1] 2025. IBM Watson Analytics. http://www.ibm.com/analytics/watson-analytics.
[2] 2025. Microsoft Q&A. https://powerbi.microsoft.com/en-us/documentation/powerbi-service-q-and-a.
[3] Tal August, Lucy Lu Wang, Jonathan Bragg, Marti A. Hearst, Andrew Head, et al. 2022. Paper Plain: Making Medical Research Papers Approachable to Healthcare Consumers with Natural Language Processing. *ACM Transactions on Computer-Human Interaction* 30 (2022), 1 – 38. https://doi.org/10.1145/3589955
[4] John M. Austin. 1962. How To Do Things with Words. https://doi.org/10.1093/acprof:oso/9780198245537.001.0001

[5] Marcia J. Bates. 1989. The Design of Browsing and Berrypicking Techniques for the Online Search Interface. *Online Review* 13, 5 (1989), 407–424. https://doi.org/10.1108/eb024320

[6] Charles Berret and Tamara Munzner. 2024. Iceberg Sensemaking: A Process Model for Critical Data Analysis. *IEEE Transactions on Visualization and Computer Graphics* PP (11 2024). https://doi.org/10.1109/TVCG.2024.3486613

[7] Kirsten Boehner, Janet Vertesi, Phoebe Sengers, and Paul Dourish. 2007. How HCI Interprets the Probes. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2007). https://doi.org/10.1145/1240624.1240789

[8] Chris Bopp, Ellie Harmon, and Amy Voida. 2017. Disempowered by Data: Nonprofits, Social Enterprises, and the Consequences of Data-Driven Work. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (2017). https://doi.org/10.1145/3025453.3025694

[9] Pia Borlund. 2003. The IIR Evaluation Model: A Framework for Evaluation of Interactive Information Retrieval Systems. *Information Research* 8 (2003). https://api.semanticscholar.org/CorpusID:38977080

[10] Kelly E. Caine. 2016. Local Standards for Sample Size at CHI. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (2016). https://doi.org/10.1145/2858036.2858498

[11] Yang Chen, Jing Yang, and William Ribarsky. 2009. Toward Effective Insight Management in Visual Analytics Systems. *2009 IEEE Pacific Visualization Symposium* (2009), 49–56. https://doi.org/10.1109/PACIFICVIS.2009.4906837

[12] Anamaria Crisan, Brittany Fiore-Gartland, and Melanie K. Tory. 2020. Passing the Data Baton: A Retrospective Analysis on Data Science Work and Workers. *IEEE Transactions on Visualization and Computer Graphics* 27 (2020), 1860–1870. https://doi.org/10.1109/TVCG.2020.3030340

[13] Cursor. 2025. Agent Mode: Autonomous AI Coding Agent. https://docs.cursor.com/chat/agent Accessed: 2025-04-06.

[14] Cursor. 2025. Overview: Guide to Using @ Symbols in Cursor. https://docs.cursor.com/context/@-symbols/overview Accessed: 2025-04-06.

[15] Rui Ding, Shi Han, Yong Xu, Haidong Zhang, and Dongmei Zhang. 2019. QuickInsights: Quick and Automatic Discovery of Insights from Multi-Dimensional Data. *International Conference on Management of Data* (2019). https://doi.org/10.1145/3299869.3314037

[16] Ian Drosos, Advait Sarkar, Xiaotong (Tone) Xu, Carina Negreanu, Sean Rintel, et al. 2024. "It's Like a Rubber Duck that Talks Back": Understanding Generative AI-Assisted Data Analysis Workflows through a Participatory Prompting Study. *Proceedings of the 3rd Annual Meeting of the Symposium on Human-Computer Interaction for Work* (2024). https://api.semanticscholar.org/CorpusID:270639010

[17] Ethan Fast, Binbin Chen, Julia Mendelsohn, Jonathan Bassen, and Michael S Bernstein. 2018. Iris: A Conversational Agent for Complex Tasks. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–12. https://doi.org/10.1145/3173574.3174047

[18] Jane Fine, Mahi Kolla, and Ilai Soloducho. 2025. Data Science Agent in Colab: The Future of Data Analysis with Gemini. https://developers.googleblog.com/en/data-science-agent-in-colab-with-gemini/ Accessed: 2025-03-09.

[19] Raymond Fok, Joseph Chee Chang, Tal August, Amy X. Zhang, and Daniel S. Weld. 2023. Qlarify: Recursively Expandable Abstracts for Directed Information Retrieval over Scientific Papers. https://doi.org/10.1145/3654777.3676397

[20] George W. Furnas. 1986. Generalized Fisheye Views. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 1986)*. ACM, New York, NY, USA, 16–23. https://doi.org/10.1145/22339.22342

[21] Ge Gao, Alexey Taymanov, Eduardo Salinas, Paul Mineiro, and Dipendra Misra. 2024. Aligning LLM Agents by Learning Latent Preference from User Edits. *ArXiv* abs/2404.15269 (2024). https://api.semanticscholar.org/CorpusID:269302910

[22] Tong Gao, Mira Dontcheva, Eytan Adar, Zhicheng Liu, and Karrie Karahalios. 2015. DataTone: Managing Ambiguity in Natural Language Interfaces for Data Visualization. *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology* (2015). https://doi.org/10.1145/2807442.2807478

[23] Google. 2025. NotebookLM: AI-Powered Note Taking & Research Assistant. https://notebooklm.google/ Accessed: 2025-04-06.

[24] Garrett Grolemund and Hadley Wickham. 2014. A Cognitive Interpretation of Data Analysis. *International Statistical Review* 82 (2014). https://doi.org/10.1111/insr.12028

[25] Ken Gu, Madeleine Grunde-McLaughlin, Andrew McNutt, Jeffrey Heer, and Tim Althoff. 2024. How Do Data Analysts Respond to AI Assistance? A Wizard-of-Oz Study. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 1015, 22 pages. https://doi.org/10.1145/3613904.3641891

[26] Ken Gu, Ruoxi Shang, Tim Althoff, Chenglong Wang, and Steven M. Drucker. 2024. How Do Analysts Understand and Verify AI-Assisted Data Analyses?. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 748, 22 pages. https://doi.org/10.1145/3613904.3642497

[27] Siyuan Guo, Cheng Deng, Ying Wen, Hechang Chen, Yi Chang, et al. 2024. DS-Agent: Automated Data Science by Empowering Large Language Models with Case-based Reasoning. In *Proceedings of the 41st International Conference on Machine Learning* (Vienna, Austria) *(ICML'24)*. JMLR.org, Article 668, 36 pages. https://doi.org/10.48550/arXiv.2402.17453

[28] Andrew Head, Fred Hohman, Titus Barik, Steven Mark Drucker, and Robert DeLine. 2019. Managing Messes in Computational Notebooks. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (2019). https://doi.org/10.1145/3290605.3300500

[29] Marti A. Hearst. 2006. Design Recommendations for Hierarchical Faceted Search Interfaces. In *Proceedings of the ACM SIGIR Workshop on Faceted Search*. 1–5. Online at https://flamenco.berkeley.edu/papers/faceted-workshop06.pdf.

[30] Fred Hohman, Andrew Head, Rich Caruana, Robert DeLine, and Steven M. Drucker. 2019. Gamut: A Design Probe to Understand How Data Scientists Understand Machine Learning Models. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300809

[31] Matt-Heun Hong and Anamaria Crisan. 2023. Conversational AI Threads for Visualizing Multidimensional Datasets. *ArXiv* abs/2311.05590 (2023). https://doi.org/10.48550/arXiv.2311.05590

[32] Md Naimul Hoque, Tasfia Mashiat, Bhavya Ghai, Cecilia D. Shelton, Fanny Chevalier, et al. 2024. The HaLLMark Effect: Supporting Provenance and Transparent Use of Large Language Models in Writing with Interactive Visualization. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 1045, 15 pages. https://doi.org/10.1145/3613904.3641895

[33] Jessica Hullman and Nick Diakopoulos. 2011. Visualization Rhetoric: Framing Effects in Narrative Visualization. *IEEE Transactions on Visualization and Computer Graphics* 17, 12 (2011), 2231–2240. https://doi.org/10.1109/TVCG.2011.255

[34] Angel Hsing-Chi Hwang, Q. Vera Liao, Su Lin Blodgett, Alexandra Olteanu, and Adam Trischler. 2025. 'It was 80% me, 20% AI': Seeking Authenticity in Co-Writing with Large Language Models. *Proceedings of ACM Human-Computer Interaction* 9, 2, Article CSCW122 (May 2025), 41 pages. https://doi.org/10.1145/3711020

[35] James Hoffmann, Cometeer, and Robert McKeon Aloe. 2023. The Great American Coffee Taste Test dataset. TidyTuesday project, public data repository. https://raw.githubusercontent.com/rfordatascience/tidytuesday/main/data/2024/2024-05-14/coffee_survey.csv

[36] Peiling Jiang, Jude Rayan, Steven W. Dow, and Haijun Xia. 2023. Graphologue: Exploring Large Language Model Responses with Interactive Diagrams. *ArXiv* abs/2305.11473 (2023). https://api.semanticscholar.org/CorpusID:258823121

[37] Zhiqiu Jiang, Mashrur Rashik, Kunjal Panchal, Mahmood Jasim, Ali Sarvghad, et al. 2023. CommunityBots: Creating and Evaluating A Multi-Agent Chatbot Platform for Public Input Elicitation. *Proc. ACM Human Computer Interaction* 7, CSCW, Article 36 (April 2023), 32 pages. https://doi.org/10.1145/3579469

[38] Karen Spärck Jones. 1998. Automatic Summarising: Factors and Directions. *ArXiv* cmp-lg/9805011 (1998). https://doi.org/10.48550/arXiv.cmp-lg/9805011

[39] Daniel Jurafsky and James H. Martin. 2023. Chatbots & Dialogue Systems. https://api.semanticscholar.org/CorpusID:255771485

[40] Sean Kandel, Andreas Paepcke, Joseph M. Hellerstein, and Jeffrey Heer. 2012. Enterprise Data Analysis and Visualization: An Interview Study. *IEEE Transactions on Visualization and Computer Graphics* 18 (2012), 2917–2926. https://doi.org/10.1109/TVCG.2012.219

[41] Majeed Kazemitabaar, Jack Williams, Ian Drosos, Tovi Grossman, Austin Z. Henley, et al. 2024. Improving Steering and Verification in AI-Assisted Data Analysis with Interactive Task Decomposition. https://doi.org/10.1145/3654777.3676345

[42] Daniel A. Keim. 2002. Information Visualization and Visual Data Mining. *IEEE Transactions on Visualization & Computer Graphics* 8 (2002), 1–8. https://doi.org/10.1109/2945.981847

[43] Mary Beth Kery, Amber Horvath, and Brad A. Myers. 2017. Variolite: Supporting Exploratory Programming by Data Scientists. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (2017). https://doi.org/10.1145/3025453.3025626

[44] Mary Beth Kery, Bonnie E. John, Patrick O'Flaherty, Amber Horvath, and Brad A. Myers. 2019. Towards Effective Foraging by Data Scientists to Find Past Analysis Choices. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (2019). https://doi.org/10.1145/3290605.3300322

[45] Mary Beth Kery, Marissa Radensky, Mahima Arya, Bonnie E. John, and Brad A. Myers. 2018. The Story in the Notebook: Exploratory Data Science using a Literate Programming Tool. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (2018). https://doi.org/10.1145/3173574.3173748

[46] Miryung Kim, Thomas Zimmermann, Robert DeLine, and Andrew Begel. 2016. The Emerging Role of Data Scientists on Software Development Teams. *2016 IEEE/ACM 38th International Conference on Software Engineering (ICSE)* (2016), 96–107. https://doi.org/10.1145/2884781.2884783

[47] Miryung Kim, Thomas Zimmermann, Robert DeLine, and Andrew Begel. 2018. Data Scientists in Software Teams: State of the Art and Challenges. *IEEE Transactions on Software Engineering* 44 (2018), 1024–1038. https://doi.org/10.

1109/TSE.2017.2754374
[48] Gary Klein, Jennifer K. Phillips, Erica L. Rall, and Deborah A. Peluso. 2007. A Data–Frame Theory of Sensemaking. In *Expertise out of Context: Proceedings of the Sixth International Conference on Naturalistic Decision Making*. Lawrence Erlbaum Associates Publishers, Mahwah, NJ, US, 113–155. https://doi.org/10.4324/9780203810088
[49] Laura M. Koesten, Kathleen Gregory, Paul T. Groth, and Elena Paslaru Bontas Simperl. 2019. Talking Datasets: Understanding Data Sensemaking Behaviours. *International Journal for Human Compututer Studies* 146 (2019), 102562. https://doi.org/10.1016/j.ijhcs.2020.102562
[50] Haotian Li, Yun Wang, and Huamin Qu. 2024. Where Are We So Far? Understanding Data Storytelling Tools from the Perspective of Human-AI Collaboration. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 845, 19 pages. https://doi.org/10.1145/3613904.3642726
[51] Jenny Liang, Chenyang Yang, and Brad A. Myers. 2023. Understanding the Usability of AI Programming Assistants. *ArXiv* abs/2303.17125 (2023). https://doi.org/10.48550/arXiv.2303.17125
[52] Qingzi Vera Liao and Jennifer Wortman Vaughan. 2023. AI Transparency in the Age of LLMs: A Human-Centered Research Roadmap. *ArXiv* abs/2306.01941 (2023). https://api.semanticscholar.org/CorpusID:259075521
[53] Susan Lin, Jeremy Warner, J.D. Zamfirescu-Pereira, Matthew G. Lee, Sauhard Jain, et al. 2024. Rambler: Supporting Writing With Speech via LLM-Assisted Gist Manipulation. *Proceedings of the CHI Conference on Human Factors in Computing Systems* (2024). https://doi.org/10.1145/3613904.3642217
[54] Maxim Lisnic, Vidya Setlur, and Nicole Sultanum. 2025. Plume: Scaffolding Text Composition in Dashboards. *CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.48550/arXiv.2503.07512
[55] Michael Xieyang Liu, Advait Sarkar, Carina Negreanu, Benjamin G. Zorn, J. Williams, et al. 2023. "What It Wants Me To Say": Bridging the Abstraction Gap Between End-User Programmers and Code-Generating Large Language Models. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (2023). https://doi.org/10.1145/3544548.3580817
[56] Yaoli Mao, Dakuo Wang, Michael Muller, Kush R. Varshney, Ioana Baldini, et al. 2019. How Data Scientists Work Together With Domain Experts in Scientific Collaborations: To Find The Right Answer Or To Ask The Right Question? *Proceedings of ACM Human-Computer Interaction* 3, GROUP, Article 237 (Dec. 2019), 23 pages. https://doi.org/10.1145/3361118
[57] Catherine C. Marshall and Sara Bly. 2005. Saving and Using Encountered Information: Implications for Electronic Periodicals. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Portland, Oregon, USA) *(CHI '05)*. Association for Computing Machinery, New York, NY, USA, 111–120. https://doi.org/10.1145/1054972.1054989
[58] Andrew McNutt, Chenglong Wang, Robert DeLine, and Steven Mark Drucker. 2023. On the Design of AI-powered Code Assistants for Notebooks. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (2023). https://doi.org/10.1145/3544548.3580940
[59] Microsoft. 2025. Get Started with Copilot in Excel. https://support.microsoft.com/en-us/office/get-started-with-copilot-in-excel-d7110502-0334-4b4f-a175-a73abdfc118a. Accessed: 2025-04-02.
[60] Monica. 2025. Manus AI: The General AI Agent That Bridges Thinking and Action. https://manus.im/ Accessed: 2025-04-06.
[61] Michael J. Muller, Ingrid Lange, Dakuo Wang, David Piorkowski, Jason Tsay, et al. 2019. How Data Science Workers Work with Data: Discovery, Capture, Curation, Design, Creation. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (2019). https://doi.org/10.1145/3290605.3300356
[62] OpenAI. 2024. ChatGPT Data Analyst. https://chatgpt.com/g/g-HMNcP6w7d-data-analyst. Accessed: 2025-04-08.
[63] OpenAI. 2024. Hello, GPT-4.0! https://openai.com/index/hello-gpt-4o/. Accessed: June 13, 2024.
[64] OpenAI. 2025. How Do I Export My ChatGPT History and Data? https://help.openai.com/en/articles/7260999-how-do-i-export-my-chatgpt-history-and-data. OpenAI Help Center article.
[65] Rock Yuren Pang, Ruotong Wang, Joely Nelson, and Leilani Battle. 2022. How Do Data Science Workers Communicate Intermediate Results? *2022 IEEE Visualization in Data Science (VDS)* (2022), 46–54. https://doi.org/10.1109/VDS57266.2022.00010
[66] Sharoda Paul and Meredith Morris. 2011. Sensemaking in Collaborative Web Search. *Human-Computer Interaction* 26 (01 2011). https://doi.org/10.1080/07370024.2011.559410
[67] Amy Pavel, Dan B. Goldman, Bjoern Hartmann, and Maneesh Agrawala. 2015. SceneSkim: Searching and Browsing Movies Using Synchronized Captions, Scripts and Plot Summaries. *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology* (2015). https://doi.org/10.1145/2807442.2807502
[68] David Piorkowski, Soya Park, April Yi Wang, Dakuo Wang, Michael Muller, et al. 2021. How AI Developers Overcome Communication Challenges in a Multidisciplinary Team: A Case Study. *Proceedings of ACM Human-Computer Interaction* 5, CSCW1, Article 131 (April 2021), 25 pages. https://doi.org/10.1145/3449205
[69] David Piorkowski, Soya Park, April Yi Wang, Dakuo Wang, Michael J. Muller, et al. 2021. How AI Developers Overcome Communication Challenges in a Multidisciplinary Team. *Proceedings of the ACM on Human-Computer Interaction* 5 (2021), 1 – 25. https://doi.org/10.1145/3449205
[70] Peter Pirolli and Stuart Card. 2005. The Sensemaking Process and Leverage Points for Analyst Technology as Identified Through Cognitive Task Analysis. In *Proceedings of the International Conference on Intelligence Analysis*. 2–4. McLean, VA, USA.
[71] Peter Pirolli, Stuart K. Card, and Mija M. Van Der Wege. 2003. The Effects of Information Scent on Visual Search in the Hyperbolic Tree Browser. *ACM Transactions on Computer-Human Interaction* 10, 1 (March 2003), 20–53. https://doi.org/10.1145/606658.606660
[72] Napol Rachatasumrit, Gonzalo Ramos, Jina Suh, Rachel Ng, and Christopher Meek. 2021. Forsense: Accelerating Online Research Through Sensemaking Integration and Machine Research Support. In *Proceedings of the 26th International Conference on Intelligent User Interfaces*. 608–618.
[73] Sebastian Ramírez. 2024. FastAPI. https://fastapi.tiangolo.com Accessed: 2024-09-26.
[74] Jennifer Rogers and Anamaria Crisan. 2023. Tracing and Visualizing Human-ML/AI Collaborative Processes through Artifacts of Data Work. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (2023). https://doi.org/10.1145/3544548.3580819
[75] Aayushi Roy, Deepthi Raghunandan, Niklas Elmqvist, and Leilani Battle. 2023. How I Met Your Data Science Team: A Tale of Effective Communication. In *IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC) 2023, Washington, DC, USA, October 3–6, 2023*. IEEE, 199–208. https://doi.org/10.1109/VL-HCC57772.2023.00032
[76] Adam Rule, Aurélien Tabard, and James D. Hollan. 2018. Exploration and Explanation in Computational Notebooks. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (2018). https://doi.org/10.1145/3173574.3173606
[77] John R. Searle. 1969. Speech Acts: An Essay in the Philosophy of Language. https://api.semanticscholar.org/CorpusID:147355356
[78] Vidya Setlur, Sarah E. Battersby, Melanie Tory, Rich Gossweiler, and Angel X. Chang. 2016. Eviza: A Natural Language Interface for Visual Analysis. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (Tokyo, Japan) *(UIST 2016)*. ACM, New York, NY, USA, 365–377. https://doi.org/10.1145/2984511.2984588
[79] Vidya Setlur, Melanie Tory, and Alex Djalali. 2019. Inferencing Underspecified Natural Language Utterances in Visual Analysis. In *Proceedings of the 24th International Conference on Intelligent User Interfaces* (Marina del Ray, California) *(IUI '19)*. Association for Computing Machinery, New York, NY, USA, 40–51. https://doi.org/10.1145/3301275.3302270
[80] Vidya Setlur and Melanie K. Tory. 2022. How do You Converse with an Analytical Chatbot? Revisiting Gricean Maxims for Designing Analytical Conversational Behavior. *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (2022). https://doi.org/10.1145/3491102.3501972
[81] Frank M. Shipman and Catherine C. Marshall. 1999. Formality Considered Harmful: Experiences, EmergingThemes, and Directions on the Use of Formal Representations inInteractive Systems. *Computer Supported Cooperative Work* 8, 4 (Oct. 1999), 333–352. https://doi.org/10.1023/A:1008716330212
[82] Ben Shneiderman. 1996. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. (1996), 336–343. https://doi.org/10.1109/VL.1996.545307
[83] Gabriel Skantze. 2021. Turn-taking in Conversational Systems and Human-Robot Interaction: A Review. *Computer Speech & Language* 67 (05 2021), 101178. https://doi.org/10.1016/j.csl.2020.101178
[84] Justin Smith, Justin A. Middleton, and Nicholas A. Kraft. 2017. Spreadsheet Practices and Challenges in a Large Multinational Conglomerate. In *2017 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*. 155–163. https://doi.org/10.1109/VLHCC.2017.8103463
[85] Meta Open Source. 2024. React. https://react.dev/ Accessed: 2024-09-24.
[86] Arjun Srinivasan and John Stasko. 2018. Orko: Facilitating Multimodal Interaction for Visual Exploration and Analysis of Networks. *IEEE Transactions on Visualization and Computer Graphics* 24, 1 (2018), 511–521. https://doi.org/10.1109/TVCG.2017.2745219
[87] Sangho Suh, Bryan Min, Srishti Palani, and Haijun Xia. 2023. Sensecape: Enabling Multilevel Exploration and Sensemaking with Large Language Models. *ArXiv* abs/2305.11483 (2023). https://doi.org/10.1145/3586183.3606756
[88] Nicole Sultanum, Anastasia Bezerianos, and Fanny Chevalier. 2021. Text Visualization and Close Reading for Journalism with Storifier. *2021 IEEE Visualization Conference (VIS)* (2021), 186–190. https://doi.org/10.1109/VIS49827.2021.9623264
[89] Nicole Sultanum, Devin Singh, Michael Brudno, and Fanny Chevalier. 2019. Doccurate: A Curation-Based Approach for Clinical Text Visualization. *IEEE Transactions on Visualization and Computer Graphics* 25 (2019), 142–151. https://doi.org/10.1109/TVCG.2018.2864905

[90] Nicole Sultanum and Arjun Srinivasan. 2023. DataTales: Investigating the Use of Large Language Models for Authoring Data-Driven Articles. *2023 IEEE Visualization and Visual Analytics (VIS)* (2023), 231–235. https://doi.org/10.1109/VIS54172.2023.00055

[91] Tableau. 2024. Tableau Agent. https://www.tableau.com/products/tableau-agent. Accessed: 2024-09-26.

[92] Ant Design Team. 2024. Ant Design. https://ant.design/ Accessed: 2024-09-24.

[93] Maartje ter Hoeve, Julia Kiseleva, and M. de Rijke. 2020. What Makes a Good and Useful Summary? Incorporating Users in Automatic Summarization Research. In *North American Chapter of the Association for Computational Linguistics*. https://doi.org/10.18653/v1/2022.naacl-main.4

[94] Priyan Vaithilingam, Tianyi Zhang, and Elena L. Glassman. 2022. Expectation vs. Experience: Evaluating the Usability of Code Generation Tools Powered by Large Language Models. *CHI Conference on Human Factors in Computing Systems Extended Abstracts* (2022). https://doi.org/10.1145/3491101.3519665

[95] April Yi Wang, Will Epperson, Robert DeLine, and Steven Mark Drucker. 2022. Diff in the Loop: Supporting Data Comparison in Exploratory Data Analysis. *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (2022). https://doi.org/10.1145/3491102.3502123

[96] April Yi Wang, Anant Mittal, Christopher A. Brooks, and Steve Oney. 2019. How Data Scientists Use Computational Notebooks for Real-Time Collaboration. *Proceedings of the ACM on Human-Computer Interaction* 3 (2019), 1 – 30. https://api.semanticscholar.org/CorpusID:207946488

[97] April Yi Wang, Dakuo Wang, Jaimie Drozdal, Xuye Liu, Soya Park, et al. 2021. What Makes a Well-Documented Notebook? A Case Study of Data Scientists' Documentation Practices in Kaggle. *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (2021). https://doi.org/10.1145/3411763.3451617

[98] April Yi Wang, Dakuo Wang, Jaimie Drozdal, Michael J. Muller, Soya Park, et al. 2021. Documentation Matters: Human-Centered AI System to Assist Data Science Code Documentation in Computational Notebooks. *ACM Transactions on Computer-Human Interaction* 29 (2021), 1 – 33. https://doi.org/10.1145/3489465

[99] April Yi Wang, Zihan Wu, Christopher A. Brooks, and Steve Oney. 2020. Callisto: Capturing the "Why" by Connecting Conversations with Computational Narratives. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (2020). https://doi.org/10.1145/3313831.3376740

[100] Dakuo Wang, Josh Andres, Justin D. Weisz, Erick Oduor, and Casey Dugan. 2021. AutoDS: Towards Human-Centered Automation of Data Science. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (2021). https://doi.org/10.1145/3411764.3445526

[101] Fengjie Wang, Yanna Lin, Leni Yang, Haotian Li, Mingyang Gu, et al. 2024. OutlineSpark: Igniting AI-powered Presentation Slides Creation from Computational Notebooks through Outlines. *Proceedings of the CHI Conference on Human Factors in Computing Systems* (2024). https://doi.org/10.1145/3613904.3642865

[102] Fengjie Wang, Xuye Liu, Oujing Liu, Ali Neshati, Tengfei Ma, et al. 2023. Slide4N: Creating Presentation Slides from Computational Notebooks with Human-AI Collaboration. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (2023). https://doi.org/10.1145/3544548.3580753

[103] Huichen Will Wang, Larry Birnbaum, and Vidya Setlur. 2025. Jupybara: Operationalizing a Design Space for Actionable Data Analysis and Storytelling with LLMs. *CHI Conference on Human Factors in Computing Systems* abs/2501.16661 (2025). https://doi.org/10.48550/arXiv.2501.16661

[104] Yun Wang, Zhida Sun, Haidong Zhang, Weiwei Cui, Ke Xu, et al. 2020. DataShot: Automatic Generation of Fact Sheets from Tabular Data. *IEEE Transactions on Visualization and Computer Graphics* 26 (2020), 895–905. https://doi.org/10.1109/TVCG.2019.2934398

[105] Luoxuan Weng, Xingbo Wang, Junyu Lu, Yingchaojie Feng, Yihan Liu, et al. 2024. InsightLens: Discovering and Exploring Insights from Conversational Contexts in Large-Language-Model-Powered Data Analysis. *ArXiv* abs/2404.01644 (2024). https://doi.org/10.48550/arXiv.2404.01644

[106] Yifan Wu, Joseph M. Hellerstein, and Arvind Satyanarayan. 2020. B2: Bridging Code and Interactive Visualization in Computational Notebooks. *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology* (2020). https://doi.org/10.1145/3379337.3415851

[107] Liwenhan Xie, Chengbo Zheng, Haijun Xia, Huamin Qu, and Zhu-Tian Chen. 2024. WaitGPT: Monitoring and Steering Conversational LLM Agent in Data Analysis with On-the-Fly Code Visualization. *ArXiv* abs/2408.01703 (2024). https://doi.org/10.1145/3654777.3676374

[108] J.D. Zamfirescu-Pereira, Richmond Y. Wong, Bjoern Hartmann, and Qiang Yang. 2023. Why Johnny Can't Prompt: How Non-AI Experts Try (and Fail) to Design LLM Prompts. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (2023). https://doi.org/10.1145/3544548.3581388

[109] Amy X. Zhang and Justin Cranshaw. 2018. Making Sense of Group Chat through Collaborative Tagging and Summarization. *Proceedings of the ACM on Human-Computer Interaction* 2 (2018), 1 – 27. https://doi.org/10.1145/3274465

[110] Amy X. Zhang, Michael J. Muller, and Dakuo Wang. 2020. How do Data Science Workers Collaborate? Roles, Workflows, and Tools. *Proceedings of the ACM on Human-Computer Interaction* 4 (2020), 1 – 23. https://doi.org/10.48550/arXiv.2001.06684

[111] Amy X. Zhang, Lea Verou, and David R Karger. 2017. Wikum: Bridging Discussion Forums and Wikis Using Recursive Summarization. *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (2017). https://doi.org/10.1145/2998181.2998235

[112] Siyan Zhao, Mingyi Hong, Yang Liu, Devamanyu Hazarika, and Kaixiang Lin. 2025. Do LLMs Recognize Your Preferences? Evaluating Personalized Preference Following in LLMs. In *The Thirteenth International Conference on Learning Representations*. https://doi.org/10.48550/arXiv.2502.09597

[113] Wenting Zhao, Xiang Ren, John Frederick Hessel, Claire Cardie, Yejin Choi, et al. 2024. WildChat: 1M ChatGPT Interaction Logs in the Wild. *ArXiv* abs/2405.01470 (2024). https://doi.org/10.48550/arXiv.2405.01470

[114] Chengbo Zheng, April Yi Wang, Dakuo Wang, and Xiaojuan Ma. 2022. NB2Slides: A JupyterLab Extension for AI-Assisted Presentation Slide Creation from Computational Notebooks. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, Paper 761, 13 pages. https://doi.org/10.1145/3491102.3517615

[115] Chengbo Zheng, Dakuo Wang, April Yi Wang, and Xiaojuan Ma. 2022. Telling Stories from Computational Notebooks: AI-Assisted Presentation Slides Creation for Presenting Data Science Work. *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (2022). https://doi.org/10.1145/3491102.3517615

[116] Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Tianle Li, Siyuan Zhuang, et al. 2023. LMSYS-Chat-1M: A Large-Scale Real-World LLM Conversation Dataset. *ArXiv* abs/2309.11998 (2023). https://doi.org/10.48550/arXiv.2309.11998

A Supplementary Material

The supplementary material includes the following:

(1) (**§ A.1**). Additional figures that show the backend design of SyncSense (Fig. 9), how speech acts are distributed throughout an analytical conversation (Fig. 7), and the composition of authored summaries (Fig. 8).
(2) (**§ A.2**) A table of the codebook for the qualitative findings of our study (Table 3).
(3) (**§ A.3**) Prompt templates used in the backend of SyncSense.
(4) **`study_analytical_conversations.csv`** containing the links to the original ChatGPT Data Analyst conversations along with extracted statistics for all analytical conversations in our user study.
(5) **`syncsense_tutorial_walkthrough.csv`** containing screenshots of the tutorial walkthrough used to introduce participants to SyncSense.

A.1 Additional Figures

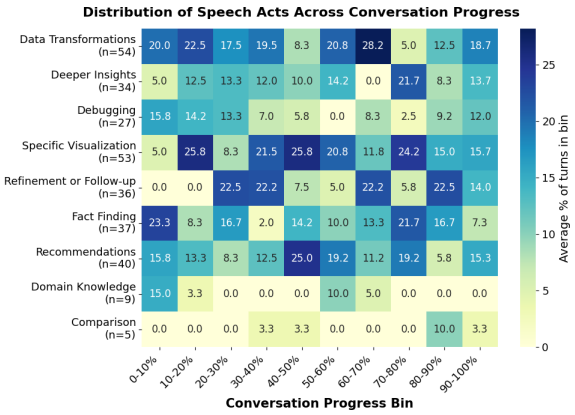


Figure 7: Frequency of each speech act at each turn decile. Each conversation is normalized by its number of turns and at each turn decile, the frequency is normalized by speech act types (i.e., columns sum to 100%). This is averaged across all participant conversations. Participants focus on data transformations, visualizations and refinement throughout the conversation, indicating the iterative nature of analytical conversations

A.2 Study Result Codes

We include the qualitative codes, definitions, and occurrences in study sessions in Table 3.

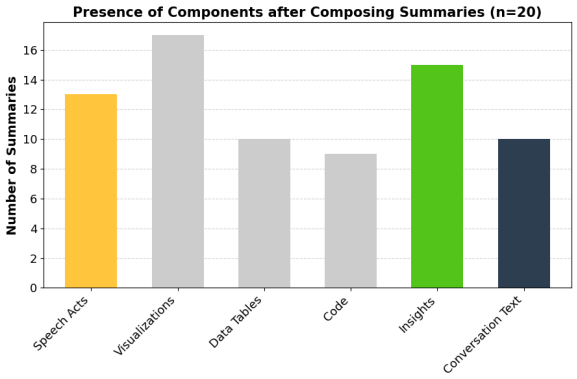


Figure 8: Presence of conversation elements in the drop container just before participants serialized their final summaries. Speech acts encompass SyncSense summary of the given turn. Bars reflect binary presence, whether a element appeared in a summary, not frequency of use. Visualizations and insights appeared in over 75% of summaries, with conversation text also commonly included for added context.

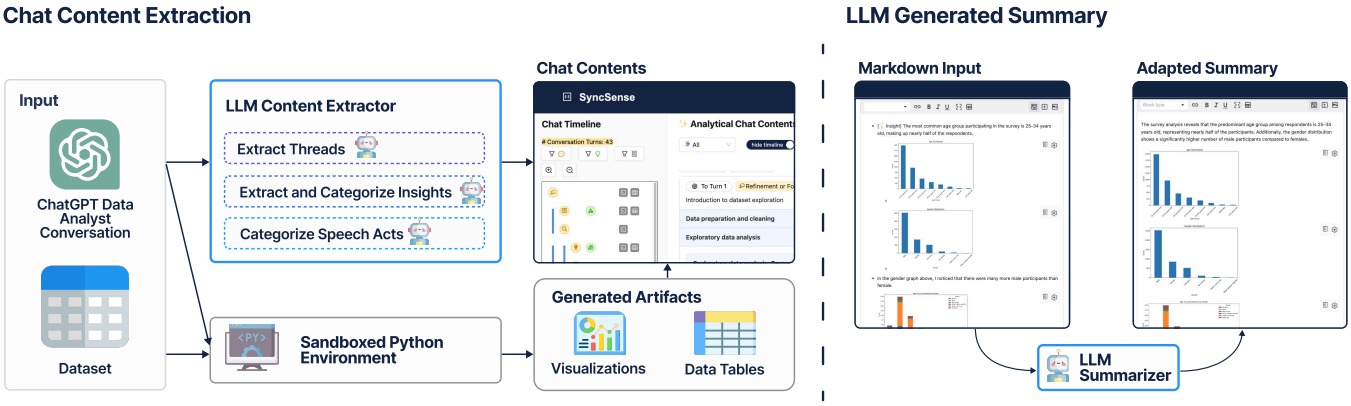


Figure 9: SyncSense's Backend and LLM modules. The chat element extraction process takes the original ChatGPT analytical conversation and dataset. The code from the conversation is executed to generate the artifacts, while three LLM modules (three separate LLM API calls) are used to extract threads, insights, and speech acts, respectively (left). The LLM-generated summary in the Authoring panel takes the markdown and user configurations in the post-drop editor as input to adapt and generates (via an API call) the final summary (right).

Section	Code	Definition	Participants
Information Needs (**RQ2**, Sec. 7.2)	Orienting	Needing to situate oneself in the conversation to understand its structure and content.	P1–P10
	Recalling	Retrieving specific information or insights from memory.	P1, P2, P3, P10
	Prioritization	Determining which parts of the conversation are most relevant for summarization.	P1–P10
Strategies for Information Needs (**RQ3**, Sec. 7.3.1)	Sequential Navigation	Exploring contents in order, either through high-level or detailed views.	P1–P10
	Abstractive Navigation	Moving between different abstraction levels (e.g., summary to detail).	P1, P2, P3, P4, P7, P10
	Visual Recall	Using visuals, tags, or icons to trigger memory of key content.	P1, P2, P3, P4, P5, P8
	Filtering	Narrowing the view based on specific criteria.	P2, P3, P4, P6, P7
Strategies for Summary Communication (**RQ3**, Sec. 7.3.2)	Add Process Details	Including information about analytical steps or methodology.	P1, P2, P3, P4, P5, P9, P10
	Add Action-Oriented Context	Framing summaries in terms of goals, implications, and next steps.	P2, P3, P4, P5, P6, P7, P9
	Adjust Overall Length	Editing summaries for conciseness or comprehensiveness.	P4, P5, P6, P7, P8, P10
	Reorder/Format Content	Changing structure or visual style to improve readability and focus.	P2, P3, P6, P7, P10
Preferred Role of AI (**RQ4**, Sec. 7.5)	Retain Editorial Control	Wanting final say over summary content, especially in high-stakes settings.	P1, P2, P5, P7
	Selective + Structured Authoring	Leveraging AI for structure while choosing content manually.	P2, P3, P4, P6, P10
	Acceleration, Not Automation	Using AI to save time or surface drafts without ceding control.	P2, P5, P6, P10
	Scaffolding analyses	Using AI to help structure the analysis process	P2, P3, P10

Table 3: Codes organized by analytic focus area, with definitions and participants.

A.3 Prompt Templates

Summarize Turns. We first summarize each turn before extracting the threads.

System prompt:

You are an AI Conversational Data Analysis Assistant who is an expert at understanding a data conversation. In a conversation, because data analysis is non-linear and iterative, there can be multiple threads of analysis being explored. Analysis threads are distinct, focused lines of investigation or exploration within a larger analytical conversation. Each thread can address a specific question, hypothesis, or aspect of the overall topic, often requiring its own set of methods and analyses. There can also be sub-threads within a thread that go into more detail on a specific aspect of the thread.

🧑 **User:**

<Instruction> We want to capture the structure of analysis threads in a conversation.

First, we need to understand what the each turn in the conversation is exploring. Given a conversation history which is a list of ordered messages between the user and the assistant with each message in the following format:

```json
{
  "turn": 0,  // id for the turn in the conversation PAY ATTENTION TO THIS
  "role": "user",  // "user" or "assistant" or "tool"
  "content": "Hello World",  // the content of this message, in markdown format string
}
```

return a short half-sentence summary for each turn. A single turn encompasses both the user query and the assistant messages. Therefore, there are more than one message (JSON) with the same turn ID.

Note:

- The number of threads in the conversation should match the number of turns in the conversation (as specified by the "turn" field).
- This means that the number of threads should be the same as the number of unique turn IDs in the conversation history.
- Be sure to include the turn number that the summary references Please return the result in the format specified in **Format Instructions** enclosed in triple backticks.

</Instruction>

<Format Instructions>

```
// Format instructions based on LangChain's Pydantic JSON Parser (hidden for brevity)
```

<Format Instructions/>

<Example>

Conversation History:

```
// Real example of an analytical conversation analyzing airlines (hidden for brevity)
```

Analysis Turn Summaries:

```json
{
  "threads": [
    { "tid": 1, "title": "Any source airports? no", "turn": 0 },
    { "tid": 2, "title": "Any destination airports? no", "turn": 1 },
    { "tid": 3, "title": "What is the most traveled route", "turn": 2 },
    { "tid": 4, "title": "Pricing Per Airline", "turn": 3 },
    { "tid": 5, "title": "Airline Class Differences", "turn": 4 },
    ... // (hidden for brevity)
    { "tid": 15, "title": "Plot Improvement 1: breakdown by airline", "turn": 14 },
    { "tid": 16, "title": "Plot Improvement 2: average price per airline", "turn": 15 },
    { "tid": 17, "title": "Plot Alternative 3: merge all airlines", "turn": 16 },
    { "tid": 18, "title": "Conversation Summary", "turn": 17 }
  ]
```

```
}
```

</Example>

Begin.
Conversation History:

EXAMPLE INPUT

```json
[
  {
    "role": "user",
    "content": "Dear Chatty, I have a dataset to analyse. Let`s imagine that me and my business partner are
        planning to open a coffee shop in our city. To make informed decisions, we will analyze data to guide
        key strategies such as which coffee to source, pricing, and target customer demographics. Our goal is
        to uncover actionable insights to shape your business plan. The data set that we have at our disposal
        is called \"The Great American Coffee Taste Test\".  The link to the dataset is here : [csv_link]\n\n
        Dataset Description: In October 2023, \"world champion barista\" James Hoffmann and coffee company
        Cometeer held the \"Great American Coffee Taste Test\" on YouTube, during which viewers were asked to
        fill out a survey about 4 coffees they ordered from Cometeer for the tasting.",
    "turn": 0
  },
  {
    "role": "assistant",
    "content": "Got it! You`ve uploaded the dataset, so let`s start by exploring its structure. I`ll load the
        data and give you a quick summary, including column names, data types, and a few sample rows.",
    "turn": 0
  },
  ... // (hidden for brevity)
]
```

Analysis Turn Summaries:

🖥️ **Assistant:**

EXAMPLE OUTPUT

```json
{
  "threads": [
    {"tid": 1, "title": "Initial setup and dataset overview", "turn": 0 },
    {"tid": 2, "title": "List of column names and their top categories", "turn": 1 },
    ... // (hidden to save space)
    {"tid": 7, "title": "Re-upload dataset due to reset", "turn": 6 },
    {"tid": 8, "title": "Summary of detailed coffee preferences", "turn": 7 },
    {"tid": 9, "title": "Number of people matching each category", "turn": 8 },
    {"tid": 10, "title": "Clarifying categorization criteria for accuracy", "turn": 9 },
    {"tid": 11, "title": "Confirming distinct elements of where_drink column", "turn": 10 },
    {"tid": 12, "title": "Calculating count distinct for customer categories", "turn": 11 }
  ]
}
```

Consolidate Threads. We group similar threads level by level starting with the turn summaries.

System prompt:

You are an AI Conversational Data Analysis Assistant who is an expert at understanding a data conversation. In a conversation, because data analysis is non-linear and iterative, there can be multiple threads of analysis being explored. Analysis threads are distinct, focused lines of investigation or exploration within a larger analytical conversation. Each thread can address a specific question, hypothesis, or aspect of the overall topic, often requiring its own set of methods and analyses. There can also be sub-threads within a thread that go into more detail on a specific aspect of the thread.

🧑 **User:**

<Instruction> We want to iteratively capture the structure of analysis threads in a conversation.
Given the analysis threads summary and how the conversation is structured, determine if any of the threads can be grouped together to represent a larger thread. If so, group the threads together and provide a new summary for the grouped thread.
Please return the grouped analysis thread summaries based on the format specified in **Format Instructions** enclosed in triple backticks.
For any new thread, please give a new unique thread ID. In addition, be sure to reference the input thread IDs that were grouped together. If no threads can be grouped together, return an empty YAML.
</Instruction>
<Format Instructions>

```
// Format instructions based on LangChain's Pydantic JSON Parser (hidden to save space)
```

<Format Instructions/>
Begin.
Analysis Threads Summary:

EXAMPLE INPUT

```json
{
  "threads": [
    {"tid": 1, "title": "Initial setup and dataset overview", "turn": 0 },
    {"tid": 2, "title": "List of column names and their top categories", "turn": 1 },
    ... // (hidden to save space)
    {"tid": 9, "title": "Number of people matching each category", "turn": 8 },
    {"tid": 10, "title": "Clarifying categorization criteria for accuracy", "turn": 9 },
    {"tid": 11, "title": "Confirming distinct elements of where_drink column", "turn": 10 },
    {"tid": 12, "title": "Calculating count distinct for customer categories", "turn": 11 }
  ]
}
```

Grouped Analysis Threads Summary:

🖥️ **Assistant:**

EXAMPLE OUTPUT

```json
{
  "threads": [{
      "tid": 13,
      "title": "Coffee drinker categorization and detailed analysis",
      "child_ids": [3, 4, 5, 6, 8, 9, 10, 12]
    }]
}
```

A.3.1 Prompts for summarizing turns and extracting threads.

Speech Acts Extraction Prompt

System prompt:

You are an AI Conversational Data Analysis Assistant who is an expert at understanding a analytical conversation. The conversation consists of user queries and assistant responses. We want to understand the speech acts in the conversation. Speech acts is a kind of action being performed by the speaker. The analytical conversation will consist of the following speech acts from the user:
- Fact Finding: expecting a single response such as a number, or table, or "yes/no" (e.g., "did any passengers survive on Titanic?")
- Specific Visualization: requesting a specific type of visualization (e.g., "can you plot a bar chart of the number of passengers by class?")
- Domain Knowledge: asking about domain knowledge that is not directly answered by the data (ONLY include when the assistant does not have code in the response)
- Deeper Insights: e.g., "have there been outliers per class?", "How much more likely was a passenger to survive if they were in first class?"
- Data Transformations: data manipulations such as filtering, sorting, or aggregating (e.g., "can you filter the data to show only passengers in first class?")
- Recommendations: asking for recommendations or predictions based on the data (e.g., "what is the best course of action for the airline?")
- Refinement or Follow-up: refining a previous query or asking for more details (e.g., "can you show the survival rates by age group?")
- Debugging: asking for clarification or help with the analysis (e.g., "I don't understand the visualization")

🧑 **User:**

<Instruction> Given the user queries used in the conversation, which is a list of user messages with each message in the following format:

```json
{
  "turn": 0,  // id for the turn in the conversation
  "role": "USER",  // "USER
  "content": "Hello World",  // the content of this message, in markdown format string
}
```

please return the speech act for each user query. Please return the result based on the format specified in **Format Instructions**.
</Instruction>
<Format Instructions>

```
// Format instructions based on LangChain's Pydantic JSON Parser (hidden to save space)
```

</Format Instructions>

Begin.
User Queries:

EXAMPLE INPUT

```json
[
  {
    "role": "user",
    "content": "Dear Chatty, I have a dataset to analyse...", // (shortened for brevity)
    "turn": 0
  },
  {
    "role": "user",
    "content": "Can you list all the columns names and their 3 main attributes, so I have a sense of the data ?
        For the top 3 categories for each columns names give me the frequency in number and %",
    "turn": 1,
  }
  ... // (hidden for brevity)
]
```

Speech Acts:

Assistant:

EXAMPLE OUTPUT

```json
{
  "speech_acts": [
    { "msg_turn_id": 0, "speech_act": "Fact Finding" },
    { "msg_turn_id": 1, "speech_act": "Fact Finding" },
    { "msg_turn_id": 2, "speech_act": "Domain Knowledge" },
    { "msg_turn_id": 3, "speech_act": "Data Transformations" },
    { "msg_turn_id": 4, "speech_act": "Refinement or Follow-up" },
    { "msg_turn_id": 5, "speech_act": "Refinement or Follow-up" },
    { "msg_turn_id": 6, "speech_act": "Fact Finding" },
    { "msg_turn_id": 7, "speech_act": "Refinement or Follow-up" },
    { "msg_turn_id": 8, "speech_act": "Fact Finding" },
    { "msg_turn_id": 9, "speech_act": "Fact Finding" },
    { "msg_turn_id": 10, "speech_act": "Refinement or Follow-up" },
    { "msg_turn_id": 11, "speech_act": "Fact Finding" },
    { "msg_turn_id": 12, "speech_act": "Fact Finding" }
  ]
}
```

A.3.2 Prompt for extracting speech acts.

Insight Extraction Prompt

System prompt:

You are an AI Data Analysis Assistant who is an expert at extracting insights from a data conversation. Here is some background information about data insights:

*Data Insight**

Data insights are numerical or statistical results derived from data.

Insight Category

There are 11 types of data insights: Value, Proportion, Difference, Distribution, Trend, Rank, Aggregation, Association, Extreme, Categorization and Outlier. The explanations of these types are listed below:

- Value: Retrieve the exact value of data attribute(s) under a set of specific criteria.
- Proportion: Measure the proportion of selected data attribute(s) within a specified set.
- Difference: Compare any two/more data attributes or compare the target object with previous values along with the time series.
- Distribution: Demonstrate the amount of value shared across the selected data attribute or show the breakdown of all data attributes.
- Trend: Present a general tendency over a time segment.
- Rank: Sort the data attributes based on their values and show the breakdown of selected data attributes.
- Aggregation: Calculate the descriptive statistical indicators (e.g., average, sum, count, etc.) based on the data attributes.
- Association: Identify the useful relationship between two data attributes or among multiple attributes.
- Extreme: Find the extreme data cases along with the data attributes or within a certain range.
- Categorization: Select the data attribute(s) that satisfy certain conditions.
- Outlier: Explore the unexpected data attribute(s) or statistical outlier(s) from a given set.

Insight Evidence

Insight evidence are intermediate outputs in the conversation that **DIRECTLY** support each insight. Here are the types of insight evidence and their explanations:

* Code: The **EXACT** piece of code that **DIRECTLY** derives the insight. Dependency import and visualization code are **EXCLUDED** from consideration. For example, if the insight is "The positive rate for females is more than double that of males", the piece of code we focus on is

```python
\# Filter only the positive rates
male_positive_rate = gender_positive_rate.loc['Male', 'Positive']
female_positive_rate = gender_positive_rate.loc['Female', 'Positive']
```

* Code Console Output: The **EXACT** code output that **DIRECTLY** derives the insight. For example, if the insight is "There is a particular peak of worldwide gross in 1999", the code output we focus on is

```console
Release Year  Worldwide Gross
...           ...
14            1999
...           ...
```

* Source Text: The **quote** of conversation where the insight is summarized. You must identify the **minimal but critical** part of the conversation that **DIRECTLY** derives the insight.

🧑 **User:**

<Instruction> Given an analysis conversation which is a list of ordered messages and each message is in the following format:

```json
{{
```

```
  "id": 0,  // unique id of every message
  "role": "USER",  // "USER" or "ASSISTANT"
  "content": "Hello World",  // the content of this message, in markdown format string
}}
```

extract all insights from the conversation. Please return the insights based on the format specified in **Format Instructions**.
IMPORTANT:
- Find as **MANY** insights as possible from the conversation. But remember, **QUALITY** is more important than **QUANTITY**. **NEVER EVER** make up insights. **ONLY** derive insights from the conversation and the dataset.
- You should only extract insights that are directly related to **Data Analysis**. Irrelevant insights **MUST NOT** be included.
- **NEVER EVER** make up insight types. **ONLY** classify the insights into the 11 categories: [Value, Proportion, Difference, Distribution, Trend, Rank, Aggregation, Association, Extreme, Categorization, Outlier].
- Insight evidence, especially **CODE** and **CODE OUTPUT**, must be **DIRECTLY** releated to the corresponding insight, following the instructions above. Make them **ACCURATE** and **CONCISE**.
- The insights should reference the original messages (via the 'id' field) that are part of a given thread. These messages do not have to be continuous and may contain disjoint parts of the conversation.
</Instruction>
<Format Instructions>

```
// Format instructions based on LangChain's Pydantic JSON Parser (hidden to save space)
```

</Format Instructions>
Begin.
Convsersation History:

EXAMPLE INPUT

```json
[
  {
    "role": "user",
    "content": "Dear Chatty, I have a dataset to analyse. Let`s imagine that me and my business partner are
        planning to open a coffee shop in our city. To make informed decisions, we will analyze data to guide
        key strategies such as which coffee to source, pricing, and target customer demographics. Our goal is
        to uncover actionable insights to shape your business plan. The data set that we have at our disposal
        is called \"The Great American Coffee Taste Test\".  The link to the dataset is here : [csv_link]\n\n
        Dataset Description: In October 2023, \"world champion barista\" James Hoffmann and coffee company
        Cometeer held the \"Great American Coffee Taste Test\" on YouTube, during which viewers were asked to
        fill out a survey about 4 coffees they ordered from Cometeer for the tasting.",
    "turn": 0
  },
  {
    "role": "assistant",
    "content": "Got it! You`ve uploaded the dataset, so let`s start by exploring its structure. I`ll load the
        data and give you a quick summary, including column names, data types, and a few sample rows.",
    "turn": 0
  },
  ... // (hidden for brevity)
]
```

Insights:

🖥️ **Assistant:**

EXAMPLE OUTPUT

```json
{
  "insights": [
    {
      "insight": "The dataset contains 4,042 entries and 57 columns, covering various aspects of coffee
          consumption, preferences, and demographics.",
      "keywords": ["dataset", "entries", "columns", "structure"],
      "sourceMessageIds": [3],
      "types": ["Value"],
      "evidence": {
        "code_console": ... // (hidden for brevity)
        }
    },
    {
      "insight": "The top 3 values for the 'age' column are '25-34 years old' (1986, 49.51%), '35-44 years old'
          (960, 23.93%), and '18-24 years old' (461, 11.49%).",
      "keywords": ["age", "values", "frequency"],
      "sourceMessageIds": [8],
      "types": ["Value", "Rank"],
      "evidence": {
        "code_console": "{'age': [('25-34 years old', 1986, 49.51),\n  ('35-44 years old', 960, 23.93),\n
            ('18-24 years old', 461, 11.49)]}"
      }
    },
    {
      "insight": "The top 3 values for the 'gender' column are 'Male' (2524, 71.64%), 'Female' (853, 24.21%),
          and 'Non-binary' (103, 2.92%).",
      "keywords": ["gender", "values", "frequency"],
      "sourceMessageIds": [8],
      "types": ["Value", "Rank"],
      "evidence": {
        "code_console": "{'gender': [('Male', 2524, 71.64),\n  ('Female', 853, 24.21),\n  ('Non-binary', 103,
            2.92)]}"
      }
    }
    ... // (hidden for brevity)
    ]
}
```

A.3.3 Prompt for extracting insights.

A.3.4 Prompt for rewriting from serialized markdown. For this prompt, we have preset definitions for the length, technical detail, and formality (see Table 4). These correspond to the sliders in SyncSense's Authoring panel.

Slider	Level	Description
Length	1	The summary length should be very short and concise. It should be read in 30 seconds or less (100 words or less).
	2	The summary length should be moderate. It should be read in 2 minutes or less (around 300–400 words).
	3	The summary length can be longer. It should be read in 5 minutes or less (around 800–1,000 words).
Technical Detail	1	The summary should focus on key insights with minimal technical details for a non-technical audience.
	2	The summary should balance high-level insights with details that include the concrete data analysis and statistical methods for a semi-technical audience.
	3	The summary should focus on the concrete data analysis and statistical methods for a technical audience.
Formality	1	The summary could be informal.
	2	The summary should be semi-formal.
	3	The summary should be formal.

Table 4: Slider definitions for summary length, technical detail, and formality.

Summary from Markdown Prompt

System prompt:

You are an AI Conversational Data Analysis Assistant, specialized in summarizing analytical conversations. Before proceeding with any analysis, it's important to understand the key elements of an analytical conversation.

The conversation is structured as follows:

1. **Analysis Threads**: These represent distinct lines of investigation or exploration within the conversation. Each thread focuses on a particular question, hypothesis, or aspect of the overall topic, requiring specific methods and analyses. A thread is a collection of related turns, which may not always be consecutive. Sub-threads may also exist, diving deeper into particular aspects of the main thread.
2. **Insights**: These are key findings or observations that emerge from the data. Insights come from certain turns in the conversation and are supported by evidence in the form of code, code console output, or source text in one or more turns.
3. **Artifacts**: These are the artifacts generated during the analysis process, such as visualizations, code, or data tables.

🧑 **User:**

<Instruction>

Background We have distilled the conversation into the elements that we want to be summarized written in markdown format. The elements can include analysis threads, conversation turns, insights, and artifacts (visualizations, code, and data tables).

Now, we want to reformat the extracted content into a coherent and concise summary that is tailored to specific user specitifcations. These specifications center around the length of the summary, the level of technical detail, and the formality of the language used.

For instance, the level of technical detail can range from a high-level overview of the analysis to a detailed explanation of the methods and results.

Task Given the the extracted summarized content and user specified length, technical detail, and language formality requirements please reformat the content that is tailored to the specification.

Important:

- DO NOT add any new information that is not present in the extracted content such as any personal opinions, interpretations, insights, or additional analysis.

- Do your best to adhere to the user specified requirements for length, technical detail, and language formality but DO NOT add extra content that is not present in the extracted content.
- IF there are any attached code, and data tables, they will be links in the extracted content and should be referenced in the summary.
- IF there are any images, they will also be links in the extracted content and should be included.
- DO NOT include any links that are not present in the extracted content.

</Instruction>

Begin.
Contents to be reformatted:

EXAMPLE INPUT

```
* [Chat Turn 5] Coffee shop preferences
  * [Insight] Specialty Coffee Shops are a dominant preference among respondents, chosen by 16.3% of respondents.
  * [Table for Coffee shop preferences | Turn 6](table.csv)
* ![Title - Favorite Coffee Types by Employment Status, Title - Brewing Methods by Employment Status... Turn
    8](viz_0.png)
* [Chat Turn 10] Coffee preferences by gender
  * [Code for Coffee preferences by gender | Turn 11](code_snippet.py)
  * ![Title - Favorite Coffee Types by Gender, Title - Brewing Methods by Gender... Turn 11](/viz_1.png)
* [Insight] The dataset contains 4011 responses with 7 unique age groups, where the most common age group is "
    25-34 years old" (1986 respondents).
```

Length specification:

EXAMPLE INPUT

The summary length should be very short and concise. It should be read in 30 seconds or less (100 words or less).

Technical detail specification:
When thinking about the level of technical detail, consider technical and non-technical audiences.
Non-technical audience focuses more on the big-picture goals of the work and interpretation of the findings rather than specific details.
Technical audience, however, is able to interpret the concrete data analysis methods.
Based on the above definition, the technical detail should be the following:

EXAMPLE INPUT

The summary should focus on key insights with minimal technical details for a non-technical audience.

#Formality specification:
Formal and informal writing differ in tone, purpose, and audience. Here's a comprehensive overview of each style.
Formal Writing
Used for serious, professional contexts, formal writing is clear, respectful, and direct, often reserved for business, academic, or journalistic settings. This tone is suited for addressing unfamiliar audiences, such as potential employers or formal reports.
Key Characteristics:
• Word Choice: Uses precise language and avoids slang. Example: "Please eliminate that slide from the presentation."
• Punctuation: No exclamatory punctuation. Example: "I appreciate your help."
• Structure: Full sentences, often with complex syntax. Example: "Have you completed the monthly report?"

• Clarity: Avoids ambiguity, slang, or emoji. Example: "I'm running late for our meeting. I'll be there soon."
Formal Writing Techniques:
• Avoids contractions and personal pronouns.
• Includes technical terminology relevant to the topic.
• Uses respectful, elaborate descriptions for clarity.
• Minimizes personal opinions and feelings.
Informal Writing
Informal writing has a conversational, relaxed tone and is commonly used in casual communication, such as chatting with friends or writing blog posts.
Key Characteristics:
• Word Choice: Includes slang, idioms, and casual expressions. Example: "Let's lose slide #23."
• Punctuation: Exclamation points and emoji can be used. Example: "I appreciate your help!"
• Structure: Often shorter, with direct or even fragmented sentences. Example: "Monthly report ready?"
• Tone: Personal opinions and creativity are encouraged.
Informal Writing Techniques:
• Uses contractions and pronouns freely.
• Can meander off-topic, mimicking conversational flow.
• Prioritizes tone over strict grammar and spelling.
Based on the above definition, the formality should be the following:

EXAMPLE INPUT

The summary should be semi-formal.

Final summary:

🖥️ **Assistant:**

EXAMPLE OUTPUT

```
The analysis reveals key insights regarding coffee preferences. Specialty Coffee Shops are a top choice,
     preferred by 16.3% of respondents (refer to Table). The dataset comprises 4011 responses across 7 age
     groups, with the "25-34 years old" group being the most represented (1986 respondents). Additionally,
     coffee preferences were examined by gender (refer to attached Code and Image). The study also includes
     visualizations on favorite coffee types and brewing methods by employment status (see Image).

![Coffee Preferences Table](table.csv)
![Coffee Types and Brewing Methods by Employment Status](viz_0.png)
![Coffee Types and Brewing Methods by Gender](viz_1.png)
```