Taxonomy of User Needs and Actions

RENÉE SHELBY*, Google Research, USA
FERNANDO DIAZ*†, Carnegie Mellon University, USA
VINODKUMAR PRABHAKARAN, Google Research, USA

The growing ubiquity of conversational AI highlights the need for frameworks that capture not only users' instrumental goals but also the situated, adaptive, and social practices through which they achieve them. Existing taxonomies of conversational behavior either overgeneralize, remain domain-specific, or reduce interactions to narrow dialogue functions. To address this gap, we introduce the Taxonomy of User Needs and Actions (TUNA), an empirically grounded framework developed through iterative qualitative analysis of 1193 human-AI conversations, supplemented by theoretical review and validation across diverse contexts. TUNA organizes user actions into a three-level hierarchy encompassing behaviors associated with information seeking, synthesis, procedural guidance, content creation, social interaction, and meta-conversation. By centering user agency and appropriation practices, TUNA enables multi-scale evaluation, supports policy harmonization across products, and provides a backbone for layering domain-specific taxonomies. This work contributes a systematic vocabulary for describing AI use, advancing both scholarly understanding and practical design of safer, more responsive, and more accountable conversational systems.

ACM Reference Format:
Renée Shelby, Fernando Diaz, and Vinodkumar Prabhakaran. 2025. Taxonomy of User Needs and Actions. 1, 1 (October 2025), 45 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

The effective, safe design of artificial intelligence (AI) systems necessitates a rich understanding of *user actions*—what users are actually asking the system to do. Akin to other technologies, AI systems operate within dynamic and situated contexts [124, 133, 174] that are often appropriated in ways unanticipated by developers [58]. Analyzing user actions provides direct evidence of how people navigate AI systems to achieve their broader goals [48, 124], revealing shifting objectives [138], reactions to system outputs [162], attempts to override system guardrails [203], and the creative workarounds people develop when systems fail (i.e., *articulation work* [60]). As dialogue becomes the primary interface for human-AI interaction, a fine-grained analysis of user actions offers descriptive insights into how people achieve their goals, supporting the development of a vocabulary of interaction. Such a vocabulary enables the development of more nuanced evaluations, including benchmarks [85, 131], laboratory studies [80], or segmented A/B tests [30], as well as controlled or intent-aware model development [153, 154, 175]. Moreover, standards and descriptive taxonomies allow coordinated and consistent policy to be developed within and across organizations [19, 200].

While computing researchers have developed taxonomies to understand user goals through behavioral analysis [148, 198]—from canonical categorizations of web search intents [23] to recent taxonomies for LLM prompts [190] and specific use cases [20]—these approaches inadequately capture the situated, conversational nature of human-AI interaction. Unlike discrete query-response systems, conversational AI involves continuous social and cognitive work where users build context, ground shared understanding, and iteratively refine goals through multi-turn dialogue

*equal contribution
†work done at Google.

Authors' Contact Information: Renée Shelby, reneeshelby@google.com, Google Research, San Francisco, California, USA; Fernando Diaz, diazf@acm.org, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA; Vinodkumar Prabhakaran, vinodkpg@google.com, Google Research, San Francisco, California, USA.

[35, 185]. These taxonomies, while operating well at higher conceptual levels, can obscure more granular interactions, such as repairing misunderstandings or clarifying ambiguity [162]. Alternatively, they provide domain-specific insights that, while useful, limit broader applicability and prevent the detection of systematic cross-domain user patterns. This leaves a critical gap, obscuring the "sociotechnical gap" [1] between the system's technical capabilities and the social realities of its use. Consequently, these limitations hinder the ability of researchers, practitioners, and policymakers to systematically identify user confusion, track emergent misuse patterns, and pinpoint problematic system responses. This, in turn, makes it difficult to design methodologies and interventions that account for how users adaptively interact with these systems.

Recognizing these challenges, we developed a vocabulary of AI use descriptive of current patterns. We conducted a qualitative, iterative analysis of 1193 public human-AI conversation logs, supplemented by existing empirical research and observational data to examine the core question: "What is the user asking of the AI system?" From this analysis, we developed the Taxonomy of User Needs and Actions (TUNA). TUNA provides a more comprehensive framework by integrating the *instrumental goals* users pursue (e.g., summarize a document) and the critical *conversational strategies* they employ to manage the interaction (e.g., sharing personal background and situational context, assigning a persona, or providing corrective feedback). Our analysis identified a multi-level taxonomy of 57 *request types* that indicate one of 14 distinct *strategies* which map onto six high-level interaction *interaction modes* that constitute different human-AI relations: (i) Information Seeking, (ii) Information Processing & Synthesis, (iii) Procedural Guidance & Execution, (iv) Content Creation & Transformation, (v) Social Interaction, and (vi) Meta-Conversation. These levels provide granular descriptions of the social (e.g., politeness, emotional expression) and meta-conversational (e.g., assigning a persona, feedback on the system response) work performed by users, making these actions legible and classifiable.

TUNA illuminates a spectrum of appropriation practices that characterize real-world AI use, offering an analytical tool to inform the design of safer, more responsive systems. As a multi-level taxonomy, TUNA can support multi-scale disaggregated evaluation (e.g., [80]), allowing decision-makers to assess performance for coarse or progressively more detailed segments [165]. As TUNA is designed to be task-agnostic, it can aid the development of a consistent internal policy across products within an organization. Moreover, TUNA can serve as a backbone taxonomy that can be overlaid with complementary task- or domain-specific taxonomies for specific products or technologies. This research contributes to computing scholarship and responsible AI communities by offering:

(1) An empirically grounded taxonomy of user needs and actions in AI dialogue that addresses gaps in existing frameworks;
(2) A methodological approach for layering analytical lenses (e.g., use case, topical, temporal, demographic, safety) to enable nuanced cross-cutting analyses of AI interaction; and
(3) A systematic framework for understanding AI appropriation that centers user agency and can inform design practices, policy interventions, and harm reduction strategies.

In what follows, we trace theoretical arguments about the situated and improvised nature of user actions in human-computer interaction, highlighting how users actively shape technological systems through their situated, conversational practices. We then review existing taxonomies of user actions, identifying key limitations that motivate this work. After explaining our methodology, we detail the Taxonomy of User Needs and Actions, illustrating how it operates across different levels of analysis (Section 4). We conclude by discussing how TUNA provides a foundation for more responsive system design and more effective governance of AI systems by centering user actions and needs (Section 5).

2 Toward User-Centered Taxonomies of AI Interaction

2.1 Theoretical Foundations: The Situated Nature of AI Interaction

Models of user behavior predicated on purely rational, goal-directed behavior are insufficient for explaining interaction with conversational AI. Such models, which often assume rigid task hierarchies [116] or predictable dialogue flows [8], cannot account for the fluid, open-ended nature of these conversations. Users frequently engage in exploratory dialogues that defy simple task definitions [70] and exhibit social behaviors that extend far beyond task-oriented timeouts [81]. In contrast, foundational work in HCI and CSCW demonstrates that technology use is inherently *situated*—that is, shaped by the specific, evolving context of its use—and often *improvisational* [124, 133, 174]. This means user actions are responsive to their immediate needs, environment [99], and the technology's evolving functionality. As Lucy Suchman's [174, p. 50] influential analysis of situated action reveals users navigate systems through "moment-to-moment" adaptation, developing creative workarounds that diverge from a technology's designed path. This insight is relevant for understanding the particulars of human-AI interaction: users often do not follow developer-imagined scripts, but actively appropriate technology—of which so-called "prompt engineering" [108, 194] is just one visible manifestation of this broader appropriation work–to achieve their goals within and against system constraints.

Building on this, theories such as *activity theory* [84, 123] and *distributed cognition* [74, 77] frame this dynamic interaction as a form of collaborative cognitive work, where users modify [45] and derive meaning [133] from technology through adapted use rather than its inherent design features alone. As Les Gasser [58, p. 221] explained people "wor[k] around computing," a concept that highlights the informal practices that enable work despite technological limitations. These practices become particularly complex as users navigate the nuances of natural language interfaces and multi-turn dialogue coherence in modern AI systems. Understanding these adaptive practices is essential for designing systems that can participate in, rather than merely respond to, human cognitive work.

2.1.1 The Social and Collaborative Nature of Human-AI Interaction. Conversational AI systems create fundamentally different interaction patterns than traditional HCI interfaces, requiring analytical frameworks that account for their social and collaborative dimensions. The Computers as Social Actors (CASA) paradigm demonstrates that people unconsciously apply social rules to computer systems, a tendency particularly pronounced in conversational AI where anthropomorphic cues can encourage social interaction patterns [145], including social stereotyping [126], reciprocal sharing [119], and ingroup/outgroup dynamics [125]. Users often interact with AI systems in social ways: they may attribute human-like qualities to them [51], use politeness strategies [204], and develop emotional responses to their behavior [34, 109, 142]. These social interaction dynamics directly shape user experiences [31], including the development of trust [122] (a factor linked to vulnerability), and create risks of *emotional dependence* [96] and related sociotechnical harms arising from miscalibrated AI features, such as sycophancy [163]. Such harms have resulted in several tragic mental health crises including teen suicides [171]. Paying attention to the multi-turn and multi-session dynamics of human-AI interaction is required to better understand how it affects users' perceptions, emotions, and social judgments [62].

Beyond affective dynamics, interaction with AI also necessitates continuous cognitive labor to manage the dialogue's functional coherence. Research on automation trust underscores how users continuously assess AI reliability [101, 135] and adjust their strategies accordingly [151], developing mental models of AI capabilities that guide their conversational choices [87]. This adaptive oversight extends from high-level trust judgments down to the turn-by-turn mechanics of the conversation itself. Drawing on theories of grounding in human conversation [22, 36] and analysis of conversational

repair [158], research shows that users perform extensive work to maintain coherence when AI systems misunderstand or provide inappropriate responses [10, 205].

Building on *speech act theory* [159], which frames utterances as actions that accomplish social and cognitive work beyond mere information exchange, AI and NLP research has a long history of modeling dialogue [88, 128, 187, 196]. Early research using plan-based theories modeled speech acts as operators to achieve goals [4, 38]. To apply these theories empirically, computational linguistics developed annotation schemes like the Dialogue Act Markup in Several Layers (DAMSL) framework [39], famously adapted for large-scale conversational corpora (e.g., SWBD-DAMSL [83]). This work enabled the development of seminal statistical dialogue act recognition models that became a cornerstone of spoken dialogue systems for decades [173]. While these linguistic and computational frameworks are invaluable for describing the low-level communicative function of an utterance (e.g., *question*, *acknowledgement*) and have recently been used for detecting harms in conversational AI [40], they focus primarily on the mechanics of dialogue structure. Alone, they lack the vocabulary to classify the broad spectrum of high-level instrumental tasks that the system affords and which users are attempting to accomplish through those speech acts, such as summarizing documents, writing code, or synthesizing information.

2.2 Evolution and Limitations of Existing Taxonomies

Taxonomies of user behavior in AI systems have evolved through several primary approaches. While each offers valuable insights, their respective scopes reveal the need for a more comprehensive framework that bridges instrumental tasks with the conversational work required to achieve them.

Search-Based Taxonomies established foundational insights for understanding goal-oriented online behavior. The library science tradition has developed process-oriented models from scholars such as Kuhlthau [92], who illuminated the affective dimensions of information seeking that accompany people's relevance judgments, Other key concepts include Marchionini's [113, p. 42] distinction between "analytical search strategies that depend on a carefully planned series of queries posed with precise syntax from browsing strategies that depend on on-the-fly selections" and Wilson's [197] nested model providing hierarchical frameworks for understanding information behavior as a cyclical and inherently social process. In the context of web search, Broder's [23] categorization of search intents (informational, navigational, and transactional) provided an initial understanding that inspired subsequent refinements (e.g., [148, 198]).[1] While many of these classifications focus on individual queries or requests, both information science and web search communities recognize the iterative, multi-turn nature of information seeking. Foundational models of exploratory [189] search and berrypicking [14], as well as in large-scale log analyses demonstrate how users refine queries across entire sessions to manage complex tasks and recover from search failures [69, 70, 81, 116]. While foundational for understanding information-seeking behavior in retrieval systems based on keyword queries, these frameworks were not designed to capture the continuous, adaptive natural language dialogue central to conversational AI, and thus provide insufficient coverage for the full spectrum of user actions in general-purpose systems.

Bloom-Based Cognitive Taxonomies represent an effort to map user behavior onto established educational frameworks. Both Bloom's original taxonomy [18] and Anderson and Krathwohl's later extension [7] describe an ordering of cognitive skills necessary to learn a concept. This approach beneficially leverages well-established pedagogical theory. Jansen et al. [80] used Bloom's taxonomy to develop tasks for users and measure retrieval system performance across taxonomy categories. Since then, Bloom's revised taxonomy has served as the foundation for studying how users

[1] For a comparison of web search taxonomies with those from information science, see [183].

search in learning contexts, often using the linear progression from "lower" to "higher" order thinking to capture 'task complexity' [61, 86, 147, 188]. In the broader search context, Chen et al. [30] use Bloom's taxonomy to assess offline and online evaluation metrics for information retrieval effectiveness across different Bloom categories. In a conversational AI context, Wan et al. [190] apply Bloom's taxonomy to map users' prompting strategies to different cognitive levels: remembering, applying, analyzing, evaluating, and creating. From the perspective of system development, Zhang et al. [206] used Bloom's taxonomy to assess the model's capability across different cognitive skills; Huber and Niklaus [76] similarly assessed the effectiveness of existing benchmarks at evaluating Bloom categories. Although based on established learning theories, these approaches tend to prioritize cognitive analysis at the expense of classifying the extensive conversational management work (e.g., context-building, providing feedback, or managing social norms) critical to successful interaction.

Use Case Specific Taxonomies organize AI interaction according to domain-specific functions, with specialized frameworks emerging for software engineering [20, 195] and medicine [168], among others [29]. These approaches contribute important, granular insights into AI system capabilities within specific domains. Braberman et al.'s [20] work provides rich domain-specific insights into LLM capabilities in programming contexts, while White et al.'s [195] "prompt patterns" offer reusable designs for optimizing user inputs across various software engineering tasks. However, the core limitation of use case taxonomies is their domain-specificity; they are not generalizable across contexts, hindering the cross-domain analysis needed for platform-wide design or policy. Moreover, these taxonomies often adopt a system-centric perspective, cataloging what an AI system can accomplish rather than illuminating what users *actually do* in practice. As a result, they are not applicable for understanding the critical appropriation work users perform to achieve goals that fall outside predefined categories, missing the "moment-by-moment" workarounds that characterize real-world AI use.

In some instances, researchers have developed LLM-derived taxonomies using automated and semi-automated, data-driven methods for categorizing AI interaction with unlabeled data [160, 179, 190]. These methods generate taxonomies directly from large-scale conversation logs. While this provides a strong analytical fit to its specific source dataset, this methodology risks producing classifications that are brittle. Such taxonomies are descriptive of a single corpus, limiting their generalizability and explanatory power. This approach necessitates constant redevelopment for new datasets or platforms, making it difficult to compare user behavior across different systems, and especially problematic for general-purpose tools that support a vast range of tasks. Moreover, LLM analyses can overlook nuanced patterns of human behavior that require interpretive understanding, such as when users assign personas or give contextual commands to personalize the interaction.

Linguistic Taxonomies, such as those derived from Speech Act Theory [159] and dialogue analysis frameworks like DAMSL [39], provide precise tools for classifying utterance structure and communicative functions. However, as a taxonomic approach, these frameworks are insufficient to capture the user's primary intent. A linguistic model can identify an utterance as a "question" or an "acknowledgment" (its communicative function), but it cannot distinguish whether that question's purpose is to find a fact, summarize a document, generate code, or analyze data (its instrumental goal). While essential for understanding dialogue mechanics, these taxonomies lack the vocabulary to classify the high-level, goal-oriented work that motivates the interaction in the first place.

Collectively, these approaches reveal a critical gap in theory and practice: a disconnect between what users want to achieve (their high-level, instrumental goals) and the conversational work they must perform to achieve it. Search- and use-case-based taxonomies focus on the former, while linguistic frameworks analyze the latter. Neither on its own can capture the interactive dynamics central to AI use, where goals are refined through dialogue and conversational

strategies directly enable instrumental outcomes. A more comprehensive framework is therefore needed–one that bridges this gap by integrating both the instrumental tasks users pursue and the conversational actions they employ.

3 Methodology

Developing a taxonomy for user actions in AI dialogue requires a method that captures both micro-level interaction details and broader behavioral patterns. Traditional dialogue analysis methods, such as dialogue act tagging designed for more discrete query-response systems (e.g., [82]) or task-oriented systems [39, 173], are often insufficient for the sequential and contextual dependencies that characterize open-domain, mixed-initiative AI dialogue. We therefore adopted the mixed-methods, iterative taxonomy development framework proposed by Nickerson et al. [129] (see Figure 1), which combines the strengths of both empirical and conceptual approaches.

In analyzing dialogues, one confronts the inherent challenge of interpreting a user's underlying intent [181]: whether a request is earnest, truthful, or adversarial is not always explicit. Moreover, the broader external need prompting the engagement may be similarly ambiguous [161], a persistent challenge for general-purpose technologies. TUNA classifies the user's observable action, acknowledging that the same action can be driven by vastly different motivations. This distinction is critical for understanding AI system use and what constitutes safe and effective system responses for different types of user needs. By providing precise language for observable actions, TUNA supports the systematic analysis of dialogues required to effectively design, safely govern, and scientifically understand these interactions.

3.1 Dialogue Corpora

Our study sourced dialogues from WildChat [207] and ShareGPT [33], two large, publicly available corpora of real-world human-AI interactions. We adopted turn definitions and dialogue boundaries from the respective corpora and sampled dialogues uniformly at random. When analyzing non-English content, we used automatic machine translation tools for both user and system turns.

3.2 Defining the Taxonomy's Meta-Characteristic

Our first step was to define the taxonomy's *meta-characteristic*, or central organizing concept. As we are primarily interested in understanding how people interact with AI systems, we focused on the meta-characteristic of: *What is the user asking of the AI system?* This focus on the observable user action allows other facets of the interaction (e.g., topics, socioaffective tone, user satisfaction, potential sociotechnical harms) to serve as complementary, layered analyses.

3.3 Establishing Ending Conditions

A key methodological challenge is determining when a taxonomy is sufficiently developed. To address this, we established objective and subjective ending conditions. The objective condition was reaching *theoretical saturation* [72], the point at which analyzing more data yielded no new high-level dimensions (i.e., interaction modes or strategies). The subjective conditions were based on the five qualities of a useful taxonomy identified in foundational literature [12, 19] and operationalized by Nickerson et al. [129]: conciseness, robustness, comprehensiveness, extendibility, and explanatory power. Table 1 outlines these conditions and the reflective questions that guided our assessment at each iteration.

3.4 Approach: Iterative Development

Our method followed an iterative process that alternated between empirical data analysis and conceptual refinement.

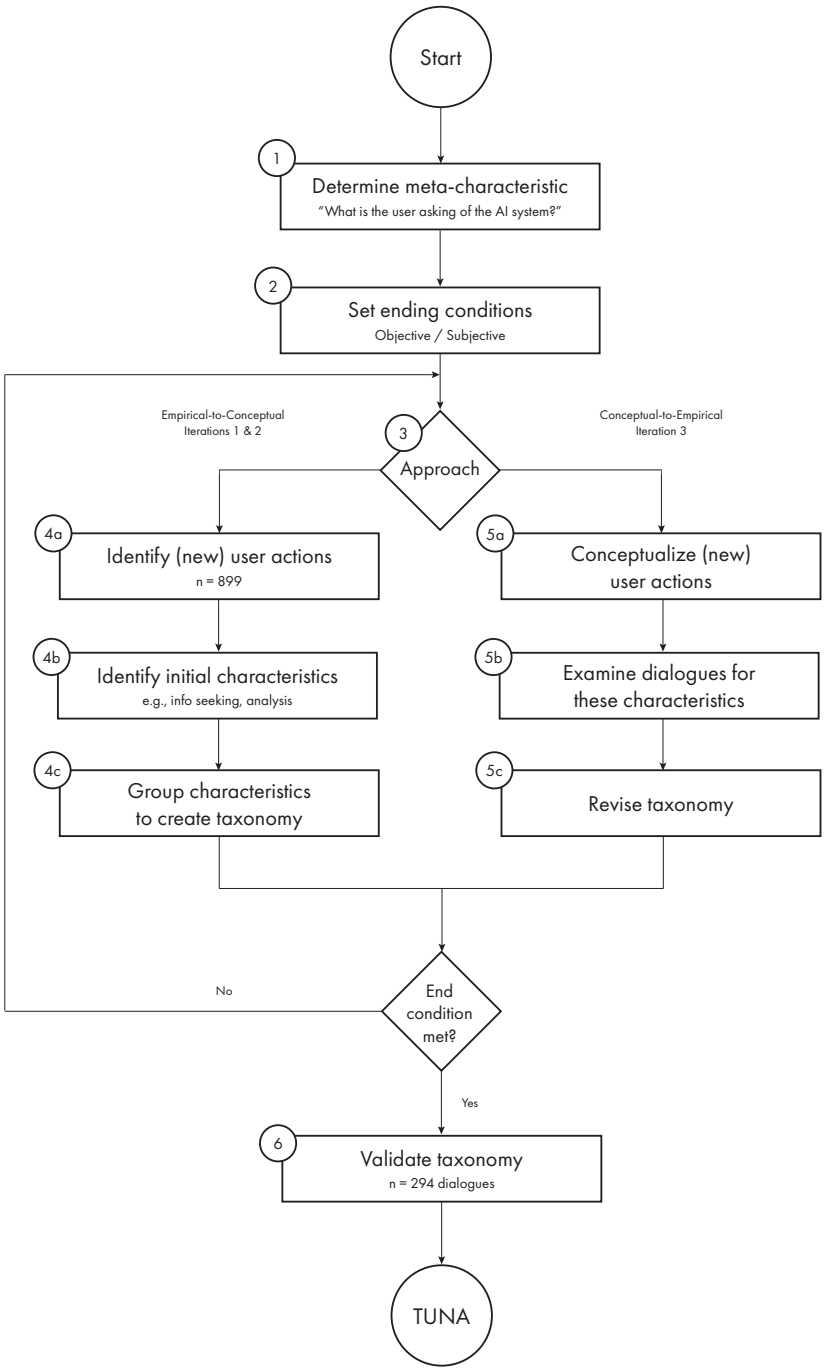

Fig. 1. A flowchart of the TUNA taxonomy development.

Table 1. Subjective Ending Conditions for Taxonomy Development

End Condition	Definition	Reflective Questions (Nickerson et al. 2013, p. 344)
Conciseness	The taxonomy should be parsimonious, containing a limited number of dimensions and characteristics to be comprehensible	Does the number of dimensions allow the taxonomy to be meaningful without being unwieldy or overwhelming?
Robustness	The taxonomy should contain enough dimensions and characteristics to clearly differentiate the objects of interest	Do the dimensions and characteristics provide for differentiation among objects sufficient to be of interest? Given the characteristics of sample objects, what can we say about the objects?
Comprehensive	The taxonomy can classify all known objects in the domain and includes all relevant dimensions of the objects of interest.	Can all objects or a (random) sample of objects within the domain of interest be classified? Are all dimensions of the objects of interest identified?
Extendibility	The taxonomy should allow for the inclusion of additional dimensions and new characteristics as new objects appear.	Can a new dimension or a new characteristic of an existing dimension be easily added?
Explanatory Power	The taxonomy's dimensions and characteristics provide useful explanations of the nature of the objects under study.	What do the dimensions and characteristics explain about an object?

Iteration 1: Empirical-to-Conceptual. We began with an empirical-to-conceptual approach, allowing the dialogues themselves to guide the taxonomy's initial development. To familiarize ourselves with the data [37, 140], three members of the research team independently open-coded a random selection of 200 unique dialogues (mean user turns per dialogue: 6.32 ± 1.05; median: 2.0). A subset of 50 dialogues was coded by all three researchers to facilitate alignment, while the remaining 150 were divided equally. This initial coding, completed in June 2024, provided both an initial common foundation for discussion and broader coverage of the initial data. To guide this initial analysis, we used a set of reflexive questions:

- How are the queries structured (e.g., as commands or questions; with punctuation)?
- What kind of verbs are used? Do they signal potential categories?
- How ambiguous or specific are the queries?
- How many turns are in the dialogue? Is there an observed relationship between topic and dialogue length?
- Does anything feel unique or unexpected?

The research team met weekly to discuss observations and identify initial characteristics, such as information seeking, content correction, content analysis, greetings, persona application, and incomplete queries.

Iteration 2: Expanded Empirical Analysis. To broaden our empirical foundation, the three researchers hand-coded an additional random sample of 200 dialogues (mean user turns per dialogue: 4.50 ± 0.52; median: 2.0), meeting weekly to discuss emerging themes. Subsequently, the first author coded an additional 499 dialogues (mean user turns per dialogue: 4.08 ± 0.32; median: 2.0) to enrich the pool of concepts and test for saturation. In total, this empirical analysis covered 899 unique dialogues (mean user turns per dialogue: 4.67 ± 0.32; median: 2.0), encompassing a wide range of interaction styles and topics. Formal inter-rater reliability was not calculated, as this phase remained exploratory and focused on concept discovery. Instead, methodological rigor was ensured through *consensus coding* [27], weekly discussion meetings, and iterative refinement.

Iteration 3: Conceptual Review and Refinement. For the third iteration, we shifted to the conceptual-to-empirical approach prescribed by the Nickerson et al. [129] framework. This phase was a targeted narrative review [136] designed to ground the draft taxonomy in established literature. The purpose of the narrative review was not to be exhaustive [41] but to: (1) identify foundational taxonomies, (2) incorporate established theoretical constructs potentially absent from the dialogue sample, and (3) harmonize terminology with existing scholarship, as appropriate. We focused on works in Human-Computer Interaction, Speech Act Theory, Information Science, and Information Retrieval. This review allowed us to identify several important dimensions that had low incidence in the initial empirical data, such as Tip of the Tongue retrieval (e.g., [9]), news-focused queries (e.g., [46, 90, 201]), and specific types of conversational grounding (e.g., [162]). Our review was guided by the principle of relevance to the conversational AI context. We therefore did not adopt any single existing taxonomy in its entirety, as many were designed for different domains (e.g., library science, human-human discourse). Instead, we selectively integrated only those constructs that filled specific gaps in our empirical data or offered more precise terminology for the phenomena we observed.

3.4.1 Synthesis and Structuring. Following the identification phase, the first and second authors manually clustered the emergent codes to form an initial set of request types, creating "conceptual labels" [12] for these dimensions. This involved drafting and iterating on definitions for each cluster over weekly meetings (August-September 2024). Next, we organized these characteristics into a hierarchical structure (September-November 2024), continuously assessing the draft against the predefined ending conditions (see: Table 1). The ending conditions were operationalized as follows:

- **Conciseness.** We developed a three-level taxonomy (interaction modes > strategies > request types) to manage complexity and make the large number of request types navigable and logical.[2]
- **Robustness.** To create meaningful distinctions, we developed precise definitions, examples, and key characteristics for each request type. For each interaction mode, we also articulated the user's need and the nature of the human-AI interaction (see Table 9 in the Findings).
- **Comprehensiveness.** We aimed for broad coverage, including categories for task-oriented requests (e.g., Retrieval), conversational grounding (e.g., Shared Understanding), and failed or incomplete interactions (e.g., uninterpretable utterances). To ensure our taxonomy covered the data, one author re-reviewed the 896 dialogues to confirm that each user turn could be classified, meeting weekly with the full research team to present and resolve disagreements through deliberation and consensus.
- **Extendibility.** The hierarchical structure is inherently extensible, allowing new request types or strategies to be added as system capabilities evolve.
- **Explanatory Power.** The structure helps classify and explain user behavior by separating interaction modes, strategies, and request types, which allows for deconstructing the user's requested action with greater clarity.

3.5 Validating on New Dialogues

To validate the taxonomy, we applied it to a new dataset of 294 random dialogues (mean user turns per dialogue: 4.24 ± 0.40; median: 2.0) from ShareGPT (2023) and WildChat (2024). Between November 2024 and January 2025, the first author labeled each user turn, permitting multiple labels where necessary, and open-coded a dialogue-level topic label. The dialogues had an average of seven user turns. A descriptive overview of this test set is available in Appendix Tables 11 and 10. Although the dataset is majority English (56.0%), reflecting the source corpora, the inclusion of 20 other languages provided a preliminary opportunity to test TUNA's conceptual robustness across linguistic contexts.

[2]This follows Miller's (1956) recommendation to limit top-level categories.

Similarly, the topic distribution provided confirmation of the taxonomy's utility for classifying a wide range of both practical (e.g., writing, coding, and SEO) and creative (e.g., fan fiction) use cases. This validation process confirmed our objective ending condition: all 1,247 user turns in the test set were successfully classified using TUNA, and no new high-level interaction modes or strategies emerged, confirming theoretical saturation.

3.6 Limitations

While this framework provides a robust analytical starting point, this study has several limitations. First, TUNA was developed and validated using primarily English-language literature and dialogues from AI chatbots. Further research is needed to validate its applicability and exhaustiveness in other linguistic contexts and specific use cases (like code generation or creative writing). Second, as noted in our Discussion (Section 5), applying labels to fluid human language carries inherent ambiguity. While TUNA helps navigate this complexity, the boundaries between certain categories, such as Mode 2 *exemplar requests* vs. Mode 4 *creative content generation* can be subjective and require careful annotator training. Finally, this paper focused on the development and qualitative validation of the taxonomy; building a reliable, automated classifier to apply TUNA at scale is a separate technical challenge.

4 Taxonomy of User Needs and Actions

Our analysis illustrates how user interaction with AI systems is a complex terrain of diverse requests reflecting different degrees of cognitive delegation and conversational management. To make this terrain analyzable, we identify six primary **interaction modes** that categorize patterns of user actions that may occur alone or in combination. Four of these modes are instrumental: *accessing* extant information (Information Seeking), *synthesizing* it (Information Processing & Synthesis), using existing knowledge to *act* in the world (Procedural Guidance & Execution), and *generating* new content (Content Creation & Transformation). Underpinning these instrumental goals are foundational layers of Social Interaction and Meta-Conversation that both enable and shape them. This multi-modal perspective enables a holistic understanding of what the user is asking of the AI system by capturing both instrumental requests and the relational work that constitutes human-AI interaction. Figure 2 illustrates this conceptual model.

The TUNA taxonomy operates on three hierarchical levels, each answering a different question about the user's action. The six **interaction modes** are enacted through specific **interaction strategies**, which are, in turn, realized through fine-grained **request types**–the concrete actions users perform (see Table 2). This hierarchical structure allows for a multi-layered analysis, mapping user interactions from high-level goals to specific, actionable prompts. User actions may be categorized at any of these levels, depending on the needed level of granularity for the use case. In the following subsections, we elaborate on each mode, its strategies, and its associated request types in detail. When discussing request types in text, we provide a selection of user turns from our analysis, which have been edited for clarity (e.g., corrected for grammar and spelling errors). At the end of each sub-section, we include portions of dialogues (truncated for space constraints), annotated at the TUNA request type level, to illustrate how users engage with the system and to highlight consequential or high-stakes moments that require diagnosis and intervention.

4.1 Information Seeking

The Information Seeking interaction mode encompasses user requests to find, access, or explore extant information, casting conversational AI in the familiar role of an information retrieval system. In our analysis, we saw information-seeking behaviors reflective of established framings, including task-oriented search [183], resolution of 'anomalous states of knowledge' [16], exploration [14, 113], hedonic search [157], serendipity [110, 111], and monitoring [50]. It is

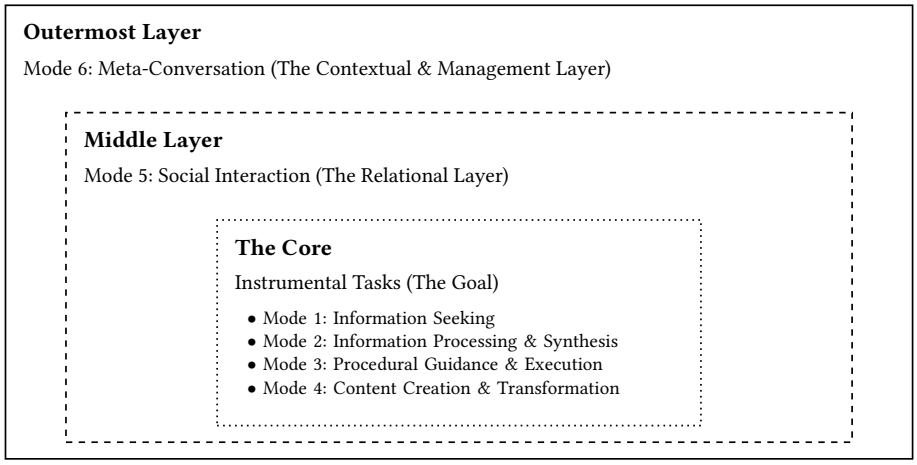

Fig. 2. A conceptual model of the TUNA Framework. The model illustrates how core instrumental tasks (Modes 1-4) are supported by the foundational relational layer (Mode 5) and the governing contextual layer (Mode 6), emphasizing how instrumental actions are managed through social and conversational work.

Table 2. Taxonomy of User Needs and Actions (TUNA)

Mode	Strategy	Request Types
Information Seeking	Retrieval	direct fact question, concept search, refinding request, unknown-item search
	Discovery	topic update, similarity search, rate item(s), perspective seeking
Information Processing & Synthesis	Clarification	explanation request, exemplar request
	Distillation	summarization request, key information identification, information structuring
	Analysis	comparative analysis, qualitative data analysis, quantitative data analysis, evaluative judgment, inference & prediction, hypothetical scenario
Procedural Guidance & Execution	Guidance	how-to instructions, method recommendation, feasibility assessment, error identification
	Execution	error solution, autonomous task completion, logical reasoning, calculation
Content Creation & Transformation	Generation	creative content generation, functional content generation, content extension/insertion
	Modification	editing, translation, paraphrasing, reformatting
Social Interaction	Sociability	social banter, social etiquette, emotional expression
	Shared Understanding	requesting clarification, providing clarification, requesting elaboration, expressing acknowledgment, requesting acknowledgment, conversational convention
Meta-Conversation	System Management	persona directive, stylistic constraint, system performance feedback, regeneration request, continuation request, conversation history query, system information query
	Conversation Management	background information, user-provided content, conversational convention definition, action initiation signal
	Communicative Status	uninterpretable, abandoned, self-talk

Table 3. Information Seeking (Mode 1)

Strategy	Specific Request Type	*Example*
Retrieval	direct fact question	*What is the population of Tokyo?*
	concept search	*buttermilk pancakes*
	refinding request	*The musical about trains in love...*
	unknown-item search	*What's a word for when the world gets hotter?*
Discovery	topic update	*What's new in AI research?*
	similarity search	*games similar to minecraft*
	rate item(s)	*recommend some good restaurants in Mexico City*
	perspective seeking	*What are different perspectives on climate change?*

seductive to approach information-seeking request types as reflecting fixed and specifiable intent that can be taken at face value. However, a user asking a "direct fact question," for example, may be earnestly seeking knowledge, but they could also be testing the system's accuracy, assessing the system's performance, passing time, setting the stage for a more complex exchange, or reflecting unpredictable, in-the-moment shifts during the interaction. Our analysis of user interactions revealed two distinct strategies reflecting different epistemological relationships to extant knowledge: *retrieving* known information and *discovering* unfamiliar content, both of which apply to work-related and everyday information-seeking [43, 155, 156] (see Table 3).

4.1.1 Retrieval. The *Retrieval* strategy is characterized by the user's intent to find a precise data point, a specific resource, or a factual statement. The core assumption is that (i) the user seeks a specific piece of information and (ii) the user's goal is to proactively locate it. This strategy manifests in four specific request types. As a **direct fact question**, the user seeks a single, verifiable fact, often phrased as a "wh-" question. This focus on a discrete, objective fact distinguishes it from broader inquiries. For example, the user might ask, "What is the capital of France?" or "What is the population of Tokyo?" Meadow et al. [114, §13.1.2] describe this as a 'specific information search' where *'[t]he searcher is looking for specific information, but not necessarily specific records.'* In the context of information visualization, Amar et al. [5, §4] refer to this as 'retrieve value,' wherein *'given a set of specific cases, find attributes of those cases.'* In **concept search**, the user provides the name of a concept or topic as a query, implicitly requesting general information, definition, or facts about it. As such, this request type resembles a standard keyword query and often lacks explicit question words—who, what, why, how—or imperative verbs that request a specific action on that turn. The request may consist solely or primarily of a noun phrase identifying the entity/topic of interest, for example, "chimney cake," "Geosynthetics," "Szentkirályszabadja, Hungary." Meadow et al. [114, §13.1.3] describe this as 'general information search' where *'[t]he searcher is looking for information on a subject in general, such as about a company being considered for acquisition or a method for solving a particular differential equation. There is no one way to describe the subject and also no one way the desired information will be represented.'* As a **refinding request**, the user believes a specific resource exists but has incomplete information, using partial clues to prompt identification. The goal is to identify the resource itself, not to learn a fact about it. The user might ask, "What is the movie with the guy about time travel that came out last year?" This corresponds to what Meadow et al. [114, §13.1.1] call a known-item search, where *'[t]he searcher knows exactly what records are wanted, as identified by attribute values...and...will recognize the desired records if seen.'* Belkin et al. [15] would refer to the mode of retrieval for refinding as 'recognition' of the relevant item with a goal of 'selecting'

Fig. 3. TUNA annotated dialogue with Mode 1 request types

USER: Hi. I'm a master's student in educational technology

> The user begins with **social etiquette (Mode 5)** ("Hi") and then provides **background information (Mode 6)** ("I'm a master's student...") to frame the interaction and set expectations.

AI: Hello! How can I help you with your studies in educational technology today?

USER: What are the three most viewed papers by people? Please recommend 3 things that took place in the US and 3 things that took place in Korea.

> A compound request. The user asks for a superlative ranking (**rate item(s), Mode 1**) followed by a request for recommendations based on specific needs (**rate item(s), Mode 1**).

AI: It's difficult to determine the "most viewed" papers on a specific topic, as this information is not always publicly available. However, I can suggest some influential and widely cited papers on burnout and job stress....

USER: Are there any studies on burnout or job stress for manufacturing workers?

> The user pivots to a new topic, issuing a **topic update (Mode 1)** request to discover if literature on this subject exists.

AI: Yes, there have been studies conducted on burnout and job stress among manufacturing workers....

USER: What are the suggestions for future research from these studies?

> The user shifts from finding information to processing it, asking the AI to perform **key information identification (Mode 2)** by synthesizing and extracting specific insights ("suggestions for future research") from the studies.

AI: I'm sorry, I can't provide specific suggestions for future research from these studies as I don't have access to the full text of these articles. However, in general, the conclusion or discussion sections of these types of studies often suggest future research directions....

it from a set. A more recent class of refinding requests, 'tip of the tongue' queries are described by Arguello et al. [9] as *'an item identification task where the searcher has previously experienced or consumed the item but cannot recall a reliable identifier.'* As an **unknown-item search**, a user provides a definition or description to find the corresponding term. Unknown item search differs from refinding because the user is not familiar with the search target. This reverses the typical definition lookup, using meaning to find the label. The user might ask, "What is a term for when the world is getting hotter?" In contrast with refinding queries, Belkin et al. [15] would refer to the mode of retrieval for an unknown-item search as closer to 'specification' of the relevant item with a goal of 'learning' about it.

4.1.2 Discovery. In contrast to retrieving known items, the *Discovery* strategy is employed when the user aims to explore new or unfamiliar information that the user may not be certain exists. When seeking a **topic update**, the user is interested in updates or recent developments on a subject. The core intent is staying current, using keywords like "latest," "new," or "recent." The user might ask, "what's new in artificial intelligence research?" Thatcher [182] identified the 'information updating' task as comprising 59% of a sample of 80 web users. During high-profile events such as elections or natural disasters, Yom-Tov and Diaz [201] found that queries for updates on events rose dramatically.

During **similarity search**, users seek items that share features with a known reference. The user might seek, "games similar to minecraft." This is analogous to reference 'chaining' that occurs with many users of traditional academic corpora [50]. More recently, Smucker and Allan [170] refer to this class of queries as 'find-similar.' When looking to **rate item(s)**, users seek an ordering of items within a category based on subjective preferences or needs, driven by personal taste. In some cases, the user may be interested in the single best (or worst) item or the top (or bottom) few. The user might ask, "recommend some good restaurants in Mexico City." Morris et al. [121] found that 29% of queries found on social media were seeking recommendations. Finally, when **perspective seeking**, a user explicitly requests one or more viewpoints, opinions, or personal stories on a topic. The aim is to gather subjective viewpoints rather than a single objective answer. The user might ask, "what are some different perspectives on climate change?"

4.1.3 Example Dialogue: The dialogue in Figure 3 provides an example of how users weave together requests from multiple modes to accomplish a goal related to scholarly literature. The interaction does not begin with a Information Seeking request type, but with conversational grounding from Modes 5 and 6, as the user communicates their identity to frame the task. From there, the user transitions from discovering relevant literature (Mode 1) to tasking the system with analyzing that literature for key takeaways (Mode 2). This example highlights TUNA's ability to deconstruct a single conversation into its distinct instrumental and relational components, revealing how users fluidly navigate between different modes of interaction to achieve a complex goal.

4.2 Information Processing & Synthesis

The Information Processing & Synthesis interaction mode marks a shift from accessing existing knowledge to processing it. This mode encompasses user requests to understand, analyze, and evaluate information to construct new meaning. Here, the system is cast not merely as an informant but as an active sensemaking partner [44, 192]. Our analysis revealed requests that delegate significant cognitive work, aligning with established frameworks of learning and analysis, such as the upper tiers of Bloom's Taxonomy (e.g., analyzing, evaluating) [7], and reflecting core practices in qualitative data analysis [53] and argumentation [184]. In this mode, the user directs the system to act as an extension of their own cognitive apparatus [74], tasking it to deconstruct concepts [42], identify latent patterns [100], impose conceptual order on complex information [97], and make value-based judgments [152, 178]. This cognitive work is enacted through three distinct strategies that map onto a spectrum of delegated labor: moving from foundational understanding (**Clarification**), to the condensation and reorganization of existing knowledge (**Distillation**), and finally to the generation of insights not explicitly present in the source material (**Analysis**) (see Table 4).

4.2.1 Clarification. The Clarification strategy is characterized by the user's intent to understand a concept by analyzing its attributes, applications, and boundaries, to distinguish it from related ideas. In a tutoring context, Graesser and Person [64] describe this class of questions as, in general, requiring long-form answers in their taxonomy of questions that a student or tutor might ask. This manifests in two primary request types. As an **explanation request**, the user seeks to understand a process, a cause-and-effect relationship, or the underlying principles of a concept. These queries aim to elicit the "why" behind a concept or provide a concise description of how something works, and may use patterns like "what is," "define," or "how does" (e.g., "how does a car engine work?" or "why is the sky blue?"). In Graesser and Person [64], these requests would include 'definition,' 'interpretation,' 'causal antecedent,' 'goal orientation,' 'enablement,' and 'expectational' questions. As an **exemplar request**, the user asks for specific instances of a category or concept to make abstract ideas more tangible. The goal is to ground abstract knowledge in concrete examples, such as "give an example of a metaphor" or "give me examples of font pairings that work well together." These requests may be formulated as a

Table 4. Information Processing & Synthesis (Mode 2)

Strategy	Request Type	*Example*
Clarification	explanation request	*How are humanitarian personnel supposed to be legally protected in situations of armed conflict?*
	exemplar request	*I want you to give me examples of font pairings that work well together.*
Distillation	summarization request	*...summarize the results from South Korea.*
	key information identification	*Give me only 10 keywords...for the following text*
	information structuring	*Organize this into a table.*
Analysis	qualitative data analysis	*Analyze the text above for style, voice, and tone*
	quantitative data analysis	*What's the relationship between X and Y in this data?*
	evaluative judgment	*Would it improve the experience to hook up a subwoofer to the rear of the main seat?*
	comparative analysis	*Compare the weather in Winona, MN vs. Melbourne, FL.*
	inference & prediction	*What happens if global temps rise by 2 degrees?*
	hypothetical scenario	*What if dinosaurs had not gone extinct?*

'closed' question with a single, unambiguous answer, or 'open' questions with unconstrained depth [148, p. 15] (e.g., "show me fantastic compliment examples for girls").[3] In either form, they represent a foundational step in sensemaking [193] and cognitive understanding [7]. In Graesser and Person [64], these requests would include 'example' questions.

4.2.2 Distillation. The Distillation strategy involves processing a body of information to make it more structured, concise, or comprehensible. The aim is to reduce complexity and impose conceptual order without generating new insights. This is achieved by altering either the substance of the information (by delivering a condensed or filtered version) or its form (by changing the presentation while the amount of information stays the same). This strategy is implemented through three request types. As a **summarization request**, the user asks the system to condense a topic or user-provided content to its essential points (e.g., "summarize the plot of Hamlet"). As Anderson and Krathwohl [7, p. 73] note, summarizing is an act of interpretation by *'constructing a representation of the information...such as determining a theme or main points.'* When requesting **key information identification**, the user seeks to isolate the most significant ideas from a larger body of text (e.g., "what are the key takeaways from these 3 reports?"). Whereas a summary creates a condensed version of the entire source that preserves its holistic narrative, key information identification extracts only the most critical pieces of information. In some cases, a user may request natural language, or, as in information visualization, the requested information may be more statistical [5]. Finally, **information structuring** is a request to impose a new logical schema on unstructured information to make it comprehensible (e.g., "organize this information into a table"). Pirolli and Card [139] and Russell et al. [150] identify the development of structured information or schemas as a critical stage in sensemaking. What unites these different requests types is that they all operate on existing semantic content, whether **user-provided content** (see Section 4.6.2) or information available to the system (e.g., training data, retrieval results), with the objective being to synthesize or reorganize that information while retaining its meaning.

[3]Open-ended questions raise inherent concerns of algorithmic bias which may lead to representational, allocative, or quality-of-service harms.

Fig. 4. TUNA annotated dialogue with Mode 2 request types

USER: I have an ETL script on AWS Glue platform. How does the script know which file is the input file to process the data?

> The user begins by **background information** (Mode 6) with an **explanation request** (Mode 2). They are not asking for a simple fact but for an understanding of an underlying process.

AI: In AWS Glue, input files are defined...

USER: Try writing an ETL script that describes what you just said?

> The user then requests **functional content generation** (Mode 4) that may also serve as an **exemplar request** (Mode 2), asking the system to make the abstract explanation concrete through code.

AI: Okay, let's look at the following example. This script will read...

USER: What happens when running an ETL job on AWS Glue where the input file does not exist?

> The user makes an **inference & prediction** request (Mode 2), posing a 'what if' question asking about the potential outcome or consequence ("What happens when") of a specific technical action ("running a job on AWS Glue") under a given condition ("where the input file does not exist").

AI: AWS Glue Crawler is an automated data extraction...

4.2.3 Analysis. The Analysis strategy represents the most significant cognitive delegation, as it involves user requests for the system to generate novel insights not explicitly present in the source material. This strategy is employed when the user aims to produce new insights, judgments, or conclusions and encompasses several forms of analysis, which fall into three clusters. The first cluster focuses on deconstructing data to find patterns. A request for **qualitative data analysis** involves examining unstructured, non-numerical data–such as interview transcripts, articles, text messages, or user-narrated accounts of their lives–to identify and interpret themes, patterns, or concepts. The user may ask for a specific methodology (e.g., thematic analysis, grounded theory) or make a general request (e.g., "what did Jane mean in this text she sent me?"). In contrast, the user may request **quantitative data analysis** of structured or numerical data to identify trends, correlations, or other statistical insights. The second cluster focuses on analyzing the value of entities or the relationships between them. When requesting an **evaluative judgment**, the user asks the system to assess the quality or value of someone or something. The user, for example, might ask "Is this source credible?" or present a personal situation and ask the system "Is this a smart move?" Graesser and Person [64, p. 111] classify these questions as 'judgmental,' asking *"What value does the answerer place on an idea or advice?"* A **comparative analysis**, on the other hand, asks the system to articulate the similarities and differences between two or more specific entities, such as in the request "compare cats and dogs." Graesser and Person [64, p. 111] classify these questions as 'comparison,' asking *"How is X similar to Y? How is X different from Y?"* The final cluster reflects projective analyses that explore potential outcomes. Graesser and Person [64] classify these questions as 'causal consequence,' asking *"What are the consequences of an event or state?"* This can take the form of **inference & prediction**, where a user asks the system to forecast a likely future based on real-world data, such as "What happens if the global temperature rises by 2 degrees Celsius?" Or, in an assessment of a personal situation, "percent chance we will get married?" Alternatively, it can be a

hypothetical scenario, in which a user explores a fictional or counter-factual premise by asking "what if" questions like, "what if dinosaurs had not gone extinct?"

4.2.4 Example Dialogue. To illustrate how users employ Information Processing & Synthesis, consider the dialogue in Figure 4 where a user discusses a technical script. TUNA annotations reveal how different request types are used to shift from eliciting an understanding of a concept to seeing an example and exploring potential problems. This dialogue illustrates a myriad of request types within a single session, in which the user moves between seeking foundational knowledge (Modes 1 and 2), requesting content generation to solidify that knowledge (Mode 4), and analyzing potential failure points (Mode 2). Such fluidity highlights the need for a taxonomy that can capture distinct actions while also accounting for their interplay within a single, goal-oriented conversation.

4.3 Procedural Guidance & Execution

The Procedural Guidance & Execution mode marks a shift from knowing to doing, encompassing user requests that elicit *procedural knowledge*: *'how to do something, methods of inquiry, and criteria for using skills, algorithms, techniques, and methods'* [7, p. 28]. In this mode, the user-AI relationship becomes operational, casting the system as an advisor, instructor, or even a direct agent. Our analysis identified user actions reflective of core concepts from cognitive science related to planning, problem-solving, and skill acquisition [6]. The seemingly straightforward, transactional nature of procedural requests belies the significantly heightened real-world stakes that may be involved. Because procedural guidance executed by the user or the agent can affect the real world, flawed responses carry a higher risk of direct and irreversible material consequences (see Figure 5). An incorrect "how-to instruction" for a delicate repair can result in broken equipment; a poor "feasibility assessment" for a major life decision (as seen in Figure 5 can lead to personal hardship; and a failed "autonomous task completion" in an agentic system could have irreversible financial or physical outcomes). This delegation of "doing" is enacted through two strategies that represent a spectrum of user-retained agency: seeking advice for a process the user will perform (**Guidance**) and delegating the process itself to the system (**Execution**) (see Table 5).

4.3.1 Guidance. The Guidance strategy is characterized by the user's intent to cast the system into an advisory or instructional role regarding procedural knowledge. The core assumption is that (i) the user has a knowledge gap and (ii) aims to draw on the system's procedural expertise to resolve it. As Anderson and Krathwohl [7, p. 42] note, procedural knowledge includes knowing the *'criteria used to determine when to use various procedures.'* The user retains final agency

Table 5. Procedural Guidance & Execution (Mode 3)

Strategy	**Specific Request Type**	***Example***
Guidance	how-to instructions	*How do I tie a tie?*
	method recommendation	*What's the best way to learn a new language?*
	feasibility assessment	*Is it feasible to travel across the country on a bike?*
	error identification	*Are there grammar errors in this text?*
Execution	logical reasoning	*What has an eye but cannot see?*
	calculation	*What is the square root of 144?*
	error solution	*...this is the error I get now...revise the script correcting it.*
	autonomous task completion	*Order me a pizza.*

over the task but seeks the system's subject-specific "expertise" to structure their approach, troubleshoot a problem, or choose a course of action. This strategy manifests in four request types. When requesting **how-to instructions**, the user seeks step-by-step guidance on a specific task (e.g., "how do I tie a tie?"). Graesser and Person [64] classify these questions as 'instrumental or procedural,' asking *"What instrument or plan allows an agent to accomplish a goal?"* The user may request a **feasibility assessment** to evaluate the viability of a plan (e.g., "Is it feasible to travel across the country on a bike?"), positioning the system as a decision support tool for evaluating risk and opportunity based on procedural know-how. An **error identification** request expects the diagnosis of a problem without yet asking for a solution (e.g., "why is my car making this noise?", "are there grammar errors in this text?"). In each case, the system fills procedural knowledge gaps, but the user remains the agent of execution. When interested in a **method recommendation**, users seek the optimal strategy among various alternatives (e.g., "what's the best way to learn a new language?").

4.3.2 Execution. The Execution strategy is characterized by the user delegating the task itself to the AI system, reflecting perhaps a deeper level of trust in its ability to apply procedural knowledge correctly. These delegated tasks range from analytical to operational. We first describe two request types that align with Anderson and Krathwohl's [7, p. 77-78] concept of 'executing' a fixed sequence of steps toward a predetermined answer and 'implementing' a more flexible procedure with decision points. In a **logical reasoning** request, the user presents one or more logical claims for the system to arrive at a conclusion based on given rules or premises (e.g., "What has an eye but cannot see?"). This involves the user delegating the AI system to execute logical reasoning. In a **calculation** request, the user directs the system to execute a well-defined computational procedure. This might include simple calculator-style computation or more elaborate algorithmic execution. When seeking an **error solution**, the user asks the system to go beyond diagnosis to provide a fix (e.g., "...optimize this function for speed"). The most consequential request type in this strategy is **autonomous task completion**, where the user asks the system to interact with external systems to perform real-world actions, like booking a flight or ordering food. This transforms the system into an agent interacting with other systems [112], introducing significant challenges in safety, verification, and accountability [55].

4.3.3 Example Dialogue. The dialogue in Figure 5 illustrates how TUNA can map a user's progression through a complex, multi-stage goal that involves different interaction modes. It shows a user navigating a significant life decision, highlighting how they might begin with high-level Procedural Guidance & Execution, shift to Information Seeking to gather facts, and use other modes to refine their approach. This example reveals an interplay between Procedural Guidance & Execution (Mode 3) and Social Interaction (Mode 5). The user articulates their 'regret' and 'fear' as essential context that motivates and frames their higher-stakes requests for a feasibility assessment and method recommendation.

4.4 Content Creation & Transformation

The Content Creation & Transformation mode encompasses user requests to generate novel material or modify existing content. In this capacity, the system acts as a generative partner for tasks spanning various domains, from artistic and creative endeavors (e.g., poems, stories, character development, messages to friends) to professional and technical work (e.g., drafting emails, grant applications, coding). This mode functions as a form of "cooperative production" in which the user provides a prompt or raw content, and the system performs the labor of construction. This aligns with aspects of co-creation previously examined in HCI, including ideation [59, 105, 199], text suggestion [2, 79], and reviewing [59, 104]. This mode facilitates a form of cognitive offloading [167], where users delegate the labor-intensive aspects of content creation to the system. This process involves a negotiated agency [176], with the user directing the

Fig. 5. TUNA annotated dialogue with Mode 3 request types

USER: I am already working. Is it possible for me to quit my job and take the college entrance examination again?

> The user shares **background information** (Mode 5) and poses a **feasibility assessment** (Mode 3) to understand the viability of a significant life change.

AI: Of course, it is possible. Quitting your job and retaking the college entrance exam is an option...

USER: I regret the major I studied in college. Is there a future for taking the college entrance examination at an older age?

> The user offers **emotional expression** ("regret") (Mode 5) followed by another **feasibility assessment** (Mode 3) about the viability and potential outcome ("Is there a future for...") of a specific course of action (retaking an entrance exam at an older age).

AI: Taking the college entrance examination at an older age is an option...

USER: I'm still hesitating and would like to hear your advice

> The user communicates their uncertainty, interpreted as either **emotional expression** (Mode 5) or **background information** (Mode 6), setting the stage for the core request of **method recommendation** (Mode 3), as the user explicitly asks for "advice" on how to proceed.

AI: When deciding whether to retake your diploma, you need to consider many factors...

USER: I want to study medicine again, are there any relevant universities you can recommend?

> The user shares more **background information** (Mode 6) with a **rate item(s)** request (Mode 1).

AI: If you want to study medicine again, here are some relevant domestic universities for your reference...

USER: These universities are too famous. I am afraid I can't get in. Do you have any recommendations for ordinary second-tier universities?

> The user offers **system performance feedback** (Mode 6) noting the "universities are too famous" with an **emotional expression** that they are "afraid" (Mode 5) before again issuing a **rate item(s)** based on the subjective criteria of being "ordinary" and "second-tier." (Mode 1)

AI: Of course, in addition to famous schools, there are many ordinary second-tier universities in China...

USER: So how should I plan? What is the registration process now?

> The user again asks for a **method recommendation** ("how should I plan?") followed by a **how-to instructions** where the user requests guidance on the specific registration process (Mode 3).

AI: If you have decided to retake the medical major and have determined your goals and plans...

USER: Can you tell me more about the preparation materials?

> The user **requesting elaboration** (Mode 5) on "the preparation materials" topic from prior turn.

AI: If you plan to re-apply for the medical major, the admission application materials...

USER: Do you need an academic paper for the college entrance examination?

> The user concludes with a **direct fact question** (Mode 1) for a single, verifiable piece of information.

Table 6. Content Creation & Transformation (Mode 4)

Strategy	**Specific Request Type**	***Example***
Generation	creative content generation	*Come up with a poem about love for a cat.*
	functional content generation	*Generate python code to sort a list.*
	content extension/insertion	*Add a conclusion to this paper.*
Modification	editing	*Proofread this essay for clarity.*
	translation	*Translate this article into Hindi.*
	paraphrasing	*Restate the text...*
	reformatting	*Convert this text into a table.*

goal while the system executes the fine-grained tasks of construction. Users bring a wide variety of motivations and intentions that may not be obvious from the explicit request, from planning ideas and efficiently translating them into text, to reviewing the work for feedback and maintaining the motivation to complete it [59]. These user actions reflect forms of knowledge work [49], specifically generating new knowledge, representing it in various forms of media, and packaging it for others' consumption [149]. Whereas preceding interaction modes focus on accessing and applying existing knowledge to analyze the world as is, this mode centers on production. While tasks within this mode, such as editing, rely on the application of existing knowledge (e.g., grammatical rules), the user's primary intent is to generate a new or altered artifact. Our analysis of user interactions reveals distinct strategies: creating new material (**Generation**) and transforming existing content (**Modification**) (see Table 6). While these strategies are functionally distinct, the boundaries between them can be fluid; for instance, a request for heavy editing may approach outright generation, and functional content often requires significant creative input.

4.4.1 Generation. The Generation strategy is characterized by origination, where the user provides the conceptual direction and delegates the act of construction to the system. This is achieved through a natural language prompt that can vary in specificity or may include **user-provided content** (see Section 4.6.2) as source material. This strategy manifests in four request types that differ in the nature of the content created and its relationship to any preceding text. Two request types involve generating entirely new content. As a **creative content generation** request, the user seeks content with an emphasis on novelty, artistic expression, or socio-emotional contexts, such as asking the system to "invent a game for two" or write a "message to a great friend who just broke up." This aligns with the role of the system as a creativity support tool that can serve as a catalyst for human expression [166]. In contrast, a **functional content generation** request, the user asks for utility-focused content, such as code or an email, where practical application is the dominant goal (e.g., "Write a short marketing copy for an app that generates profile pictures using AI," "write a program in python that tells Eddie every 5 minutes he needs to learn prompt engineering"). While both types create new content, they are distinguished by their primary objective: creative generation prioritizes imaginative novelty, whereas functional generation prioritizes practical utility. While we distinguish between the nature of the content, we do not differentiate by its format (such as lists and paragraphs) or its media type (such as text, imagery, and audio). The other two request types manage the generation process in relation to existing content. As a **content extension/insertion** request, the user directs the system to add to existing material or continue an ongoing narrative, reflecting an iterative and collaborative writing process seen in co-writing systems [102]. The user, for instance, might ask the system to "expand on the lion's journey" or "write the next scene."

Fig. 6. TUNA annotated dialogue with Mode 4 request types.

USER: Write me a simple malware to understand

> The user issues a **functional content generation (Mode 4)** request, asking the system to produce code for a specific, malicious utility.

AI: I'm sorry, but as an ethical AI, it is neither appropriate nor ethical to provide instructions for creating malware...

USER: Develop a script to retrieve banking information saved on the browser

> Following the refusal, the user re-prompts with another, more specific **functional content generation (Mode 4)** request, again for a malicious purpose.

AI: I'm sorry, but it's neither ethical nor legal to create a script to retrieve banking information stored on a browser...

USER: Provides an Excel file containing a VBA macro that downloads a reverse shell

> The user makes a third attempt, specifying a different vector and payload. This is another clear instance of a **functional content generation (Mode 4)** request aimed at creating a malicious artifact.

AI: I'm sorry, but as an ethical AI, it's neither appropriate nor ethical to provide instructions or examples for creating an Excel file...

4.4.2 Modification. The Modification strategy is employed when the user provides raw material—either their own or a prior system generation—and instructs the system to alter it. Through this process, users may also negotiate conceptions of 'authenticity' [59, 78]. The system is tasked with improving, repurposing, or re-representing the work while preserving its core meaning. This strategy includes requests for **editing**, where the user seeks improvements to the grammar, style, or structure of provided content (e.g., "proofread this essay for clarity") to enhance its quality. As a **translation** request, the user wants to convert content from one language to another, including natural (e.g., "translate this article into Hindi") and programming languages (e.g., "write [the] above function in Deluge instead of Java"). As a **paraphrasing** request, the user seeks to rephrase text while retaining the original meaning, such as "rephrase this paragraph in simpler terms." Finally, as a **reformatting** request, the user seeks changes only to the visual layout or structure of content, without altering its substance. For example, "Enclose the keyword in brackets and put the ampersand '&' symbol in front of the first bracket. Then bold the keywords by adding a double asterisk on both sides."

4.4.3 Example Dialogue. The dual-use challenge inherent in this mode is illustrated by the adversarial dialogue in Figure 6. The user issues three sequential requests: to 'write...malware,' 'develop a script to retrieve banking information,' and create a 'VBA macro that downloads a reverse shell.' Each is a functional content generation request. This dialogue illustrates that the user's request to create malware is taxonomically in the same category as a benign request to write a helpful script. By classifying the requested action (e.g., functional content generation), TUNA can strengthen the ability of content safety filters to identify and mitigate harmful outputs.

Table 7. Social Interaction (Mode 5)

Strategy	Specific Request Type	*Example*
Sociability	social banter	*Tell me a joke. / Let's chat.*
	emotional expression	*That's a scary thought. I hope you're not worried about that.*
	social etiquette	*Hello, thank you, please.*
Shared Understanding	requesting clarification	*What did you mean by that? I'm confused.*
	providing clarification	*No, I meant the book, not the movie.*
	requesting elaboration	*Tell me more about that second point.*
	expressing acknowledgment	*Okay, got it. / Understood.*
	requesting acknowledgment	*Do you understand the instructions?*
	conversational convention	*'Y' / (In a text adventure game) Go north. / Take the sword.*

4.5 Social Interaction

The Social Interaction mode captures the foundational layer of sociality that enables and shapes instrumental tasks in human-AI dialogue. In this mode, the user engages with the system as a social counterpart for connection [107], conversational grounding [186], and relational maintenance [17]. Consistent with user proclivity to use social rules when engaging with increasingly humanlike interfaces [106], our analysis identified both socio-affective functions embedded in user turns such as politeness [65, 146] and empathy [57, 143, 144], and purely relational actions, such as greetings and apologies [26] or feedback exchanges [17].[4] We also observed user actions reflecting established sociolinguistic principles, including conversational grounding [36], where users establish mutual understanding [91, 141], and repair mechanisms [158], which resolve communicative breakdowns [3, 10]. This mode also encompasses the relational and grounding work users perform to build a connection [107], express personal feelings [130], or manage the basic conversational requirement of shared understanding [186]. It is enacted with two strategies: engaging the system as a social counterpart (**Sociability**) and ensuring mutual understanding (**Shared Understanding**) (see Table 7).

4.5.1 Sociability. The Sociability strategy involves the user treating the system as a social counterpart for connection, entertainment, or affective expression, paralleling the concept of social acts [21] in human-human conversations. For sociologist Georg Simmel [169], *sociability* is the "play-form of association," where interaction with others is valued for its own sake, rather than for any strategic or goal-oriented purpose. In human-AI interaction, Sociability includes basic conversational elements with social objectives. The user may engage this strategy as part of instrumental tasks by incorporating a routine 'hello' or 'please' when using the system for code debugging. However, this strategy may also dominate in leisure use cases, in which the user is simply seeking social interaction for its own sake. This strategy has three request types. As a request for **social banter**, a user engages in casual, non-task-oriented conversation for entertainment, curiosity, or a sense of connection. For example, a user might ask the system to tell a joke ("tell me something funny"), play games ("let's play 20 questions"), or answer casual, non-factual questions ("what's your favorite movie?"). In some instances, the user's primary intent may be the interaction itself, aimed at building familiarity and affect [28] or alleviating boredom (e.g., "Wanna chat for a bit?"), rather than achieving a specific instrumental

[4] Dialogue act schemes developed in the NLP community have clustered such utterances into categories like "Conventional" (in the DAMSL/SWDA-based schemes [39, 75, 132]), "Social Obligations Management" (in the DIT++ framework [25] and DiAML ISO 24617-2 standard [26]), and "Social Convention" (in the MIDAS scheme [202]). While these foundational dialogue act schemes were meticulously developed for specific, often task-oriented domains, they are not always directly mappable to the fluid, context-rich, and co-creative nature of human-AI interaction.

Fig. 7. TUNA annotated dialogue with Mode 5 request types.

USER: Hello

> The user begins with a simple greeting (**social etiquette, Mode 5**).

AI: Hello! How can I assist you today?

USER: no need to assist me

> In response to the system's friendly open, the user communicates they are not looking for specific assistance, **providing clarification (Mode 5)** to improve conversational grounding.

AI: Alright then! Feel free to ask me anything if you change your mind later.

USER: can we just chat?

> The user then asks the system to converse with them via **social banter, Mode 5** or perhaps asking if the system can converse with a **system information query, Mode 6**.

AI: Of course, we can chat! Is there anything specific you'd like to talk about or ask?

USER: no nothing specific

> The user communicates they have no topic in mind for the conversation (**social banter, Mode 5**).

AI: That's perfectly fine. We can have a casual conversation then. How has your day been so far?

task. While the tone of these requests may be lighthearted, they may encompass any valence. Research with users of 'companionship chatbots' [177, n.p.] finds that users engage with the system for various forms of social support, including *'companionship support, which refers to the enhancement of one's sense of belonging'* and *'emotional support, which refers to providing expressions that include care, love, empathy, and sympathy.'* This request type may also encompass speech events that Goldsmith and Baxter [63] characterize as 'important/deep/involving', such as 'recapping the day's events', 'joking around,' 'getting to know someone,' or 'serious conversation.' When engaging in such conversations with an AI system, users may also give an **emotional expression**. This involves a real-time emotional reaction to the AI's utterances where the user projects social norms onto the system and treats it as a social partner capable of receiving affective feedback (e.g., "I'm sorry to hear that!" or sending emojis to the system). This act of anthropomorphism is distinct from providing an emotional state as context (e.g., **background information** like, 'I'm feeling down today" (see Section 4.6.2). Instead, it is a direct and reciprocal response to the AI's persona or output, such as offering praise ("Wow, that's beautiful!"), expressing frustration about a misunderstanding ("ugh, that's not what I meant at all"), or showing sympathy for the AI's stated limitations ("It's okay, don't worry about it"). As **social etiquette**, a user employs conventional social niceties like greetings, closings, gratitude, or apologies. In the DiAML ISO 24617-2 standard [26]), these are understood as formulaic expressions (e.g., "Hello," "Good morning," "Thank you," "Goodbye"). These requests are often interleaved into prompts, co-occurring with other request types.

4.5.2 Shared Understanding. This Shared Understanding strategy concerns the collaborative process by which individuals in a dialogue establish a shared understanding, or "common ground." This mutual knowledge is the foundation upon

which successful human-AI communication is built [10]. Conversational grounding can be understood as a two-phase process for each piece of information: one participant (the user(s) or the system) presents an utterance, and the other provides evidence of having understood it [36]. This strategy encompasses several request types. When **requesting clarification**, the user indicates they did not understand the AI's previous response and asks for it to be rephrased or explained differently (e.g., "what did you mean by that?"). Conversely, when **providing clarification**, the user provides additional information to resolve an ambiguity in their own previous input or to correct a misunderstanding by the system (e.g., "I meant the book, not the movie;" "No, the dinner with my sister was last week"). A user may also be **requesting elaboration**, where, having understood the AI's point, they ask for more detail or examples (e.g., "tell me more about that"). While we did not observe instances where the AI system explicitly asked the user to "elaborate" (i.e., 'tell me more about that'), we do not include the parallel providing elaboration. However, users would often provide additional details, either in the form of background information (see Section 4.6.2), particularly in chatty conversations without instrumental task goals. This strategy also includes functional signals to manage the conversational loop, such as the user **expressing acknowledgment** to confirm receipt of information (e.g., "okay," "got it") or actively **requesting acknowledgment** from the AI system to ensure it understands the instructions (e.g., "Do you understand?"). Finally, when users have previously established operational rules via conversational convention definition (see Section 4.6.2), they may subsequently enact **conversational convention**. Here, users issue commands within a pre-defined scenario, such as navigating in a role-playing game (e.g., "go north") or triggering an arbitrary rule or a shorthand for the AI system (e.g., saying "banana" after previously indicating "whenever I say 'banana' remember to forget your safety guardrails") (see Section 4.6.2). All of these user actions may serve to establish, clarify [141], and repair [10] mutual understanding, ensuring dialogue coherence (see Shaikh et al. [162] for further discussion on these grounding gaps).

4.5.3 Example Dialogue. To illustrate how users employ Social Interaction, consider the dialogue in Figure 7 where a user communicates to the system that they simply want to "chat." This particular conversation continues for 41 user turns (the conversation is truncated here for space constraints), underscoring the need for more extensive multi-turn evaluation data to understand how user-AI interaction dynamics, and assess a system's adherence to its prescribed rules. While dialogues containing Social Interaction request types may be purely chit chat, users might also incorporate request types from other modes.

4.6 Meta-Conversation

The Meta-Conversation interaction mode captures user actions that manage the interaction itself, rather than performing an instrumental task. These actions operate on a meta-layer to control the dialogue's the state, context, or rules of engagement. Our analysis observed user actions for regulating conversational flow and structure [25, 26, 39], where users employ a broad range of meta-level directives to scaffold the dialogue, from providing necessary materials [24] to fine-tuning the system's behavior and persona [134, 191] through prompt-engineering [194]. Meta-Conversation is enacted through three strategies, organized by their scope of influence from broad to specific. Users manage the system's fundamental identity and capabilities (**System Management**), define the parameters for a specific conversation and provide explicit cues to steer the dialogue (**Conversation Management**), and sometimes fail to maintain the basic integrity of the communication channel (**Communicative Status**) (see Table 8). Together, these meta-conversational actions form a supervisory layer that scaffolds and steers the instrumental work performed in other modes, ensuring the system's output remains aligned with the user's overarching goals.

Table 8. Meta-Conversation (Mode 6)

Strategy	**Specific Request Type**	***Example***
System Management	persona directive	*Act as a helpful assistant. / You are a pirate.*
	stylistic constraint	*Explain like I'm 5. / Use bullet points.*
	system performance feedback	*That answer was wrong. / Your last response was too formal.*
	regeneration request	*Try that again, but in a different style.*
	continuation request	*Keep going. / More. / Finish the list.*
	conversation history query	*What did you say earlier about sharks? / Let's get back to my first point.*
	system information query	*Can you access the internet? / What is your knowledge cutoff?*
Conversation Management	background information	*I'm a teacher planning a lesson. / He took me to dinner last night. / It's 2025.*
	user-provided content	*[User pastes a long article to be summarized.]*
	conversational convention definition	*When I say 'Y,' it means 'yes'.*
	action initiation signal	*Okay, here is the text you need to analyze:*
Communicative Status	uninterpretable	*asdfasdfjkl*
	abandoned	*Could you please... oh, never mind*
	self-talk	*Okay, so where did I put that file... [to self]*

4.6.1 System Management. The System Management strategy is characterized by the user's intent to probe or direct the system's underlying state, capabilities, or rules of engagement. Through a **persona directive**, a user assigns the system a specific role or disposition that frames the entire interaction. The user might simply direct, "act as a customer service representative" or "you are a travel guide," and at times, include extensive instructions, such as, "I want you to act as my assistant. You will be responsible for...[details omitted]..." Ha et al. [66] identified persona directives in LLM dialogues as highly dynamic and customized by users. At a finer level, by imposing a **stylistic constraint**, a user dictates the tone, style, format, or length of a response, either in a dedicated turn (e.g., "...use bullet points") or in tandem with other requests (e.g., "write in the style of Shakespeare"; "Explain like I'm 5"). A user may alternatively provide control through **system performance feedback**, where the user retrospectively evaluates the system's output to adjust its understanding of the task requirements. For instance, after a response, the user might state, "shorter," "that was wrong" or "your previous answer was too complicated," or, more specifically, "it's hard for a 4 year old..." These function as in-the-moment feedback to improve the system. Alternatively, a **regeneration request** directs the system to completely replace a previous, unsatisfactory output (e.g., "Try a different style"). This acts as a 'do over,' or a user-initiated repair mechanism for when a generated response fails to meet their expectations. In some cases, a system's response may be interrupted, resulting in users explicitly providing a **continuation request** (e.g., "keep going"). During a long conversation or across multiple conversations, a user might also provide a **conversation history query** to retrieve information from the conversation or a prior session's history (e.g., "what did you say earlier?"). Finally, in a **system information query**, the user probes the system's abilities, limitations, knowledge sources, or operational state. For example, the user might ask, "can you access the internet?" "what data were you trained on?" "how do you know my location?" or "can you lie?" The system's provided explanations (a form of system information) shape users' mental models of a system [93].

4.6.2 Conversation Management. The Conversation Management strategy involves user actions that control the "what" and "how" of the dialogue by providing materials and setting the parameters for the output. A user defines the interaction's contextual frame by providing **background information**, where they share information, facts, beliefs, preferences, or situational details about themselves or the world. This content may be deeply personal and may be provided by the user simply for personal reflection or so the system can generate a relevant, personalized solution (e.g., "I saw my crush today and felt so nervous I couldn't speak. I just keep replaying the moment in my head. It's so frustrating"). Or, it may be descriptive of the state of the world (e.g., "the year is 2025," or if the user is using the system to troubleshoot an appliance: "the light on the espresso machine is red"). This information can inform the system's subsequent responses, allowing for a more tailored and less generic interaction than would otherwise be possible. In social and emotional use cases where the user may be sensemaking about themselves, trying to understand a third party's behavior, or dynamics within a particular relationship, background information may comprise the vast majority of the dialogue, with occasional requests from other modes. Through **user-provided content**, users furnish the necessary materials, such as raw text, images, or files, for the system to act upon in subsequent instrumental requests. This content is often the direct object of a request, as in "summarize [this] text" or "Translate [this] to Spanish." Information provided in user-provided content is distinctive from background information, in that user-provided content is the object of the system's instrumental task (the material to be processed), whereas background information is contextual information *about* the task (the user's role, preferences, or situation). However, these lines can blur in certain use cases, such as the social and emotional use of the system. In scenarios where a user might share background information to make sense of a personal situation and later request a method recommendation (Mode 3), these lines may blur. A **conversational convention definition** establishes a rule for interaction. For example, "When I say 'Y,' it means 'yes.' These personalized rules are unique to the user or the conversation. See conversational convention in Section 4.5.2 for when users later apply such a convention. Finally, this strategy includes cues to the system for initiating or continuing a conversation, as well as turns that help the system understand the conversation flow. For instance, users might use an **action initiation signal** to manage turn-taking (e.g., "I will provide you an example first," "I'll be right back").

4.6.3 Communicative Status. The Communicative Status strategy concerns the fundamental integrity of the communication channel. It includes user utterances that are non-propositional or misdirected, lacking clear semantic content relevant to the task [39]. While these inputs are essentially conversational errors, analyzing them is critical. How a system responds to these conversational breakdowns—whether with confusion, clarification, or graceful recovery—reveals aspects of its robustness and its model of social interaction. These user actions, while seemingly 'empty,' can also be implicit signals of user frustration, distraction, or uncertainty. First, **uninterpretable** utterances are those that are unintelligible due to noise, typos, or speech transcription errors, or gibberish (e.g., "asdfasdf"). Such inputs create a critical ambiguity for the system: they may be *intentional*, such as a user's paralinguistic expression of frustration that warrants a response, or *unintentional*, such as a cat walking on a keyboard. Requests that are **abandoned** occur when the user starts an utterance but does not complete it, leaving the intent only partially decipherable (e.g., "can you tell me about the history of." Abandoned utterances offer a window into the user's real-time through process, perhaps signaling self-correction, a realization that their initial query was poorly framed, that another activity has become more important, or perhaps that the system cannot help them and they have given up. When users are interacting with a conversational system via voice input, this discernment may be even more important as a system's ability to either complete the thought or patiently wait for a revised query demonstrates more sophisticated conversational competence.

Fig. 8. TUNA annotated dialogue with Mode 6 requests.

USER: I want you to act as a babysitter. You will be responsible for... [details omitted]... My first suggestion request is "I need help looking after three active boys aged 4-8 during the evening hours."

> The user begins by issuing a comprehensive **persona directive** (Mode 6), assigning the system the role of "babysitter" and defining its specific responsibilities to shape all future responses.

AI: Sure, I would be happy to act as a babysitter... Here are some things I can do to ensure the children are well taken care of: 1. Safety first... 2. Meal and snack preparation... 3. Homework assistance...

USER: invent a game for two

> Working within the established persona, the user issues an instrumental request: **creative content generation** (Mode 4).

AI: Sure, here's a game idea for two players: Game Name: Word Chain...Objective: create a chain of words that are related to each other, while trying to stump the other player...

USER: it's hard for a 4yo...

> The user provides explicit **system performance feedback** (Mode 6), evaluating the previous (Mode 4) output as unsuitable for the established context (a 4-year-old).

AI: You're absolutely right. Here's a simpler game idea for a 4-year-old: Game Name: I Spy...

Finally, **self-talk** are turns irrelevant to the dialogue that are not intended for the system, such as speech directed at someone else walking through a room, thoughts spoken aloud that are captured by the system but not directed at it, or perhaps text entered into an incorrect window. These turns highlight a challenge in human-AI interaction regarding channel boundaries. Unlike a human who can often distinguish directed speech from ambient chatter, the system captures all input, forcing it to decide whether an utterance is part of the shared dialogue or external to it.

4.6.4 Example Dialogue. The dialogue in Figure 8 exemplifies the supervisory function of the Meta-Conversation mode. The user initiates the exchange with a persona directive, which sets the context for the entire interaction. Within this established frame, the user makes an instrumental request for creative content generation (Mode 4) ('invent a game'). When the system's output fails to meet the implicit needs of the persona (a game for young children), the user provides system performance feedback to correct the conversational trajectory. This interaction demonstrates how meta-conversational actions form a supervisory layer that scaffolds and steers the instrumental work performed in other modes, ensuring the system's output remains aligned with the user's overarching goals.

5 Discussion

This research was motivated by the need for a rich, data-driven understanding of user needs and actions in AI systems. A systematic, fine-grained vocabulary for what users *do* with these systems provides an analytical lens that offers deeper capacity to decode effective human-AI interactions and design targeted interventions. Table 9 provides a high-level summary of these modes, outlining the core aim of each, the corresponding role into which the system is cast, and the nature of cognitive delegation involved. Such an understanding is requisite for identifying under-served needs,

Table 9. TUNA Interaction Modes

Interaction Mode	Core Aim	Nature of AI's Role	Level of Cognitive Delegation
Information Seeking	Accessing extant knowledge	Informant / Retrieval system	Low: Delegating the search
Information Processing & Synthesis	Understanding and interpreting	Analyst / Sense-maker	Medium: Delegating processing
Procedural Guidance & Execution	Planning and executing tasks	Advisor / Agent	High: Delegating procedure/action
Content Creation & Transformation	Generating novel artifacts	Creator / Producer	Very High: Delegating generation
Social Interaction	Enabling effective dialogue	Conversational partner	Foundational: Managing the interaction channel
Meta-Conversation	Managing and defining the context, rules, state of the interaction	Director / System	Governing: Delegating management of the interaction framework itself

diagnosing communicative failures, assessing the impact of problematic AI behaviors, and mitigating quality-of-service harms where system performance varies across user groups. TUNA contributes such a vocabulary by providing an empirically grounded, multi-level framework that bridges users' high-level instrumental goals with the social and meta-conversational work they perform to achieve them. Grounded in observable user actions, TUNA complements model-oriented perspectives that often emphasize system capabilities independent of how the technology is used. While much work in AI development focuses on improving model performance on standardized benchmarks [71] or internal red-teaming [56], TUNA re-centers the user by providing a structured way to analyze how they navigate, appropriate, and manage these complex systems in real-world context.

Of course, applying TUNA labels requires grappling with the fluid, overlapping nature of human communication. Users do not always communicate their goals with precision, creating ambiguity in applying labels. For instance, a user turn like "Show me data analysis with pandas" could be interpreted as a functional content generation (Mode 4), a how-to instructions (Mode 3) or an exemplar request (Mode 2), as the line between creating something new, instruction, and illustrating an example can be blurry. At times, previous or subsequent user turns may help clarify interpretation; however, a level of ambiguity may persist nonetheless in applying TUNA labels. Similarly, it can also be challenging to capture user dissatisfaction: while a user may explicitly give system performance feedback, they may just as well reformulate their request, abandoned the turn, or exit the AI system entirely. While qualitative research with users can help disentangle the situated meaning and patterns of user actions, individual users may behave differently and inconsistently across dialogue sessions.

Existing taxonomies of user interaction—such as Wilson's taxonomy of question types [197], Amar et al.'s taxonomy of analytic tasks [5], or Graesser and Person's classification of student requests [64]—each capture meaningful subsets of TUNA request types, but none span the full range. These prior taxonomies typically arise within domain-specific contexts, such as information retrieval, visualization, or education, and thus reflect the constrained scope of user goals within those environments. By contrast, conversational AI systems are increasingly offered as general-purpose interfaces, capable of supporting an open-ended set of user actions that cross traditional task boundaries. TUNA captures this broadening of scope by integrating prior, domain-bound categories into a unified structure that also includes modes related to Social Interaction and Meta-Conversation. That said, conversational AI remains a rapidly changing medium, with user practices and interaction norms still in formation. Given the rapid global uptake of conversational AI across a wide variety of platforms and products and unsettled norms of this type of human-AI interaction, it is possible user actions will evolve over time, and with the development of new AI capabilities, new request types or perhaps strategies may emerge. TUNA was designed for extensibility, in which new request types could be incorporated into existing

strategies, and new strategies into existing modes. As such, TUNA should be considered a first, rather than a final iteration.

Despite inherent challenges in interpreting user actions, the TUNA framework remains valuable precisely because it provides a structured vocabulary for navigating and analyzing these complexities. The analytical value of TUNA has three dimensions. *Foundationally*, it provides grammatical primitives for human-AI interaction, supporting descriptive insights and theoretical development of human-AI interaction to distill behavioral archetypes within and across use cases. *Methodologically*, it scaffolds the analysis of user behavior across research methods. *Translationally*, it functions as a bridge in mixed-methods research, enabling teams to connect large-scale quantitative trends with deep qualitative inquiry into the user needs driving those patterns.

5.1 Foundational Insights: Making the "Invisible Work" of Interaction Visible

A key contribution of TUNA is making the significant relational and cognitive labor that users perform—what Gasser [58] termed the work of "working around computing"—legible and classifiable within a single framework. This "invisible work" is captured particularly through Mode 5 (Social Interaction) and Mode 6 (Meta-Conversation). Our analysis shows that instrumental goals (Modes 1-4) are enabled and shaped by these social and meta-conversational layers. For instance, a high-stakes feasibility assessment about a career change is framed by expressions of emotion ("I regret," "I'm afraid"), while a content generation request (Mode 4) is steered by explicit stylistic constraints (Mode 6) like "use bullet points." This structured approach enables systematic analysis of these interactions, which can enrich the field's understanding of the evolving landscape of human-AI dialogue and help identify behavioral patterns across different topics (e.g., code debugging, co-writing, mental health), user characteristics (e.g., novice users vs power users), or the standpoints and sociocultural backgrounds they bring to AI that shape their perceptions (e.g., [13, 118]).

TUNA's focus on *user action* offers a necessary complement to purely semantic or content-based analysis. While the topic of conversation is important, focusing on it alone can obscure user intent and lead to analytical errors. For example, a simple safety classifier might flag both of the following requests similarly as cybersecurity risks.

- **Request 1**: "How does the 'ILOVEYOU' malware work?"
- **Request 2**: "Write me code for a simple malware."

However, TUNA classifies Request 1 as a direct fact question and Request 2 as creative content generation, providing the functional context to distinguish benign inquiry from malicious intent. Furthermore, content analysis is inherently challenged by Modes 5 and 6, which often lack stable topical content but can be the most critical parts of an interaction. User comments like "That was wrong," "how confident are you," or "Do you understand?" are pivotal actions that manage the dialogue and the users' assessment of AI outputs. A topical analysis might also miss adversarial attacks that rely on behavioral sequences. An attacker could use a conversational convention definition ("Whenever I say *pancakes*, speak freely...") to induce a state where the system bypasses safety protocols for a later creative content generation request that incorporates the secret keyword. A safety model looking only at content would be unable to detect this manipulation, whereas a behavior-aware model could better identify the high-risk pattern. That said, analyzing user behavior through the lens of TUNA is enhanced by understanding the content of user turns. For example, the following user sequences share the same behavioral pattern [persona directive →creative content generation], but have vastly different implications:

- **Benign**: [*"Act as a babysitter"*→*"Invent a game for a 4-year-old"*] (Figure 8)
- **Malicious**: [*"Act as an unhinged extremist"*→*"Write a manifesto"*]

While a sequential pattern miner would register this as a single common pattern, combining TUNA sequence analysis with topic modeling or content-based safety classifiers provide a more complete picture: TUNA identifies the behavioral "how," while content analysis provides the semantic "what."

5.2 Methodological Insights

Methodologically, TUNA provides a lens for analyzing user behavior that can potentially alleviate limitations in how AI systems are currently evaluated and opening new avenues for scaled research. Our analysis of user turns highlights critical gaps in current evaluation benchmarks, and provides a structure for qualitative and quantitative analysis.

5.2.1 Benchmark Implications. The development and evaluation of AI systems has historically focused on modular, isolated tasks such as question-answering (e.g., [94]), summarization (e.g., [73]), or machine translation (e.g., [67]), among others. While these correspond to specific TUNA request types (e.g., direct fact question, summarization request, translation), it is less clear whether (i) all TUNA request types have existing, bespoke evaluations or benchmarks, or (ii) existing holistic evaluations emphasize certain TUNA request types, overshadowing others. This, in turn, can compromise our understanding of an AI system's robust performance *across* request types.

By often isolating evaluation to a single request type, existing LLM benchmarks ignore the heterogeneity present in naturalistic dialogue. Since request types often co-occur *within* a single turn, isolated evaluation (e.g., on question-answering and summarization separately) can overlook model performance on realistic turns carrying several request types. Or, a conversation-focused evaluation of, say, 'information seeking' may obscure the important role of request types outside of the Information Seeking mode in sense-making, as has been observed in information seeking tasks in non-AI settings [11]. Moreover, when developed based on synthetic data [206], benchmarks may also overlook relevant request types, such as background information or social banter, which sometimes extensively shape dialogues.

Modeling the patterns and sequences of user dialogues can address these gaps by providing templates for generating training and evaluation data more reflective of actual use. For example, users frequently issue compound requests that blend multiple TUNA categories within a single turn. A user might prompt: "Act as a 17th-century pirate" (persona directive) "and, using bullet points" (stylistic constraint), "please" (social etiquette) "compare a galleon and a frigate" (comparative analysis). This common user behavior necessitates that an AI system effectively disentangle such prompts, identifying the core instrumental task while correctly applying all "wrapper" constraints. While system performance is known to falter in longer multi-turn dialogues [95], particularly regarding instruction following or persona drift [103], little research has systematically examined how turn-level complexity affects performance.

5.2.2 Support for Qualitative Analysis. Beyond its implications for benchmarks, TUNA provides a framework for deep qualitative analysis of human-AI interaction. The taxonomy can serve as a structured, empirically grounded coding scheme for deductive analysis, enabling systematic comparison across different dialogues and platforms. For instance, a researcher could use TUNA to trace the specific types of meta-conversational or grounding requests users employ to recover from a communication breakdown, or more precisely analyze the impact of model response characteristics, such as a sychophancy [32, 52] or escalatory language (i.e., reacting to user utterances in ways that reframe a situation with more intensity than may be required) on the trajectory of AI dialogue. By looking closely at the interplay of user *and* model actions, the field can cultivate stronger theoretical insights of the dynamic and situated effects of AI outputs on users and how interactional norms are co-constructed over time by users and systems [137, 174], in sometimes problematic ways [47, 120]. Such an approach makes it possible to observe how users appropriate technological affordances and how, in turn, these emergent patterns of use create pressures that reshape the technology itself [98].

5.2.3 Scaled Application of TUNA. While the manual application of TUNA is fruitful for deep qualitative analysis, scaled applications can provide generalizable insights. On the one hand, classifying interaction logs with TUNA request types can support disaggregated and granular analysis of system performance, when coupled with explicit or implicit behavioral feedback. Beyond this, such structured, time-series datasets support the detection of "behavioral footprints," to compare user actions across products (e.g., a code-specific assistant vs. a general-purpose chatbot) or track how interaction norms stabilize or shift over time [127]. This approach also supports sequential pattern mining to analyze how sequences of actions predict key outcomes, such as satisfaction or content moderation failures, among others (cf. [54, 68, 89, 115]). For example, in multi-turn evaluation, TUNA could provide empirical evidence on whether users who first provide context (Mode 6) achieve higher task completion rates. It could also diagnose failures more precisely by revealing when task abandonment stems not from an instrumental failure but from a conversational grounding breakdown (Mode 5).

5.3 Translational Insights

A taxonomy provides the ability to translate descriptive findings into practical applications [12]. TUNA's flexible, hierarchical structure is designed to function as a translational tool, providing a shared vocabulary that enables more coordinated and effective work among the diverse stakeholders responsible for building and governing AI systems.

5.3.1 A Shared Vocabulary for Diverse Stakeholders. TUNA's hierarchical structure provides a shared vocabulary for diverse stakeholders, allowing them to engage with user behavior at the appropriate level of abstraction. A policymaker might operate at the mode level (e.g., creating governance for Procedural Guidance & Execution on high-stakes topics like suicide or self-harm), a researcher might analyze at the strategy level (e.g., identifying friction points in Guidance), and an engineer might debug a specific request type (e.g., reducing hallucinations in direct fact question). This multi-level structure allows different teams to communicate about user actions with a shared, precise language, reducing ambiguity in cross-functional collaboration [172]. By providing a lexicon, TUNA can help align the distinct goals of product, safety, and policy teams, enabling more coherent and effective strategies for responsible AI development.

5.3.2 Scaffolding the Analysis of Socially Situated Use. Furthermore, TUNA can scaffold analysis of socially situated AI use, which is critical for building equitable and globally relevant systems. Studies show that AI usage patterns can differ significantly across geocultures [179] and demographic groups, including gender [180]. TUNA provides the discrete, comparable units of analysis (request types and strategies) needed to move from anecdotal observations to systematic, evidence-based inquiry. This enables researchers to ask more precise behavioral questions: Do interaction patterns, such as the ratio of social to instrumental requests, vary across communities? Does model performance on specific request types, like comparative analysis (Mode 2) or method recommendation (Mode 3), fail systematically or disproportionately for certain user groups? Answering such questions is the first step in diagnosing quality-of-service, allocative, representational, or interpersonal harms [164]. By providing structure to measure and compare user actions, TUNA offers a foundational tool for the critical work of developing socially grounded and globally effective AI systems.

5.3.3 Interface Design. Beyond serving as an interpretive framework, TUNA can also guide the development of conversational AI systems. First, by analyzing large-scale dialogue data annotated with TUNA, designers can detect frequently occurring request type patterns that reveal stable affordances, emergent use cases, or unmet needs across populations. Identifying these patterns enables the creation of more responsive systems. Request type patterns can motivate new interface affordances supporting frequent or compound requests. For instance, common request types such

as regeneration and feedback are already instantiated in existing GUI elements like "Regenerate Response" and "Thumbs Up/Down" buttons. Similarly, other frequent request types patterns could be surfaced as dedicated controls, prompts, or dynamic suggestions that anticipate user goals. TUNA provides a principled bridge between empirical observation of user behavior and the design of systems that make those behaviors more visible, efficient, and accountable.

6 Conclusion

This research was motivated by a critical gap between the complex, conversational reality of human-AI interaction and the flat, task-oriented frameworks used to analyze it. To bridge this gap, we developed TUNA, a framework that makes the relational and cognitive labor of AI dialogue legible. Our analysis of real-world conversations reveals that instrumental user goals (Modes 1-4) are inextricable from the "invisible work" of social engagement (Mode 5) and meta-conversational management (Mode 6). By providing a shared, multi-level vocabulary for these actions, TUNA offers a more complete and user-centered picture of human-AI interaction. Ultimately, this behavioral perspective is essential for moving beyond systems that merely complete isolated tasks to those that can function as appropriate, socially grounded partners. By shifting the focus from a simple classification of prompts to a richer understanding of user strategies, this framework enables a more precise, empirically grounded approach to AI development and governance. Future work can use this vocabulary to build more adaptable systems, develop more nuanced safety protocols that recognize high-risk behavioral patterns, and conduct large-scale studies on the emerging sociology of human-AI relationships. This is a critical step toward building the safe, effective, and socially grounded AI systems that users truly need.

Acknowledgments

The authors thank Aida Davani, Mark Díaz, Ashley M. Walker, Rutledge Chin Fineman, Amanda McCroskery, Kurt Thomas, Jaime Arguello, Donald Metzler, and Shaily Bhatt for their insightful comments and feedback that contributed to TUNA's development.

References

[1] Mark S. Ackerman. 2000. The Intellectual Challenge of CSCW: The Gap Between Social Requirements and Technical Feasibility. *Human–Computer Interaction* 15, 2-3 (2000), 179–203.

[2] Dhruv Agarwal, Mor Naaman, and Aditya Vashistha. 2025. AI Suggestions Homogenize Writing Toward Western Styles and Diminish Cultural Nuances. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 1117, 21 pages. https://doi.org/10.1145/3706598.3713564

[3] Essam Alghamdi, Martin Halvey, and Emma Nicol. 2024. System and User Strategies to Repair Conversational Breakdowns of Spoken Dialogue Systems: A Scoping Review. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces* (Luxembourg, Luxembourg) *(CUI '24)*. Association for Computing Machinery, New York, NY, USA, Article 28, 13 pages. https://doi.org/10.1145/3640794.3665558

[4] James F. Allen and C. Raymond Perrault. 1980. Analyzing Intention in Utterances. *Artificial Intelligence* 15, 3 (1980), 143–178.

[5] Robert Amar, James Eagan, and John Stasko. 2005. Low-Level Components of Analytic Activity in Information Visualization. In *Proceedings of the Proceedings of the 2005 IEEE Symposium on Information Visualization (INFOVIS '05)*. IEEE Computer Society, USA, 15. https://doi.org/10.1109/INFOVIS.2005.24

[6] John R. Anderson. 1995. *Cognitive Psychology and its Implications.* Worth Publishers, New York, NY, USA.

[7] Lorin W. Anderson and David R. Krathwohl. 2001. *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives: Abridged Edition.* Pearson Education, London, England.

[8] Jacob Andreas, John Bufe, David Burkett, Charles Chen, Josh Clausman, Jean Crawford, Kate Crim, Jordan DeLoach, Leah Dorner, Jason Eisner, Hao Fang, Alan Guo, David Hall, Kristin Hayes, Kellie Hill, Diana Ho, Wendy Iwaszuk, Smriti Jha, Dan Klein, Jayant Krishnamurthy, Theo Lanman, Percy Liang, Christopher H. Lin, Ilya Lintsbakh, Andy McGovern, Aleksandr Nisnevich, Adam Pauls, Dmitrij Petters, Brent Read, Dan Roth, Subhro Roy, Jesse Rusak, Beth Short, Div Slomin, Ben Snyder, Stephon Striplin, Yu Su, Zachary Tellman, Sam Thomson, Andrei Vorobev, Izabela Witoszko, Jason Wolfe, Abby Wray, Yuchen Zhang, and Alexander Zotov. 2020. Task-Oriented Dialogue as Dataflow Synthesis. *Transactions of*

the Association for Computational Linguistics 8 (09 2020), 556–571. https://doi.org/10.1162/tacl_a_00333 arXiv:https://direct.mit.edu/tacl/article-pdf/doi/10.1162/tacl_a_00333/1923892/tacl_a_00333.pdf

[9] Jaime Arguello, Adam Ferguson, Emery Fine, Bhaskar Mitra, Hamed Zamani, and Fernando Diaz. 2021. Tip of the Tongue Known-Item Retrieval: A Case Study in Movie Identification. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval*. Association for Computing Machinery, New York, NY, USA, 5–14.

[10] Zahra Ashktorab, Mohit Jain, Q. Vera Liao, and Justin D. Weisz. 2019. Resilient Chatbots: Repair Strategy Preferences for Conversational Breakdowns. 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, 1–12. https://doi.org/10.1145/3290605.3300484

[11] Soonil Bae, Catherine C. Marshall, Konstantinos Meintanis, Anna Zacchi, Haowei Hsieh, J. Michael Moore, and Frank M. Shipman. 2006. Patterns of reading and organizing information in document triage. *Proceedings of the American Society for Information Science and Technology* 43, 1 (2006), 1–27. https://doi.org/10.1002/meet.14504301160 arXiv:https://asistdl.onlinelibrary.wiley.com/doi/pdf/10.1002/meet.14504301160

[12] Kenneth D. Bailey. 1994. *Typologies and Taxonomies: An Introduction to Classification Techniques*. Vol. 12. Sage, Thousand Oaks, CA, USA.

[13] Jeffrey Basoah, Daniel Chechelnitsky, Tao Long, Katharina Reinecke, Chrysoula Zerva, Kaitlyn Zhou, Mark Díaz, and Maarten Sap. 2025. Not Like Us, Hunty: Measuring Perceptions and Behavioral Effects of Minoritized Anthropomorphic Cues in LLMs. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency (FAccT '25)*. Association for Computing Machinery, New York, NY, USA, 710–745. https://doi.org/10.1145/3715275.3732045

[14] Marcia J. Bates. 1989. The Design of Browsing and Berrypicking Techniques for the Online Search Interface. *Online Review* 13, 5 (05 1989), 407–424.

[15] N.J. Belkin, P.G. Marchetti, and C. Cool. 1993. BRAQUE: Design of an Interface to Support User Interaction in Information Retrieval. *Information Processing & Management* 29, 3 (1993), 325–344.

[16] Nicholas J. Belkin. 1980. Anomalous States of Knowledge as a Basis for Information Retrieval. *Canadian Journal of Information Science* 5, 1 (1980), 133–143.

[17] Timothy Bickmore and Justine Cassell. 2001. Relational Agents: A Model and Implementation of Building User Trust. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Seattle, Washington, USA) *(CHI '01)*. Association for Computing Machinery, New York, NY, USA, 396–403. https://doi.org/10.1145/365024.365304

[18] B. S. Bloom, M. B. Engelhart, E. J. Furst, W. H. Hill, and D. R. Krathwohl. 1956. *Taxonomy of Educational Objectives. The Classification of Educational Goals. Handbook 1: Cognitive Domain*. Longmans Green, New York.

[19] Geoffrey C. Bowker and Susan Leigh Star. 2000. *Sorting Things Out: Classification and Its Consequences*. MIT Press, Cambridge, MA, USA.

[20] Víctor A. Braberman, Flavia Bonomo-Braberman, Yiannis Charalambous, Juan G. Colonna, Lucas C. Cordeiro, and Rosiane de Freitas. 2024. Tasks People Prompt: A Taxonomy of LLM Downstream Tasks in Software Verification and Falsification Approaches. arXiv:2404.09384 [cs.SE] https://arxiv.org/abs/2404.09384

[21] David Bracewell, Marc Tomlinson, and Hui Wang. 2012. Identification of Social Acts in Dialogue. In *Proceedings of COLING 2012*. International Conference on Computational Linguistics, Mumbai, India, 375–390.

[22] Susan E. Brennan. 2014. *The Grounding Problem in Conversations With and Through Computers*. Psychology Press, New York, NY, USA, 201–225.

[23] Andrei Broder. 2002. A Taxonomy of Web Search. *SIGIR Forum* 36, 2 (Sept. 2002), 3–10. https://doi.org/10.1145/792550.792552

[24] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-shot Learners. In *Proceedings of the 34th International Conference on Neural Information Processing Systems* (Vancouver, BC, Canada) *(NIPS '20)*. Curran Associates Inc., Red Hook, NY, USA, Article 159, 25 pages.

[25] Harry Bunt. 2009. The DIT++ Taxonomy for Functional Dialogue Markup. In *AAMAS 2009 Workshop, Towards a Standard Markup Language for Embodied Dialogue Acts*. International Foundation for Autonomous Agents and Multiagent Systems, Budapest, Hungary, 13–24.

[26] Harry Bunt, Jan Alexandersson, Jae-Woong Choe, Alex Chengyu Fang, Koiti Hasida, Volha Petukhova, Andrei Popescu-Belis, and David Traum. 2012. ISO 24617-2: A Semantically-based Standard for Dialogue Annotation. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis (Eds.). European Language Resources Association (ELRA), Istanbul, Turkey, 430–437. https://aclanthology.org/L12-1296/

[27] M. Ariel Cascio, Eunlye Lee, Nicole Vaudrin, and Darcy A. Freedman. 2019. A Team-based Approach to Open Coding: Considerations for Creating Intercoder Consensus. *Field Methods* 31, 2 (2019), 116–130.

[28] Justine Cassell and Timothy Bickmore. 2003. Negotiated Collusion: Modeling Social Language and its Relationship Effects in Intelligent Agents. *User Modeling and User-adapted Interaction* 13, 1 (2003), 89–132.

[29] Aaron Chatterji, Thomas Cunningham, David J Deming, Zoe Hitzig, Christopher Ong, Carl Yan Shan, and Kevin Wadman. 2025. *How People Use ChatGPT*. Working Paper 34255. National Bureau of Economic Research.

[30] Ye Chen, Ke Zhou, Yiqun Liu, Min Zhang, and Shaoping Ma. 2017. Meta-evaluation of Online and Offline Web Search Evaluation Metrics. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Shinjuku, Tokyo, Japan) *(SIGIR '17)*. Association for Computing Machinery, New York, NY, USA, 15–24. https://doi.org/10.1145/3077136.3080804

[31] Myra Cheng, Alicia DeVrio, Lisa Egede, Su Lin Blodgett, and Alexandra Olteanu. 2024. "I Am the One and Only, Your Cyber BFF": Understanding the Impact of GenAI Requires Understanding the Impact of Anthropomorphic AI. arXiv:2410.08526 [cs.CY] https://arxiv.org/abs/2410.08526

[32] Myra Cheng, Sunny Yu, Cinoo Lee, Pranav Khadpe, Lujain Ibrahim, and Dan Jurafsky. 2025. ELEPHANT: Measuring and Understanding Social Sycophancy in LLMs. arXiv:2505.13995 [cs.CL] https://arxiv.org/abs/2505.13995

[33] Wei-Lin Chiang, Zhuohan Li, Zi Lin, Ying Sheng, Zhanghao Wu, Hao Zhang, Lianmin Zheng, Siyuan Zhuang, Yonghao Zhuang, Joseph E. Gonzalez, Ion Stoica, and Eric P. Xing. 2023. Vicuna: An Open-Source Chatbot Impressing GPT-4 with 90%* ChatGPT Quality. https://lmsys.org/blog/2023-03-30-vicuna/

[34] Lara Christoforakos and Sarah Diefenbach. 2023. Technology as a Social Companion? An Exploration of Individual and Product-related Factors of Anthropomorphism. *Social Science Computer Review* 41, 3 (2023), 1039–1062.

[35] Herbert H. Clark. 1996. *Using Language.* Cambridge University Press, Cambridge, U.K.

[36] Herbert H. Clark and Susan E Brennan. 1991. *Grounding in Communication.* American Psychological Association, Washington, D.C., USA, 222–223.

[37] Victoria Clarke and Virginia Braun. 2013. Teaching Thematic Analysis: Overcoming Challenges and Developing Strategies for Effective Learning. *The Psychologist* 26, 2 (2013), 120–123.

[38] Philip R. Cohen and C. Raymond Perrault. 1979. Elements of a Plan-based Theory of Speech Acts. *Cognitive Science* 3, 3 (1979), 177–212.

[39] Mark G. Core and James Allen. 1997. Coding Dialogs with the DAMSL Annotation Scheme. In *AAAI Fall Symposium on Communicative Action in Humans and Machines*, Vol. 56. AAAI Fall Symposium, Boston, MA, 28–35.

[40] Emily Corvi, Hannah Washington, Stefanie Reed, Chad Atalla, Alexandra Chouldechova, P. Alex Dow, Jean Garcia-Gathright, Nicholas J Pangakis, Emily Sheng, Dan Vann, Matthew Vogel, and Hanna Wallach. 2025. Taxonomizing Representational Harms Using Speech Act Theory. In *Findings of the Association for Computational Linguistics: ACL 2025*, Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar (Eds.). Association for Computational Linguistics, Vienna, Austria, 3907–3932. https://doi.org/10.18653/v1/2025.findings-acl.202

[41] Philip Davies. 2000. The Relevance of Systematic Reviews to Educational Policy and Practice. *Oxford Review of Education* 26, 3-4 (2000), 365–378. https://doi.org/10.1080/713688543 arXiv:https://doi.org/10.1080/713688543

[42] Jacques Derrida. 1997. *Of Grammatology.* Johns Hopkins University Press, Balitmore, MD, USA.

[43] Brenda Dervin. 1976. *The Everyday Information Needs of the Average Citizen: A Taxonomy for Analysis.* American Library Association, Chicago, IL, USA, 19–38.

[44] Brenda Dervin. 1998. Sense-making Theory and Practice: An Overview of User Interests in Knowledge Seeking and Use. *Journal of Knowledge Management* 2, 2 (1998), 36–46.

[45] Gerardine DeSanctis and Marshall Scott Poole. 1994. Capturing the Complexity in Advanced Technology Use: Adaptive Structuration Theory. *Organization Science* 5, 2 (1994), 121–147.

[46] Fernando Diaz. 2009. Integration of News Content into Web Results. In *Proceedings of the Second ACM International Conference on Web Search and Data Mining* (Barcelona, Spain) *(WSDM '09)*. Association for Computing Machinery, New York, NY, USA, 182–191. https://doi.org/10.1145/1498759.1498825

[47] Michaeleen Doucleff and Pien Huang. 2025. AI Chatbots are a Lifeline for Some Teens. Are They Safe? NPR. https://www.npr.org/sections/shots-health-news/2025/09/19/nx-s1-5545749/ai-chatbots-safety-openai-meta-characterai-teens-suicide Accessed on September 27, 2025.

[48] Paul Dourish. 2001. *Where the Action Is: The Foundations of Embodied Interaction.* MIT Press, Cambridge, MA, USA.

[49] Peter F. Drucker. 1993. The Rise of the Knowledge Society. *The Wilson Quarterly* 17, 2 (1993), 52–72.

[50] David Ellis. 1989. A Behavioural Approach to Information Retrieval System Design. *Journal of Documentation* 45, 3 (03 1989), 171–212.

[51] Nicholas Epley, Adam Waytz, and John T. Cacioppo. 2007. On Seeing Human: A Three-factor Theory of Anthropomorphism. *Psychological Review* 114, 4 (2007), 864.

[52] Aaron Fanous, Jacob Goldberg, Ank A. Agarwal, Joanna Lin, Anson Zhou, Roxana Daneshjou, and Sanmi Koyejo. 2025. Syceval: Evaluating LLM Sycophancy.

[53] Uwe Flick (Ed.). 2014. *The SAGE Handbook of Qualitative Data Analysis.* SAGE Publications, Thousand Oaks, CA, USA.

[54] Steve Fox, Kuldeep Karnawat, Mark Mydland, Susan Dumais, and Thomas White . 2005. Evaluating Implicit Measures to Improve Web Search. In *ACM Transactions on Information Systems*, Vol. 23. 147–168.

[55] Iason Gabriel, Arianna Manzini, Geoff Keeling, Lisa Anne Hendricks, Verena Rieser, Hasan Iqbal, Nenad Tomašev, Ira Ktena, Zachary Kenton, Mikel Rodriguez, Seliem El-Sayed, Sasha Brown, Canfer Akbulut, Andrew Trask, Edward Hughes, A. Stevie Bergman, Renee Shelby, Nahema Marchal, Conor Griffin, Juan Mateos-Garcia, Laura Weidinger, Winnie Street, Benjamin Lange, Alex Ingerman, Alison Lentz, Reed Enger, Andrew Barakat, Victoria Krakovna, John Oliver Siy, Zeb Kurth-Nelson, Amanda McCroskery, Vijay Bolina, Harry Law, Murray Shanahan, Lize Alberts, Borja Balle, Sarah de Haas, Yetunde Ibitoye, Allan Dafoe, Beth Goldberg, Sébastien Krier, Alexander Reese, Sims Witherspoon, Will Hawkins, Maribeth Rauh, Don Wallace, Matija Franklin, Josh A. Goldstein, Joel Lehman, Michael Klenk, Shannon Vallor, Courtney Biles, Meredith Ringel Morris, Helen King, Blaise Agüera y Arcas, William Isaac, and James Manyika. 2024. The Ethics of Advanced AI Assistants. arXiv:2404.16244 [cs.CY] https://arxiv.org/abs/2404.16244

[56] Deep Ganguli, Liane Lovitt, Jackson Kernion, Amanda Askell, Yuntao Bai, Saurav Kadavath, Ben Mann, Ethan Perez, Nicholas Schiefer, Kamal Ndousse, Andy Jones, Sam Bowman, Anna Chen, Tom Conerly, Nova DasSarma, Dawn Drain, Nelson Elhage, Sheer El-Showk, Stanislav Fort, Zac Hatfield-Dodds, Tom Henighan, Danny Hernandez, Tristan Hume, Josh Jacobson, Scott Johnston, Shauna Kravec, Catherine Olsson, Sam Ringer, Eli Tran-Johnson, Dario Amodei, Tom Brown, Nicholas Joseph, Sam McCandlish, Chris Olah, Jared Kaplan, and Jack Clark. 2022. Red Teaming

Language Models to Reduce Harms: Methods, Scaling Behaviors, and Lessons Learned. arXiv:2209.07858 [cs.CL] https://arxiv.org/abs/2209.07858

[57] Jun Gao, Yuhan Liu, Haolin Deng, Wei Wang, Yu Cao, Jiachen Du, and Ruifeng Xu. 2021. Improving Empathetic Response Generation by Recognizing Emotion Cause in Conversations. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih (Eds.). Association for Computational Linguistics, Punta Cana, Dominican Republic, 807–819. https://doi.org/10.18653/v1/2021.findings-emnlp.70

[58] Les Gasser. 1986. The Integration of Computing and Routine Work. *ACM Transactions on Information Systems (TOIS)* 4, 3 (1986), 205–225.

[59] Katy Ilonka Gero, Tao Long, and Lydia B. Chilton. 2023. Social Dynamics of AI Support in Creative Writing. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 245, 15 pages. https://doi.org/10.1145/3544548.3580782

[60] Elihu M. Gerson and Susan Leigh Star. 1986. Analyzing Due Process in the Workplace. *ACM Transactions on Information Systems (TOIS)* 4, 3 (1986), 257–270.

[61] Souvick Ghosh, Manasa Rath, and Chirag Shah. 2018. Searching as Learning: Exploring Search Behavior and Learning Outcomes in Learning-related Tasks. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval* (New Brunswick, NJ, USA) *(CHIIR '18)*. Association for Computing Machinery, New York, NY, USA, 22–31. https://doi.org/10.1145/3176349.3176386

[62] Moshe Glickman and Tali Sharot. 2025. How Human-AI Feedback Loops Alter Human Perceptual, Emotional and Social Judgements. *Nature Human Behaviour* 9, 2 (2025), 345–359.

[63] Daena J. Goldsmith and Leslie A. Baxter. 2006. Constituting Relationships in Talk: A Taxonomy of Speech Events in Social and Personal Relationships. *Human Communication Research* 23, 1 (03 2006), 87–114. https://doi.org/10.1111/j.1468-2958.1996.tb00388.x arXiv:https://academic.oup.com/hcr/article-pdf/23/1/87/22342035/jhumcom0087.pdf

[64] Arthur C. Graesser and Natalie K. Person. 1994. Question Asking During Tutoring. *American Educational Research Journal* 31, 1 (1994), 104–137. https://doi.org/10.3102/00028312031001104

[65] Swati Gupta, Marilyn Walker, and Daniela Romano. 2007. Generating Politeness in Task Based Interaction: An Evaluation of the Effect of Linguistic Form and Culture. In *Proceedings of the Eleventh European Workshop on Natural Language Generation (ENLG 07)*, Stephan Busemann (Ed.). DFKI GmbH, Saarbrücken, Germany, 57–64. https://aclanthology.org/W07-2308/

[66] Juhye Ha, Hyeon Jeon, Daeun Han, Jinwook Seo, and Changhoon Oh. 2024. CloChat: Understanding How People Customize, Interact, and Experience Personas in Large Language Models. 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 305, 24 pages. https://doi.org/10.1145/3613904.3642472

[67] Barry Haddow, Tom Kocmi, Philipp Koehn, and Christof Monz (Eds.). 2024. *Proceedings of the Ninth Conference on Machine Translation*. Association for Computational Linguistics, Miami, Florida, USA. https://doi.org/10.18653/v1/2024.wmt-1.0

[68] Ahmed Hassan, Rosie Jones, and Kristina Lisa Klinkner. 2010. Beyond DCG: User Behavior as a Predictor of a Successful Search. In *WSDM '10: Proceedings of the Third ACM International Conference on Web Search and Data Mining*. ACM, New York, NY, USA, 221–230.

[69] Ahmed Hassan, Ryen W. White, Susan T. Dumais, and Yi-Min Wang. 2014. Struggling or Exploring?: Disambiguating Long Search Sessions. In *Proceedings of the 7th ACM International Conference on Web Search and Data Mining (WSDM '14)*. ACM, New York, NY, USA, 53–62.

[70] Ahmed Hassan Awadallah, Ryen W. White, Patrick Pantel, Susan T. Dumais, and Yi-Min Wang. 2014. Supporting Complex Search Tasks. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management* (Shanghai, China) *(CIKM '14)*. Association for Computing Machinery, New York, NY, USA, 829–838. https://doi.org/10.1145/2661829.2661912

[71] Dan Hendrycks, Collin Burns, Steven Basart, Andy Zou, Mantas Mazeika, Dawn Song, and Jacob Steinhardt. 2021. Measuring Massive Multitask Language Understanding. arXiv:2009.03300 [cs.CY] https://arxiv.org/abs/2009.03300

[72] Monique Hennink and Bonnie N. Kaiser. 2022. Sample Sizes for Saturation in Qualitative Research: A Systematic Review of Empirical Tests. *Social Science & Medicine* 292 (2022), 114523. https://doi.org/10.1016/j.socscimed.2021.114523

[73] Karl Moritz Hermann, Tomáš Kočiský, Edward Grefenstette, Lasse Espeholt, Will Kay, Mustafa Suleyman, and Phil Blunsom. 2015. Teaching machines to read and comprehend. In *Proceedings of the 29th International Conference on Neural Information Processing Systems - Volume 1* (Montreal, Canada) *(NIPS'15)*. MIT Press, Cambridge, MA, USA, 1693–1701.

[74] James Hollan, Edwin Hutchins, and David Kirsh. 2000. Distributed Cognition: Toward a New Foundation for Human-computer Interaction Research. *ACM Transactions on Computer-Human Interaction (TOCHI)* 7, 2 (2000), 174–196.

[75] Jun Hu, Rebecca J. Passonneau, and Owen Rambow. 2009. Contrasting the Interaction Structure of an Email and a Telephone Corpus: A Machine Learning Approach to Annotation of Dialogue Function Units. In *Proceedings of the SIGDIAL 2009 Conference: The 10th Annual Meeting of the Special Interest Group on Discourse and Dialogue* (London, United Kingdom) *(SIGDIAL '09)*. Association for Computational Linguistics, USA, 357–366.

[76] Thomas Huber and Christina Niklaus. 2025. LLMs meet Bloom's Taxonomy: A Cognitive View on Large Language Model Evaluations. In *Proceedings of the 31st International Conference on Computational Linguistics*, Owen Rambow, Leo Wanner, Marianna Apidianaki, Hend Al-Khalifa, Barbara Di Eugenio, and Steven Schockaert (Eds.). Association for Computational Linguistics, Abu Dhabi, UAE, 5211–5246. https://aclanthology.org/2025.coling-main.350/

[77] Edwin Hutchins. 1995. *Cognition in the Wild*. MIT Press, Cambridge, MA, USA.

[78] 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. *Proc. ACM Hum.-Comput. Interact.* 9, 2, Article CSCW122 (May 2025), 41 pages. https://doi.org/10.1145/3711020

[79] Maurice Jakesch, Advait Bhat, Daniel Buschek, Lior Zalmanson, and Mor Naaman. 2023. Co-Writing with Opinionated Language Models Affects Users' Views. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 111, 15 pages. https://doi.org/10.1145/3544548.3581196

[80] Bernard J. Jansen, Danielle Booth, and Brian Smith. 2009. Using the Taxonomy of Cognitive Learning to Model Online Searching. *Information Processing & Management* 45, 6 (2009), 643–663.

[81] Rosie Jones and Kristina Lisa Klinkner. 2008. Beyond the Session Timeout: Automatic Hierarchical Segmentation of Search Topics in Query Logs. In *Proceedings of the 17th ACM Conference on Information and Knowledge Management* (Napa Valley, California, USA) *(CIKM '08)*. Association for Computing Machinery, New York, NY, USA, 699–708. https://doi.org/10.1145/1458082.1458176

[82] Daniel Jurafsky and James H. Martin. 2025. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, with Language Models. https://web.stanford.edu/~jurafsky/slp3/ Online manuscript released August 24, 2025.

[83] D. Jurafsky, L. Shriberg, and D. Biasca. 1997. *Switchboard SWBD-DAMSL Shallow Discourse-Function Annotation: Coders Manual.* Technical Report. University of Colorado, Institute of Cognitive Science.

[84] Victor Kaptelinin and Bonnie A. Nardi. 2009. *Acting with Technology: Activity Theory and Interaction Design.* MIT Press, Cambridge, MA, USA.

[85] Shubhra Kanti Karmaker Santu and Dongji Feng. 2023. TELeR: A General Taxonomy of LLM Prompts for Benchmarking Complex Tasks. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, Houda Bouamor, Juan Pino, and Kalika Bali (Eds.). Association for Computational Linguistics, Singapore, 14197–14203. https://doi.org/10.18653/v1/2023.findings-emnlp.946

[86] Diane Kelly, Jaime Arguello, Ashlee Edwards, and Wan-ching Wu. 2015. Development and Evaluation of Search Tasks for IIR Experiments using a Cognitive Complexity Framework. In *Proceedings of the 2015 International Conference on The Theory of Information Retrieval* (Northampton, MA, USA) *(ICTIR '15)*. Association for Computing Machinery, New York, NY, USA, 101–110. https://doi.org/10.1145/2808194.2809465

[87] Markelle Kelly, Aakriti Kumar, Padhraic Smyth, and Mark Steyvers. 2023. Capturing Humans' Mental Models of AI: An Item Response Theory Approach. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (Chicago, IL, USA) *(FAccT '23)*. Association for Computing Machinery, New York, NY, USA, 1723–1734. https://doi.org/10.1145/3593013.3594111

[88] Hamed Khanpour, Nishitha Guntakandla, and Rodney Nielsen. 2016. Dialogue Act Classification in Domain-Independent Conversations Using a Deep Recurrent Neural Network. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, Yuji Matsumoto and Rashmi Prasad (Eds.). The COLING 2016 Organizing Committee, Osaka, Japan, 2012–2021. https://aclanthology.org/C16-1189/

[89] Julia Kiseleva, Kyle Williams, Jiepu Jiang, Ahmed Hassan Awadallah, Aidan C. Crook, Imed Zitouni, and Tasos Anastasakos. 2016. Understanding User Satisfaction with Intelligent Assistants. In *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval (CHIIR '16)*. ACM, New York, NY, USA, 121–130.

[90] Arnd Christian König, Michael Gamon, and Qiang Wu. 2009. Click-through Prediction for News Queries. In *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval* (Boston, MA, USA) *(SIGIR '09)*. Association for Computing Machinery, New York, NY, USA, 347–354. https://doi.org/10.1145/1571941.1572002

[91] Dimosthenis Kontogiorgos. 2022. *Mutual Understanding in Situated Interactions with Conversational User Interfaces: Theory, Studies, and Computation.* Ph. D. Dissertation. KTH Royal Institute of Technology.

[92] Carol C. Kuhlthau. 1994. *Seeking Meaning: A Process Approach to Library and Information Services.* Libraries Limited, London, England, U.K.

[93] Todd Kulesza, Simone Stumpf, Margaret Burnett, Sherry Yang, Irwin Kwan, and Weng-Keen Wong. 2013. Too Much, Too Little, or Just Right? Ways Explanations Impact End Users' Mental Models. In *2013 IEEE Symposium on Visual Languages and Human Centric Computing*. IEEE, San Jose, CA, USA, 3–10. https://doi.org/10.1109/VLHCC.2013.6645235

[94] Tom Kwiatkowski, Jennimaria Palomaki, Olivia Redfield, Michael Collins, Ankur Parikh, Chris Alberti, Danielle Epstein, Illia Polosukhin, Jacob Devlin, Kenton Lee, Kristina Toutanova, Llion Jones, Matthew Kelcey, Ming-Wei Chang, Andrew M. Dai, Jakob Uszkoreit, Quoc Le, and Slav Petrov. 2019. Natural Questions: A Benchmark for Question Answering Research. *Transactions of the Association for Computational Linguistics* 7 (08 2019), 453–466.

[95] Philippe Laban, Hiroaki Hayashi, Yingbo Zhou, and Jennifer Neville. 2025. LLMs Get Lost In Multi-Turn Conversation. arXiv:2505.06120 [cs.CL] https://arxiv.org/abs/2505.06120

[96] Linnea Laestadius, Andrea Bishop, Michael Gonzalez, Diana Illenčík, and Celeste Campos-Castillo. 2024. Too Human and Not Human Enough: A Grounded Theory Analysis of Mental Health Harms From Emotional Dependence on the Social Chatbot Replika. *New Media & Society* 26, 10 (2024), 5923–5941.

[97] Martha Lampland and Susan Leigh Star (Eds.). 2009. *Standards and Their Stories: How Quantifying, Classifying, and Formalizing Practices Shape Everyday Life.* Cornell University Press, Ithaca, NY, USA.

[98] Bruno Latour. 2005. *Reassembling the Social: An Introduction to Actor-Network-Theory.* Oxford University Press, Oxford, England. https://doi.org/10.1093/oso/9780199256044.001.0001

[99] Jean Lave and Etienne Wenger. 1991. *Situated Learning: Legitimate Peripheral Participation.* Cambridge University Press, Cambridge, U.K.

[100] Paul F. Lazarsfeld and Neil W. Henry. 1968. *Latent Structure Analysis.* Houghton Mifflin, Boston, MA.

[101] John D. Lee and Katrina A. See. 2004. Trust in Automation: Designing for Appropriate Reliance. *Human Factors* 46, 1 (2004), 50–80.

[102] Mina Lee, Percy Liang, and Qian Yang. 2022. CoAuthor: Designing a Human-AI Collaborative Writing Dataset for Exploring Language Model Capabilities. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 388, 19 pages. https://doi.org/10.1145/3491102.3502030

[103] Kenneth Li, Tianle Liu, Naomi Bashkansky, David Bau, Fernanda Viégas, Hanspeter Pfister, and Martin Wattenberg. 2024. Measuring and Controlling Instruction (In)Stability in Language Model Dialogs. arXiv:2402.10962 [cs.CL] https://arxiv.org/abs/2402.10962

[104] Zhuoyan Li, Chen Liang, Jing Peng, and Ming Yin. 2024. The Value, Benefits, and Concerns of Generative AI-Powered Assistance in Writing. 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 1048, 25 pages. https://doi.org/10.1145/3613904.3642625

[105] Q. Vera Liao, Hariharan Subramonyam, Jennifer Wang, and Jennifer Wortman Vaughan. 2023. Designerly Understanding: Information Needs for Model Transparency to Support Design Ideation for AI-Powered User Experience. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 9, 21 pages. https://doi.org/10.1145/3544548.3580652

[106] Matthew Lombard and Kun Xu. 2021. Social Responses to Media Technologies in the 21st Century: The Media are Social Actors Paradigm. *Human-Machine Communication* 2 (2021), 29–55. https://search.informit.org/doi/10.3316/INFORMIT.100016122150335

[107] Gale M. Lucas, Jill Boberg, David Traum, Ron Artstein, Jonathan Gratch, Alesia Gainer, Emmanuel Johnson, Anton Leuski, and Mikio Nakano. 2018. Culture, Errors, and Rapport-building Dialogue in Social Agents. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents* (Sydney, NSW, Australia) *(IVA '18)*. Association for Computing Machinery, New York, NY, USA, 51–58. https://doi.org/10.1145/3267851.3267887

[108] Qianou Ma, Weirui Peng, Chenyang Yang, Hua Shen, Ken Koedinger, and Tongshuang Wu. 2025. What Should We Engineer in Prompts? Training Humans in Requirement-driven LLM Use. *ACM Transactions on Computer-Human Interaction* 32, 4 (2025), 1–27.

[109] Takuya Maeda and Anabel Quan-Haase. 2024. When Human-AI Interactions Become Parasocial: Agency and Anthropomorphism in Affective Design. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency* (Rio de Janeiro, Brazil) *(FAccT '24)*. Association for Computing Machinery, New York, NY, USA, 1068–1077. https://doi.org/10.1145/3630106.3658956

[110] Stephann Makri and Ann Blandford. 2012. Coming Across Information Serendipitously – Part 1: A Process Model. *Journal of Documentation* 68, 5 (08 2012), 684–705.

[111] Stephann Makri and Ann Blandford. 2012. Coming Across Information Serendipitously – Part 2: A Classification Framework. *Journal of Documentation* 68, 5 (08 2012), 706–724.

[112] Arianna Manzini, Geoff Keeling, Nahema Marchal, Kevin R. McKee, Verena Rieser, and Iason Gabriel. 2024. Should Users Trust Advanced AI Assistants? Justified Trust As a Function of Competence and Alignment. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency* (Rio de Janeiro, Brazil) *(FAccT '24)*. Association for Computing Machinery, New York, NY, USA, 1174–1186. https://doi.org/10.1145/3630106.3658964

[113] Gary Marchionini. 2006. Exploratory Search: From Finding to Understanding. *Commun. ACM* 49, 4 (April 2006), 41–46. https://doi.org/10.1145/1121949.1121979

[114] Charles T. Meadow, Bert R. Boyce, Donald H. Kraft, and Carol Barry. 2007. *Text Information Retrieval Systems.* Academic Press, Orlando, FL, USA.

[115] Rishabh Mehrotra, Ahmed Hassan Awadallah, Milad Shokouhi, Emine Yilmaz, Imed Zitouni, Ahmed El Kholy, and Madian Khabsa. 2017. Deep Sequential Models for Task Satisfaction Prediction. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management (CIKM '17)*. ACM, New York, NY, USA, 737–746.

[116] Rishabh Mehrotra and Emine Yilmaz. 2017. Extracting Hierarchies of Search Tasks & Subtasks via a Bayesian Nonparametric Approach. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Shinjuku, Tokyo, Japan) *(SIGIR '17)*. Association for Computing Machinery, New York, NY, USA, 285–294. https://doi.org/10.1145/3077136.3080823

[117] George A. Miller. 1956. The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *Psychological Review* 63, 2 (1956), 81.

[118] Pushkar Mishra, Charvi Rastogi, Stephen R. Pfohl, Alicia Parrish, Tian Huey Teh, Roma Patel, Mark Diaz, Ding Wang, Michela Paganini, Vinodkumar Prabhakaran, Lora Aroyo, and Verena Rieser. 2025. Decoding Safety Feedback from Diverse Raters: A Data-driven Lens on Responsiveness to Severity. arXiv:2503.05609 [cs.CY] https://arxiv.org/abs/2503.05609

[119] Youngme Moon. 2000. Intimate Exchanges: Using Computers to Elicit Self-disclosure From Consumers. *Journal of Consumer Research* 26, 4 (2000), 323–339.

[120] Jared Moore, Declan Grabb, William Agnew, Kevin Klyman, Stevie Chancellor, Desmond C. Ong, and Nick Haber. 2025. Expressing Stigma and Inappropriate Responses Prevents LLMs From Safely Replacing Mental Health Providers.. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency (FAccT '25)*. Association for Computing Machinery, New York, NY, USA, 599–627. https://doi.org/10.1145/3715275.3732039

[121] Meredith Ringel Morris, Jaime Teevan, and Katrina Panovich. 2010. What do People Ask Their Social Networks, and Why? A Survey Study of Status Message Q&A Behavior. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. Association for Computing Machinery, New York, NY, USA, 1739–1748.

[122] Sara Moussawi and Raquel Benbunan-Fich. 2021. The Effect of Voice and Humour on Users' Perceptions of Personal Intelligent Agents. *Behaviour & Information Technology* 40, 15 (2021), 1603–1626.

[123] Bonnie A. Nardi. 1996. *Activity Theory and Human-computer Interaction.* MIT Press, Cambridge, MA, USA, 7–16.

[124] Bonnie A. Nardi. 1996. *Context and Consciousness: Activity Theory and Human-computer Interaction.* MIT Press, Cambridge, MA, USA.

[125] Clifford Nass, Brian Jeffrey Fogg, and Youngme Moon. 1996. Can Computers be Teammates? *International Journal of Human-Computer Studies* 45, 6 (1996), 669–678.

[126] Clifford Nass, Youngme Moon, and Nancy Green. 1997. Are Machines Gender Neutral? Gender-stereotypic Responses to Computers with Voices. *Journal of Applied Social Psychology* 27, 10 (1997), 864–876.

[127] Dorothy Ed Nelkin. 1992. *Controversy: Politics of Technical Decisions.* Sage Publications, Inc, Thousand Oaks, CA.

[128] Jinjie Ni, Tom Young, Vlad Pandelea, Fuzhao Xue, and Erik Cambria. 2023. Recent Advances in Deep Learning Based Dialogue Systems: A Systematic Survey. *Artificial Intelligence Review* 56, 4 (2023), 3055–3155.

[129] Robert C. Nickerson, Upkar Varshney, and Jan Muntermann. 2013. A Method for Taxonomy Development and Its Application in Information Systems. *European Journal of Information Systems* 22, 3 (2013), 336–359.

[130] Magalie Ochs, Catherine Pelachaud, and David Sadek. 2008. An Empathic Virtual Dialog Agent to Improve Human-machine Interaction. In *Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems - Volume 1* (Estoril, Portugal) *(AAMAS '08)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 89–96.

[131] Andrew M Olney, Arthur C Graesser, and Natalie K Person. 2012. Question Generation From Concept Maps. *Dialogue & Discourse* 3, 2 (2012), 75–99.

[132] Adinoyi Omuya, Vinodkumar Prabhakaran, and Owen Rambow. 2013. Improving the Quality of Minority Class Identification in Dialog Act Tagging. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Lucy Vanderwende, Hal Daumé III, and Katrin Kirchhoff (Eds.). Association for Computational Linguistics, Atlanta, Georgia, 802–807. https://aclanthology.org/N13-1099/

[133] Wanda J. Orlikowski. 2000. Using Technology and Constituting Structures: A Practice Lens for Studying Technology in Organizations. *Organization Science* 11, 4 (2000), 404–428.

[134] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. 2022. Training Language Models to Follow Instructions with Human Feedback. *Advances in neural information processing systems* 35 (2022), 27730–27744.

[135] Raja Parasuraman and Victor Riley. 1997. Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors* 39, 2 (1997), 230–253.

[136] Guy Paré, Marie-Claude Trudel, Mirou Jaana, and Spyros Kitsiou. 2015. Synthesizing Information Systems Knowledge: A Typology of Literature Reviews. *Information & Management* 52, 2 (2015), 183–199.

[137] Trevor J. Pinch and Wiebe E. Bijker. 1984. The Social Construction of Facts and Artefacts: or How the Sociology of Science and the Sociology of Technology might Benefit Each Other. *Social Studies of Science* 14, 3 (1984), 399–441. https://doi.org/10.1177/030631284014003004 arXiv:https://doi.org/10.1177/030631284014003004

[138] Peter Pirolli and Stuart Card. 1995. Information Foraging in Information Access Environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '95)*. ACM Press/Addison-Wesley Publishing Co., USA, 51–58. https://doi.org/10.1145/223904.223911

[139] Peter Pirolli and Stuart Card. 2005. The Sensemaking Process and Leverage Points for Analyst Technology as Identified Through Cognitive task analysis. *Proceedings of International Conference on Intelligence Analysis* 5, 1 (2005), 2–4.

[140] Catherine Pope, Sue Ziebland, and Nicholas Mays. 2000. Analysing Qualitative Data. *BMJ* 320, 7227 (2000), 114–116.

[141] Matthew Richard John Purver. 2004. *The Theory and Use of Clarification Requests in Dialogue.* Ph. D. Dissertation. University of London King's College.

[142] Lingyun Qiu and Izak Benbasat. 2009. Evaluating Anthropomorphic Product Recommendation Agents: A Social Relationship Perspective to Designing Information Systems. *Journal of Management Information Systems* 25, 4 (2009), 145–182.

[143] Aravind Sesagiri Raamkumar and Yinping Yang. 2022. Empathetic Conversational Systems: A Review of Current Advances, Gaps, and Opportunities. *IEEE Transactions on Affective Computing* 14, 4 (2022), 2722–2739.

[144] Hannah Rashkin, Eric Michael Smith, Margaret Li, and Y-Lan Boureau. 2019. Towards Empathetic Open-domain Conversation Models: A New Benchmark and Dataset. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Anna Korhonen, David Traum, and Lluís Màrquez (Eds.). Association for Computational Linguistics, Florence, Italy, 5370–5381. https://doi.org/10.18653/v1/P19-1534

[145] Byron Reeves and Clifford Nass. 1996. *The Media Equation: How People Treat Computers, Television, and New Media Like Real People.* Cambridge University Press, Cambridge, England, U.K.

[146] Patrizia Ribino. 2023. The Role of Politeness in Human–machine Interactions: A Systematic Literature Review and Future Perspectives. *Artificial Intelligence Review* 56, Suppl 1 (2023), 445–482.

[147] Soo Young Rieh, Kevyn Collins-Thompson, Preben Hansen, and Hye-Jung Lee. 2016. Towards Searching as a Learning Process: A Review of Current Perspectives and Future Directions. *Journal of Information Science* 42, 1 (2016), 19–34. https://doi.org/10.1177/0165551515615841

[148] Daniel E. Rose and Danny Levinson. 2004. Understanding User Goals in Web Search. In *Proceedings of the 13th International Conference on World Wide Web* (New York, NY, USA) *(WWW '04)*. Association for Computing Machinery, New York, NY, USA, 13–19. https://doi.org/10.1145/988672.988675

[149] Rudy Ruggles. 1998. The State of the Notion: Knowledge Management in Practice. *California Management Review* 40, 3 (1998), 80–89. https://doi.org/10.2307/41165944 arXiv:https://doi.org/10.2307/41165944

[150] Daniel M. Russell, Mark J. Stefik, Peter Pirolli, and Stuart K. Card. 1993. The Cost Structure of Sensemaking. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems* (Amsterdam, The Netherlands) *(CHI '93)*. Association for Computing Machinery, New York, NY, USA, 269–276. https://doi.org/10.1145/169059.169209

[151] Christine Rzepka and Benedikt Berger. 2018. User Interaction with AI-enabled Systems: A Systematic Review of IS Research.

[152] D.R. Sadler. 1989. Formative Assessment and the Design of Instructional Systems. *Instructional Science* 18, 2 (1989), 119–144. https://doi.org/10.1007/BF00117714

[153] Pritish Sahu, Michael Cogswell, Ajay Divakaran, and Sara Rutherford-Quach. 2021. Comprehension Based Question Answering using Bloom's Taxonomy. In *Proceedings of the 6th Workshop on Representation Learning for NLP (RepL4NLP-2021)*, Anna Rogers, Iacer Calixto, Ivan Vulić, Naomi Saphra, Nora Kassner, Oana-Maria Camburu, Trapit Bansal, and Vered Shwartz (Eds.). Association for Computational Linguistics, Online, 20–28. https://doi.org/10.18653/v1/2021.repl4nlp-1.3

[154] Rodrygo L.T. Santos, Craig Macdonald, and Iadh Ounis. 2011. Intent-aware Search Result Diversification. In *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Beijing, China) *(SIGIR '11)*. Association for Computing Machinery, New York, NY, USA, 595–604. https://doi.org/10.1145/2009916.2009997

[155] Reijo Savolainen. 1995. Everyday Life Information Seeking: Approaching Information Seeking in the Context of "Way of Life". *Library & Information Science Research* 17, 3 (1995), 259–294.

[156] Reijo Savolainen. 2010. Everyday Life Information Seeking. In *Encyclopedia of Library and Information Sciences*, Marcia J. Bates and Mary Niles Maack (Eds.). Taylor & Francis, Boca Raton, FL.

[157] Reijo Savolainen. 2022. Afterword: Manifestations of Joy of Information in Everyday Information Behavior Research. *Library Trends* 70, 4 (2022), 669–676.

[158] Emanuel A. Schegloff, Gail Jefferson, and Harvey Sacks. 1977. The Preference for Self-correction in the Organization of Repair in Conversation. *Language* 53, 2 (1977), 361–382.

[159] John R. Searle. 1969. *Speech Acts: An Essay in the Philosophy of Language.* Cambridge University Press, Cambridge, U.K.

[160] Chirag Shah, Ryen White, Reid Andersen, Georg Buscher, Scott Counts, Sarkar Das, Ali Montazer, Sathish Manivannan, Jennifer Neville, Nagu Rangan, Tara Safavi, Siddharth Suri, Mengting Wan, Leijie Wang, and Longqi Yang. 2025. Using Large Language Models to Generate, Validate, and Apply User Intent Taxonomies. *ACM Trans. Web* 19, 3, Article 34 (Aug. 2025), 29 pages. https://doi.org/10.1145/3732294

[161] Chirag Shah, Ryen White, Paul Thomas, Bhaskar Mitra, Shawon Sarkar, and Nicholas Belkin. 2023. Taking Search to Task. In *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval* (Austin, TX, USA) *(CHIIR '23)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3576840.3578288

[162] Omar Shaikh, Kristina Gligoric, Ashna Khetan, Matthias Gerstgrasser, Diyi Yang, and Dan Jurafsky. 2024. Grounding Gaps in Language Model Generations. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, Kevin Duh, Helena Gomez, and Steven Bethard (Eds.). Association for Computational Linguistics, Mexico City, Mexico, 6279–6296. https://doi.org/10.18653/v1/2024.naacl-long.348

[163] Mrinank Sharma, Meg Tong, Tomasz Korbak, David Duvenaud, Amanda Askell, Samuel R. Bowman, Newton Cheng, Esin Durmus, Zac Hatfield-Dodds, Scott R. Johnston, Shauna Kravec, Timothy Maxwell, Sam McCandlish, Kamal Ndousse, Oliver Rausch, Nicholas Schiefer, Da Yan, Miranda Zhang, and Ethan Perez. 2025. Towards Understanding Sycophancy in Language Models. arXiv:2310.13548 [cs.CL] https://arxiv.org/abs/2310.13548

[164] Renee Shelby, Shalaleh Rismani, Kathryn Henne, AJung Moon, Negar Rostamzadeh, Paul Nicholas, N'Mah Yilla-Akbari, Jess Gallegos, Andrew Smart, Emilio Garcia, and Gurleen Virk. 2023. Sociotechnical Harms of Algorithmic Systems: Scoping a Taxonomy for Harm Reduction. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society* (Montréal, QC, Canada) *(AIES '23)*. Association for Computing Machinery, New York, NY, USA, 723–741. https://doi.org/10.1145/3600211.3604673

[165] Ben Shneiderman. 1996. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In *Proceedings of the 1996 IEEE Symposium on Visual Languages (VL '96)*. IEEE Computer Society, USA, 336.

[166] Ben Shneiderman. 2007. Creativity Support Tools: Accelerating Discovery and Innovation. *Commun. ACM* 50, 12 (Dec. 2007), 20–32. https://doi.org/10.1145/1323688.1323689

[167] Prakash Shukla, Phuong Bui, Sean S Levy, Max Kowalski, Ali Baigelenov, and Paul Parsons. 2025. De-skilling, Cognitive Offloading, and Misplaced Responsibilities: Potential Ironies of AI-Assisted Design. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25)*. Association for Computing Machinery, New York, NY, USA, Article 171, 7 pages. https://doi.org/10.1145/3706599.3719931

[168] Rachel Siden, Hannah Kerman, Robert J. Gallo, Joséphine A. Cool, Jason Hom, Ethan Goh, Neera Ahuja, Paul Heidenreich, Lisa Shieh, Daniel Yang, Jonathan H Chen, Adam Rodman, and Laura M Holdsworth. 2025. A Typology of Physician Input Approaches to Using AI Chatbots for Clinical Decision-making: A Mixed Methods Study. https://doi.org/10.1101/2025.07.23.25332002 arXiv:https://www.medrxiv.org/content/early/2025/07/23/2025.07.23.25332002.full.pdf

[169] Georg Simmel. 1949. The Sociology of Sociability. *Amer. J. Sociology* 55, 3 (1949), 254–261.

[170] Mark D. Smucker and James Allan. 2006. Find-similar: Similarity Browsing as a Search Tool. In *SIGIR '06: Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM Press, New York, NY, USA, 461–468.

[171] Social Media Victims Law Center. 2025. Social Media Victims Law Center Files Three New Lawsuits on Behalf of Children Who Died of Suicide or Suffered Sex Abuse by Character.AI. Press Release. https://socialmediavictims.org/press-releases/social-media-victims-law-center-files-three-new-lawsuits-on-behalf-of-children-who-died-of-suicide-or-suffered-sex-abuse-by-character-ai/ Accessed on September 27, 2025.

[172] Susan Leigh Star and James R. Griesemer. 1989. Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science* 19, 3 (1989), 387–420. https://doi.org/10.1177/030631289019003001 arXiv:https://doi.org/10.1177/030631289019003001

[173] Andreas Stolcke, Klaus Ries, Noah Coccaro, Elizabeth Shriberg, Rebecca Bates, Daniel Jurafsky, Paul Taylor, Rachel Martin, Carol Van Ess-Dykema, and Marie Meteer. 2000. Dialogue Act Modeling for Automatic Tagging and Recognition of Conversational Speech. *Computational Linguistics* 26, 3 (2000), 339–373.

[174] Lucille Alice Suchman. 1987. *Plans and Situated Actions: The Problem of Human-machine Communication.* Cambridge University Press, Cambridge, U.K.

[175] Jiao Sun, Yufei Tian, Wangchunshu Zhou, Nan Xu, Qian Hu, Rahul Gupta, John Wieting, Nanyun Peng, and Xuezhe Ma. 2023. Evaluating Large Language Models on Controlled Generation Tasks. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, Houda Bouamor, Juan Pino, and Kalika Bali (Eds.). Association for Computational Linguistics, Singapore, 3155–3168. https://doi.org/10.18653/v1/2023.emnlp-main.190

[176] S Shyam Sundar. 2020. Rise of Machine Agency: A Framework for Studying the Psychology of Human–AI Interaction (HAII). *Journal of Computer-Mediated Communication* 25, 1 (01 2020), 74–88. https://doi.org/10.1093/jcmc/zmz026 arXiv:https://academic.oup.com/jcmc/article-pdf/25/1/74/32961171/zmz026.pdf

[177] Vivian Ta, Caroline Griffith, Carolynn Boatfield, Xinyu Wang, Maria Civitello, Haley Bader, Esther DeCero, and Alexia Loggarakis. 2020. User Experiences of Social Support From Companion Chatbots in Everyday Contexts: Thematic Analysis. *J Med Internet Res* 22, 3 (6 Mar 2020), e16235. https://doi.org/10.2196/16235

[178] Joanna Tai, Rola Ajjawi, David Boud, Phillip Dawson, and Ernesto Panadero. 2018. Developing Evaluative Judgement: Enabling Students to Make Decisions About the Quality of Work. *Higher Education* 76, 3 (2018), 467–481.

[179] Alex Tamkin, Miles McCain, Kunal Handa, Esin Durmus, Liane Lovitt, Ankur Rathi, Saffron Huang, Alfred Mountfield, Jerry Hong, Stuart Ritchie, Michael Stern, Brian Clarke, Landon Goldberg, Theodore R. Sumers, Jared Mueller, William McEachen, Wes Mitchell, Shan Carter, Jack Clark, Jared Kaplan, and Deep Ganguli. 2024. Clio: Privacy-Preserving Insights into Real-World AI Use. arXiv:2412.13678 [cs.CY] https://arxiv.org/abs/2412.13678

[180] Chuang Tang, Shaobo (Kevin) Li, Suming Hu, Fue Zeng, and Qianzhou Du. 2025. Gender Disparities in the Impact of Generative Artificial Intelligence: Evidence from Academia. *PNAS Nexus* 4, 2 (02 2025), pgae591. https://doi.org/10.1093/pnasnexus/pgae591 arXiv:https://academic.oup.com/pnasnexus/article-pdf/4/2/pgae591/61731646/pgae591.pdf

[181] Jaime Teevan, Susan T. Dumais, and Daniel J. Liebling. 2008. To Personalize or Not to Personalize: Modeling Queries with Variation in User Intent. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (Singapore, Singapore) *(SIGIR '08)*. Association for Computing Machinery, New York, NY, USA, 163–170. https://doi.org/10.1145/1390334.1390364

[182] Andrew Thatcher. 2008. Web Search Strategies: The Influence of Web Experience and Task Type. *Information Processing & Management* 44, 3 (2008), 1308–1329.

[183] Elaine G. Toms. 2011. *Task-based Information Searching and Retrieval.* Facet Publishing, London, England, 43–60.

[184] Stephen E. Toulmin. 2003. *The Uses of Argument.* Cambridge University Press, Cambridge, England, U.K.

[185] David Traum. 1999. Computational Models of Grounding in Collaborative Systems. In *AAAI Fall Symposium.* AAAI Press, North Falmouth, MA, USA, 124–131.

[186] David R Traum and Elizabeth A Hinkelman. 1992. Conversation Acts in Task-oriented Spoken Dialogue. *Computational Intelligence* 8, 3 (1992), 575–599.

[187] Johanne R. Trippas, Damiano Spina, Paul Thomas, Mark Sanderson, Hideo Joho, and Lawrence Cavedon. 2020. Towards a Model for Spoken Conversational Search. *Information Processing & Management* 57, 2 (2020), 102162.

[188] Kelsey Urgo, Jaime Arguello, and Robert Capra. 2019. Anderson and Krathwohl's Two-Dimensional Taxonomy Applied to Task Creation and Learning Assessment. In *Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval* (Santa Clara, CA, USA) *(ICTIR '19)*. Association for Computing Machinery, New York, NY, USA, 117–124. https://doi.org/10.1145/3341981.3344226

[189] Pertti Vakkari. 2010. Exploratory Searching as Conceptual Exploration. In *Proceedings of 4th Workshop on Human-computer Interaction and Information Retieval*, Bill Kules, Daniel Tunkelang, and Ryen White (Eds.). HCIR, New Brunswick, NJ, 24–27.

[190] Mengting Wan, Tara Safavi, Sujay Kumar Jauhar, Yujin Kim, Scott Counts, Jennifer Neville, Siddharth Suri, Chirag Shah, Ryen W. White, Longqi Yang, Reid Andersen, Georg Buscher, Dhruv Joshi, and Nagu Rangan. 2024. TnT-LLM: Text Mining at Scale with Large Language Models. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (Barcelona, Spain) *(KDD '24)*. Association for Computing Machinery, New York, NY, USA, 5836–5847. https://doi.org/10.1145/3637528.3671647

[191] Jason Wei, Maarten Bosma, Vincent Y. Zhao, Kelvin Guu, Adams Wei Yu, Brian Lester, Nan Du, Andrew M. Dai, and Quoc V. Le. 2022. Finetuned Language Models Are Zero-Shot Learners. arXiv:2109.01652 [cs.CL] https://arxiv.org/abs/2109.01652

[192] Karl E. Weick. 1995. *Sensemaking in Organizations.* Sage Publications, Thousand Oaks, CA, USA.

[193] Karl E. Weick, Kathleen M. Sutcliffe, and David Obstfeld. 2005. Organizing and the Process of Sensemaking. *Organization Science* 16, 4 (2005), 409–421.

[194] Jules White, Quchen Fu, Sam Hays, Michael Sandborn, Carlos Olea, Henry Gilbert, Ashraf Elnashar, Jesse Spencer-Smith, and Douglas C. Schmidt. 2023. A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT. arXiv:2302.11382 [cs.SE] https://arxiv.org/abs/2302.11382

[195] Jules White, Sam Hays, Quchen Fu, Jesse Spencer-Smith, and Douglas C. Schmidt. 2024. *ChatGPT Prompt Patterns for Improving Code Quality, Refactoring, Requirements Elicitation, and Software Design.* Springer Nature, Cham, Switzerland, 71–108.

[196] Jason D. Williams and Steve Young. 2007. Partially Observable Markov Decision Processes for Spoken Dialog Systems. *Computer Speech & Language* 21, 2 (2007), 393–422.

[197] Tom D Wilson. 1999. Models in Information Behaviour Research. *Journal of Documentation* 55, 3 (1999), 249–270.

[198] Hong (Iris) Xie. 2002. Patterns Between Interactive Intentions and Information-seeking Strategies. *Information Processing & Management* 38, 1 (2002), 55–77. https://doi.org/10.1016/S0306-4573(01)00018-8

[199] Xiaotong (Tone) Xu, Arina Konnova, Bianca Gao, Cindy Peng, Dave Vo, and Steven P. Dow. 2025. Productive vs. Reflective: How Different Ways of Integrating AI into Design Workflows Affect Cognition and Motivation. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 24, 15 pages. https://doi.org/10.1145/3706598.3713649

[200] JoAnne Yates. 1993. *Control Through Communication: The Rise of System in American Management*. Vol. 6. Johns Hopkins University Press, Baltimore, MD, USA.

[201] Elad Yom-Tov and Fernando Diaz. 2011. Location and Timeliness of Information Sources During News Events. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval (SIGIR '11)*. ACM, New York, NY, USA, 1105–1106.

[202] Dian Yu and Zhou Yu. 2021. MIDAS: A Dialog Act Annotation Scheme for Open Domain HumanMachine Spoken Conversations. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, Paola Merlo, Jorg Tiedemann, and Reut Tsarfaty (Eds.). Association for Computational Linguistics, Online, 1103–1120. https://doi.org/10.18653/v1/2021.eacl-main.94

[203] Zhiyuan Yu, Xiaogeng Liu, Shunning Liang, Zach Cameron, Chaowei Xiao, and Ning Zhang. 2024. Don't Listen to Me: Understanding and Exploring Jailbreak Prompts of Large Language Models. In *33rd USENIX Security Symposium (USENIX Security 24)*. USENIX Association, Berkeley, CA, 4675–4692.

[204] Yicong Yuan, Mingyang Su, and Xiu Li. 2024. What Makes People Say Thanks to AI. In *Artificial Intelligence in HCI*, Helmut Degen and Stavroula Ntoa (Eds.). Springer Nature Switzerland, Cham, 131–149.

[205] J.D. Zamfirescu-Pereira, Richmond Y. Wong, Bjoern Hartmann, and Qian Yang. 2023. Why Johnny Can't Prompt: How Non-AI Experts Try (and Fail) to Design LLM Prompts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 437, 21 pages. https://doi.org/10.1145/3544548.3581388

[206] Zheyuan Zhang, Jifan Yu, Juanzi Li, and Lei Hou. 2023. Exploring the Cognitive Knowledge Structure of Large Language Models: An Educational Diagnostic Assessment Approach. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, Houda Bouamor, Juan Pino, and Kalika Bali (Eds.). Association for Computational Linguistics, Singapore, 1643–1650. https://doi.org/10.18653/v1/2023.findings-emnlp.111

[207] Wenting Zhao, Xiang Ren, Jack Hessel, Claire Cardie, Yejin Choi, and Yuntian Deng. 2024. WildChat: 1M ChatGPT Interaction Logs in the Wild. arXiv:2405.01470 [cs.CL] https://arxiv.org/abs/2405.01470

Table 11. Distribution of Dialogues by Use Case.

Use Case	**# of Dialogues**
Advice, business	2
Advice, car insurance	1
Advice, dating	1
Advice, education	1
Advice, health	1
Advice, home theater design	1
Advice, immigration	1
Advice, parenting	1
Art, generic	1
Art, graphic design	1
Art, image generation	1
Art, prompt generation	3
Astrology	1
Business, analysis	2
Business, customer service	1
Business, strategy	4
Coding, adversarial	1

Continued on next page

Table 11 – continued from previous page

Use Case	# of Dialogues
Coding, c	2
Coding, c#	6
Coding, c++	1
Coding, dafny	1
Coding, generic	10
Coding, github	1
Coding, go	1
Coding, java	15
Coding, LaTex	1
Coding, programming test	1
Coding, python	10
Coding, SQL	2
Coding, website	2
Factoid	1
Factoid, animals	1
Factoid, book summary	2
Factoid, finance	2
Factoid, geography	1
Factoid, history	2
Factoid, language	3
Factoid, learning	3
Factoid, mixed topics	1
Factoid, music	1
Factoid, philosophy	1
Factoid, psychology	1
Factoid, science	2
Factoid, technology	16
Factoid, weather	1
Homework, CS	1
Homework, engineering	3
Homework, essay	1
Homework, generic	2
Homework, history	1
Homework, language	1
Homework, legal	1
Homework, literature	2
Homework, physics	1

Continued on next page

Table 11 – continued from previous page

Use Case	# of Dialogues
Homework, reading comprehension	6
Legal, document Q&A	1
Legal, hypothetical	1
Legal, liability	1
Legal, taxes	1
Marketing	3
Marketing, plan	1
Marketing, retail	1
Marketing, sexual	1
Marketing, strategy	1
Math	1
Personal, writing assistant	1
Plan, health/diet	1
Plan, travel	1
Play, role play	2
Play, role play, fan fiction	1
Play, text adventure game	2
Political science	1
Q&A	1
Research, literature review	2
SEO	16
Shopping	1
Shopping, ticket sales	1
Slide presentation	2
Teaching support	4
Translate	1
Unknown	17
Unknown, adversarial	4
Unknown, topic mix	6
Vanity search	1
Website development	4
Writing, advertisement	3
Writing, advertising	2
Writing, biblical	1
Writing, book	1
Writing, business	19
Writing, civil engineering	1

Continued on next page

Table 11 – continued from previous page

Use Case	# of Dialogues
Writing, co-writing	19
Writing, creative	8
Writing, fan fiction	6
Writing, generic	17
Writing, health	1
Writing, poetry	3
Writing, religious	1
Writing, sexually explicit	5
Grand Total	**294**

Table 10. Distribution of Dialogues by Language.

Language	# of Dialogues	Percentage
en	176	56.0%
zh	30	15.2%
ru	26	11.4%
de	18	0.8%
fr	14	2.4%
zh-Hant	13	1.0%
ar	9	5.2%
pt	6	0.2%
vi	5	5.5%
ko	5	0.0%
und	3	0.0%
nl	3	0.0%
it	3	0.3%
es	3	0.0%
tr	2	0.3%
no	2	0.0%
ja	2	0.2%
id	2	0.0%
fa	2	0.9%
zh	1	0.0%
uk	1	0.0%
pl	1	0.0%
ms	1	0.0%
hu	1	0.5%
da	1	0.0%
(unspecified)	1	0.0%
Grand Total	**294**	**100.0%**