ContextOS

Designing for High-Context AI Collaboration

Amer Mriziq
Aman Mehrotra
Ayushi Raj Bhatt
Manish Kumar

Advisor: Coye Cheshire

UC Berkeley, School of Information
Master of Information Management and Systems Capstone Project

May 9, 2025

UC Berkeley
School of Information

ContextOS

Executive Summary

Chat based LLM's, or AI's, such as ChatGPT, Claude, Gemini, Perplexity, etc., have proliferated everyday life to provide convenience and improved cognition to its user. Everyday more than 120 million users leverage ChatGPT's capabilities to inform all aspects of their lives from cooking, to data analysis, to personal advice, to data analysis and so much more. However, users today face significant friction when collaborating with LLM chatbots due to fragmented contexts and repetitive tasks, limiting productivity and causing frustration. To address these challenges, ContextOS streamlines LLM chatbot interactions by seamlessly integrating fragmented context into workflows, significantly reducing users effort. Through extensive research involving over 20 participants, we identified critical pain points: fragmented context across various tools, excessive cognitive load from repetitive instructions, lack of precision in contextual relevance, and limited interoperability across platforms. Informed by these insights, our team explored and prototyped a unified, intuitive context management solution, allowing users to easily inject and control context across tasks and LLM chatbot platforms. Our solution minimizes workflow disruption, empowers users with granular control, and enables seamless collaboration with multiple AI models. Ultimately, ContextOS dramatically enhances user efficiency and precision, transforming LLM chatbots from mere tools into true collaborative partners.

Introduction

The integration of LLM chatbots, or AI's, such as ChatGPT or Claude, into everyday workflows is rapidly changing how knowledge work is being done. AI tools are increasingly employed to assist with tasks ranging from drafting documents and automating communication to analyzing data and even generating creative content. The seemingly infinite possibilities of using AI creates the immense potential for boosting productivity and efficiency in all facets of an individual's life. A diverse array of AI tools are now readily accessible to individuals with varying levels of technical expertise, from those with no prior understanding of AI algorithms to seasoned professionals deeply familiar with machine learning principles. This widespread exposure to increasingly sophisticated AI capabilities creates an interesting and potentially problematic dynamic, akin to a digital divide, where the ability to effectively leverage these tools is heavily influenced by a user's understanding of their underlying mechanisms and the art of communicating with them.

Prompting, at its core, is the process of providing AI models with instructions or queries to elicit a desired response. While a simple prompt might be a single question, effective prompting often involves a nuanced approach that goes beyond basic commands. As outlined by the CTO of OpenAI in his "Anatomy of a Prompt" framework, a well-structured prompt often comprises several key elements that guide the AI towards generating the desired output. These elements can include the **instruction** (the specific task you want the AI to perform), the **context** (relevant background information to inform the AI's response), **input data** (the specific information you want the AI to process), and the **output format** (the desired structure or style of the AI's answer). Understanding and effectively utilizing these components is crucial for eliciting high-quality responses from AI models. Several key elements contribute to successful prompting, and among these, context plays a pivotal role. Context, in the realm of AI interaction, refers to the surrounding information that helps the AI model understand the user's intent and generate relevant, accurate, and useful outputs. This encompasses a wide array of information, including:

- **Background Information**: Providing the AI with necessary background details about the task, project, or subject matter.
- **Specific Instructions**: Clearly outlining the desired format, style, tone, and constraints for the AI's response.
- **Relevant Data**: Supplying the AI with relevant documents, data points, or examples to inform its output.
- **Conversational History**: In ongoing interactions, referencing previous turns in the conversation to maintain coherence and build upon prior exchanges.
- **User Role and Perspective**: Informing the AI about the user's role, expertise level, and intended audience for the output.

Ensuring that the AI model understands the user's intent and generates relevant, accurate, and useful outputs requires inputting sufficient context. However context is difficult for users to provide due to the tedious nature of determining and managing relevant context, especially across lengthy turn-by-turn interactions. This challenge of "context wrangling" – the difficulty in effectively providing and managing context for AI tools – creates a significant "invisible overhead" in AI-assisted workflows.

Objective

Our primary objective is to gain a deeper understanding of how users currently interact with AI tools. This involves investigating several key aspects of their workflows: first, we aim to understand users' existing familiarity and experience with various AI tools. Second, we seek to analyze how users construct and utilize prompts, particularly in relation to the elements of effective prompting (instruction, context, input data, output format). Finally, we intend to document the typical workflows users employ when integrating AI tools into their task processes, with a specific focus on how context is managed and utilized within these workflows. This research will provide a foundation for identifying opportunities to improve the design of AI interaction and context management.

Explorations

Literature Review

AI Collaboration and Productivity

Recent HCI research indicates that AI tools can both boost and complicate human productivity. For example, customer support agents with access to a generative AI saw a 14% average productivity increase, with lower-skilled workers improving by 34% and more experienced agents having marginal gains[11]. The effectiveness of the AI support depends on context like the users expertise and type of task[12]. One experiment with 76 software engineers found AI help using Google Bard had boosted their performance on open-ended coding problems for novices but not seasoned programmers[12]. At the same time, researchers observed risks such as “automation complacency”, where users become overly reliant on AI tools and begin trusting it too much, forgoing critical evaluation[12]. This indicates that productivity gains do occur, but only if the AI is used with oversight and appropriate context.

The impact of AI tools on tasks also hinges on how well the system understands and adapts to the user's situation. Specialized digital productivity assistants have been introduced to augment knowledge workers, bringing to light the individuals work patterns and providing suggestions for improvement[13]. AI tools illustrate their productivity value when they have deep context about users' activities. However, studies report that if an AI assistant's interpretation of context misaligned with the user's personal view of their work, the perceived effectiveness of the AI tool drops. This perceived reduction brings up notable barriers for adoption such as concerns about accuracy, transparency, and misalignment between the user and AI categorization of the task at hand[13]. AI can make people more productive but the literature is detailing that the benefit of AI tools is context dependent.

Context Management in Human-AI Interactions

Benefits of AI tools being context dependent sheds light on the criticality of managing context in human-AI collaborations. One challenge is the cognitive load of context-switching, where users often have to jump between multiple applications of AI tools. For example, knowledge workers may keep eight or more windows open and perform many application switches per hour which incurs re-entering or re-converting information and refocusing attention, increasing cognitive load[14]. Individuals depend on AI tools to carry over relevant context between taks, or even sessions, but need to repeat details. This can result in the AI giving irrelevant information because of missing context and harming the user's flow.

Conventional voice assistants often fail to maintain a broad understanding of context beyond simple commands because most interactions are one-offs[15]. Compared to legacy voice assistants, LLM systems exhibit more robust interaction due to enhanced context modeling, enabling more fluid dialogue and better handling of intent recognition errors[15]. Contextual information about AI and tasks improves human performance in collaborative decision-making. A delegation study revealed that participants with insight into an AI agent's capabilities and the task domain demonstrated significantly better team performance and more strategic AI reliance[16]. These findings show that managing context, by ensuring relevant information is accessible and persists across interactions, reduces cognitive burdens and keep the human-AI collaboration aligned.

Interoperability Across Platforms

Another theme in recent literature is the importance of interoperability among AI platforms to avoid trapping users in isolated ecosystems. When users invest time training an AI system (or adjusting to its style and preferences), switching to a different platform often means losing that accumulated context. This lack of portability creates a kind of "platform lock-in," where people stick with suboptimal tools rather than start over elsewhere. Industry analyses identify the absence of shared context standards as a major barrier; in fact, experts have called interoperability "one of the most significant barriers" to advancing AI utility across systems[17]. In response, the AI community has begun pushing for open protocols to allow context to travel with the user. A notable example is the Model Context Protocol (MCP) (introduced in late 2023 and updated in 2025), which aims to let different AI agents exchange contextual information in a common format. If widely adopted, such standards would enable AI assistants to remember a user's preferences and history across platforms[18]. In practical terms, this means a user could transition from one AI service to another without losing important context, reducing the friction and productivity loss associated with platform switching. Interoperability is key to seamless AI collaboration, ensuring that users aren't locked into a single vendor and that their contextual data can be leveraged wherever they go.

Designing User-Centered AI

Effective AI tools require a human-centered design approach. HCI scholars and practitioners have developed guidelines and methods for designing AI systems with user needs, context, and feedback in mind. For example, Amershi et al. (2019) compiled 18 validated guidelines for human-AI interaction to steer developers in creating intuitive and supportive AI features[19]. These guidelines emphasize principles like AI transparency, user control over AI outputs, and alignment with user goals. This human-centered process typically involves iterative prototyping and evaluation, where teams build and refine AI system prototypes based on user interaction. Modern generative AI has accelerated this iteration, enabling designers to use techniques like prompt engineering to quickly simulate AI behaviors and gather user feedback before full implementation[20]. This rapid cycle facilitates the exploration of alternatives and ensures AI responses integrate naturally into user workflows[20].

Participatory design is another AI development method that's gaining traction. It involves directly including end-users and other stakeholders in co-designing AI systems, such as workshops, beta testing, and feedback sessions. This is seen as super important for creating AI that's both context-aware and ethical. When designers involve users, they can really get a handle on the subtleties of how tasks are performed and what users genuinely need from the AI. Recent research highlights that this community-centered design is crucial for ensuring AI tools match human values and build trust, particularly in high-stakes areas like public services[21]. These iterative and participatory human-centered design methods help surface user worries early, and let AI solutions be customized for the specific environment and culture they'll be used in. These methods also lead to AI systems that are not only more usable and effective but also more likely to be embraced by users.

Observational Studies

The literature review highlighted the critical role of context in effective AI collaboration and underscored the importance of user-centered design methodologies. Key findings emphasized that AI tools should be designed to align with user needs, workflows, and the nuances of context within those workflows. Motivated by these insights and the identified challenges of 'context wrangling,' we adopted a user-centered approach to conduct observational studies. These studies aimed to provide a deeper understanding of how users currently manage context when interacting with AI tools, their specific pain points, and their strategies for overcoming context-related obstacles in real-world scenarios. This first-hand understanding of user behavior and context management practices directly informed the subsequent design and development of our ContextOS solution.

Methods

To gain a nuanced understanding of users' existing practices and challenges in managing context within AI-assisted workflows, we employed a qualitative observational study approach. Our goal was to observe users in their natural work environments as they interacted with AI tools, focusing on how they provided, accessed, and maintained context across various tasks and platforms. This study comprises insights from two sets of interviews: Study 1, which focused on trust in AI and involved 11 participants, and Study 2, which aimed to understand the integration of AI into knowledge workers' daily routines and involved 15 participants.

We recruited over 26 participants from diverse professional backgrounds who reported regular use of AI tools like ChatGPT, Claude, Gemini, and Perplexity. Participants included professionals in fields such as marketing, software development, research, and customer support. This variety allowed us to capture a range of use cases and context management strategies. For Study 2, participants were specifically recruited based on:

- Active use of AI tools like ChatGPT, Claude, Perplexity, or AI features in productivity apps (e.g., Notion AI, Grammarly).
- Employment in communication-heavy roles, like Product Management.

- Established recurring workflows involving AI, rather than sporadic or experimental usage.

Our observational sessions were conducted remotely via screen-sharing to allow us to witness participants' real-time interactions with AI tools and their digital workspaces. Each session lasted approximately 30-90 minutes. During these sessions, we asked participants to walk us through their typical workflows involving AI, focusing on specific tasks where context played a crucial role.

We instructed participants to think aloud as they worked, verbalizing their thought processes, the information they considered relevant as context, and the challenges they encountered in providing or managing that context. Our role as researchers was primarily observational, taking detailed notes on:

- The types of AI tools participants used.
- The tasks for which they employed AI.
- The sources of information they considered as context (e.g., documents, previous chats, emails, notes).
- The methods they used to provide context to the AI (e.g., copy-pasting, retyping, referencing previous interactions).
- Instances where context was explicitly mentioned or seemed to be a factor in the interaction's success or failure.
- Any workarounds or strategies participants employed to manage fragmented context or overcome limitations in context sharing across tools.
- The user's perceived reliability and usefulness of the AI's output, using a 7-point agreeability scale for each.
- Instances of user self-disclosure, positive acknowledgments, or expressions of displeasure, as potential indicators of trust or distrust.

Our observations were guided by the elements of a prompt outlined in the "Anatomy of a Prompt" framework (instruction, context, input data, output format) to understand how participants incorporated context into their prompts and workflows. We also paid close attention to instances of iterative refinement and back-and-forth dialogue with AI, noting how context was maintained or lost throughout these interactions.

To further understand user behavior, we also incorporated elements from the interview guides. Specifically, we explored:

- Users' familiarity with LLMs and how their approach to prompting has changed over time.
- Specific situations where the AI was particularly useful or unuseful, and what made those experiences stand out.
- How users determined if they could rely on the AI's output.
- How AI is integrated into the daily workflows of knowledge workers.
- The prompts and interaction strategies users employ.
- Usability challenges that arise during AI interactions.

The data collected from these observational sessions consisted of detailed written notes capturing user actions, verbalizations, and the context surrounding their AI interactions. This qualitative data was then analyzed using thematic analysis to identify recurring patterns, pain points, and user strategies related to context management in AI-assisted workflows.

Main Findings

Study 1 - Trust in AI

The purpose of this study's interviews was to gain a deep understanding of how individuals interact with advanced natural language AI agents like ChatGPT, Claude, and Perplexity. The core goal was to investigate how users use and perceive trust in human-AI interactions. The interviews were designed to provide rich qualitative data that could inform and validate quantitative measures, offering insights into the dynamic process of trust formation. Key research objectives guiding the interviews included assessing participants' existing familiarity with large language models (LLMs), capturing their diverse prompting strategies and interaction behaviors (both explicit prompts and other actions), and understanding their subjective perceptions of LLM reliability and usefulness in various tasks.

The methods employed involved conducting semi-structured interviews with approximately 11 participants from both student and professional backgrounds. These sessions were designed to be entirely voluntary and strictly confidential, with all identifying information slated for anonymization in any reports or publications. Participants explicitly granted permission for their conversations to be recorded and transcribed. The interview structure included initial warm-up questions to establish rapport and gauge baseline familiarity, followed by a crucial task: walking through an old chat log selected by the participant, ideally one involving multiple turns and deemed moderately to highly important to them. Participants were specifically asked to "think out loud" during this process, verbalizing their thoughts and feelings about their interactions. Throughout the task walkthroughs (or during new tasks), participants were asked to rate each AI output on a 1-7 scale for two distinct criteria: "Did the system perform as I expected it to?" (where 1 is strongly disagree and 7 is strongly agree) and "Was this response useful?" (where 1 is not useful and 7 is super useful). Post-task questions were also administered to capture broader reflections.

The findings yielded several significant insights into user interaction and trust with LLMs. Participants utilized the AI for a wide range of tasks, including deep research (e.g., summarizing academic papers, finding key themes), content revision and proofreading (including sensitive work emails), coding and debugging, brainstorming and ideation, creating documents like presentations or pitch drafts, tasks related to job applications (critiquing answers, extracting keywords), and even personal tasks like vacation planning or crafting gift ideas.

Users determined reliability through practical means such as cross-checking information provided by the AI, specifying detailed prompts with extensive context and rules, and iterating on responses to refine the output. Trust was built when the AI provided useful information or code that worked, especially on the first or second try. However, users experienced frustration when the AI provided incorrect or generic information, struggled with complex data inputs (like multi-page PDFs or handwritten notes), gave vague or surface-level answers, or seemed to "forget" previous context or instructions. Participants noted specific issues like the AI adding unnecessary formatting like long dashes or generating non-actionable links.

Aspects found to be useful included the AI's ability to save time, act as a starting point for writing or brainstorming, and offer different perspectives. Prompting strategies evolved over time, moving from vague, Google-like queries to highly specific instructions incorporating context, rules, and desired tone or style. Users learned that providing examples and detailed constraints significantly improved output quality. Some users explicitly interacted with the AI in a conversational, even emotional, manner, expressing frustration or gratitude, sometimes anthropomorphizing it as a "friend" or "human". This "feeling human" aspect could contribute to perceived usefulness in conversational exchanges. The turn-by-turn rating scales, while sometimes requiring clarification for the user, proved effective in capturing immediate feedback on whether the AI met expectations and the perceived utility of each response within the ongoing interaction. Users noted that even responses not meeting expectations could be somewhat useful if they provided hints for the next prompt

Study 2 - User Workflows

To understand how AI is integrated into the daily workflows of knowledge workers we conducted 15 qualitative interviews with the objective of answering three key questions:

- How are users currently integrating AI into their workflows?
- What prompts and interaction strategies do they use?
- What usability challenges arise during these interactions?

We sought to understand actual AI interactions in ideation, writing, and collaboration. Participants were knowledge workers, especially Product Managers at Bay Area B2B SaaS companies, using AI tools (e.g., ChatGPT, Claude) regularly. A Google Forms screener survey ensured participants met AI usage and role criteria.

The 30-minute, semi-structured interviews balanced consistency and flexibility. Participants were briefed on the research and consented to recording (or note-taking). Interviews covered: background (role, tasks, AI familiarity), recent AI usage in workflows (task, tool, context), recurring AI workflows (end-to-end interactions, challenges, AI avoidance), prompting and output refinement (prompt examples, evaluation), and open-ended reflections. Participants shared insights on their AI experiences, and confidentiality was maintained.

Investigating user workflows gave us insight into common use cases, prompting and interaction strategies, friction and usability challenges and AI workflows.

Common Use Cases

Participants reported using AI tools—primarily ChatGPT, Claude, Gemini, and Perplexity—to support a variety of daily tasks. The most common applications included:

- **Writing and editing**: users use AI to draft and refine PRDs, user flows, stakeholder emails, and internal documents. (eg., *"I write a brain dump first, then use ChatGPT for grammar, clarity, and tone tweaks."* – A.)
- **Information summarization**: users summarize long documents, user interviews, articles, or research reports. (eg., *"It helps me digest large amounts of information efficiently."* – J.)
- **Research and competitive analysis**: users use AI to support product discovery, market scans, and industry research. (eg., *"I use it to understand competitors or prepare for roadmap planning."* – G.)
- **Idea generation and brainstorming**: users ask AI to generate product ideas or evaluate early concepts. (eg., *"I prompted it with: 'I'm a PM at ABC—what are the flaws in this document?'"* – L.)
- **Presentation support**: users draft content for leadership presentations and external communications. (eg., *"I ask it to use a confident and assertive tone when prepping for meetings."* – V.)

Prompting & Interaction Strategies

Prompting strategies among participants varied widely depending on their experience level, the complexity of the task, and their personal preferences. We identified 4 approaches:

- **Goal-oriented:** Stating desired output first (e.g., J.: "I want to draw a flowchart for X. What do you need from me?").
- **Role/tone framing:** Contextualizing for audience (e.g., L.: "I'm a PM at ABC. Make this understandable for a non-technical stakeholder.").
- **Iterative exploration:** Refining through multiple rounds (e.g., I. on generating an image: "I didn't know all the details up front.").
- **Direct command-style:** Short, functional inputs (e.g., G.: "I don't want back-and-forth. If it doesn't work in two tries, I'll do it myself.").

Frictions & Usability Challenges

While participants consistently incorporated AI tools into their workflows, they also encountered recurring frictions that limited the tools' effectiveness. We identified the following 5 recurring friction and challenge points:

- **Lack of deep contextual understanding:** AI struggling with task/domain nuances (e.g., A.: "Claude doesn't get the deep data context—I usually know better," L.: "It doesn't understand the internal dynamics or feasibility constraints.").
- **Generic responses:** AI providing surface-level output, lacking creativity/depth for complex tasks (e.g., A.: "Starting a PRD from scratch gives generic results without enough nuance.").
- **Poor search integration/information retrieval:** Difficulty processing content from external sources (e.g., V.: "Parsing Google Slides fails, so I end up using screenshots.").
- **Iteration fatigue:** Frustration with repeated prompting (e.g., G.: "It's just not worth my time if I have to try three times.").
- **Trust/validation overhead:** Necessity of reviewing AI outputs (e.g., A.: "I always review the output to make sure it doesn't sound like ChatGPT," G.: "I cross-check everything—it adds extra effort.").

AI Workflows

Analyzing interviews revealed significant behavioral variations among PMs, even with similar goals. To synthesize findings, we shifted our analysis from content to interaction structure (input/output), identifying three distinct workflows:

- **Solved & Refine:** Users with completed outputs (e.g., emails) use AI for minor edits with low-context prompts (e.g., "Improve this"). PMs used this for high-stakes external communication (e.g., polishing emails to leadership).
- **Coworker:** Users co-develop outputs (e.g., PRDs) with AI, providing detailed guidance in multi-turn interactions.

- **T-Research:** Users explore broad, open-ended questions, using AI to gain clarity and refine research directions. Input is vague, and the goal is often discovery rather than a specific deliverable.

Workflow Type	Input	Output
Solved & Refine 🪄	- Solved work (e.g. first draft of an email) - Simple instructions *("improve this")*	Modified input (e.g. Polished text - email ready to be sent)
T- Research 🕵️	- Broad domain question *("explain me MCP..")*	User has a direction (e.g. Collected information)
Coworker 🤖	- Context *("I am a PM at a fintech company…")* - Instruction (*"Help me to prepare PRD.."*) - *context + Instructions are complex and dynamic* - Incomplete output *("Use this PRD structure..")*	Specific completed output (e.g. final PRD version)

Study Takeaways

Our two user studies provided valuable insights into how individuals interact with and integrate AI tools into their workflows. Study 1, focusing on trust in AI, revealed that users assess reliability through practical verification and build trust through consistent usefulness. Frustrations arose from inaccuracies, generic responses, and limitations in handling complex inputs or retaining context. Notably, trust was also influenced by the perceived "humanness" of the AI in conversational exchanges.

Study 2 broadened this perspective by examining the patterns of AI integration in knowledge workers' daily routines. We identified common use cases spanning writing, summarization, research, ideation, and presentation support. Prompting strategies varied based on user experience and task complexity, ranging from direct commands to iterative refinement. However, several key frictions emerged, including a lack of deep contextual understanding by the AI, the generation of generic outputs, challenges with information retrieval from diverse sources, iteration fatigue, and a significant overhead associated with trust and validation.

Interestingly, our analysis of user workflows in Study 2 highlighted that even with similar goals, users exhibited diverse interaction patterns. By focusing on the structure of these interactions (input and expected output), we identified three distinct workflows: Solved & Refine, Coworker, and T-Research.

Scoping a Solution: Addressing the Needs of Complex AI Collaboration

The insights gleaned from our user studies, particularly the dynamics of the **Coworker** workflow, highlight a critical area for innovation in human-AI interaction. This workflow, characterized by complex tasks, rich contextual requirements, and iterative refinement, represents a prime opportunity to significantly enhance user productivity and unlock the full potential of AI in professional settings.

As our findings illustrate, the ability for users to effectively manage and leverage relevant context is paramount for successful collaboration on these sophisticated tasks. The current limitations of siloed information and the burden of manual context curation present significant barriers to achieving seamless and productive AI partnerships. These limitations manifest in several key pain points, as summarized below:

Pain Point	User Need	Why It Matters
Siloed Context	Unified access to all relevant info	Fragmented tools prevent AI from acting as a true collaborator; user has to manually stitch context.
Instruction Fatigue	Get high-quality AI output without crafting perfect prompts	High prompt friction brakes user flow and prevents them from developing better prompting habits leading to weaker outcomes over time.
Relevance Changes by Task	Dynamically establish the most relevant context for the task	Persistent memory isn't useful without precision irrelevant info leads to bloated or generic output
"Context" Lock-In	Context portability across AI tools	Personalization creates stickiness, but users want the freedom to move across different LLMs.

Therefore, to truly support users in these high-value collaborative scenarios, a solution must address these core challenges. This presents a clear opportunity to:

- Unify context across platforms: Breaking down information silos to provide a holistic view.
- Give users control: Providing intuitive mechanisms for users to select and apply specific contextual elements to their AI interactions.
- Lower the prompt burden: Reducing the need for extensive manual prompting by leveraging system intelligence to anticipate and suggest relevant information.

By focusing on these key areas, we can envision a future where users can engage in complex AI collaborations with greater ease and efficiency, moving beyond the limitations of current tools. The following section will introduce one such conceptual approach designed to address these critical needs and facilitate high-context AI collaboration.

Conceptual Prototype: ContextOS

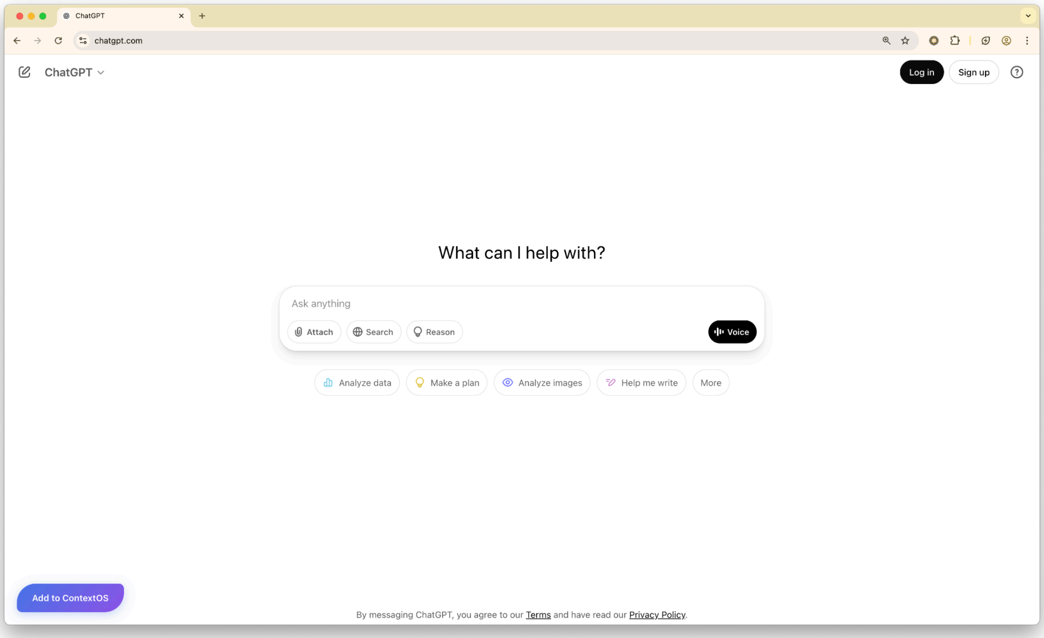
The core idea of our design is to lower the barriers to working within the Co-worker workflow. If we can help users manage huge amounts of context, then users can focus on getting the results they require. Our goal then becomes designing a user experience that scales with user skills, enabling a smooth shift from simple prompts to high-context, high-impact collaboration. If we seamlessly allow the injection of context, we can make advanced AI workflows effortless, fast and accessible to all users. We've created a set of design recommendations or a framework for designers looking to support context-heavy workflows in AI systems.

The following design recommendations are as follows:

1. Context must be centralized and organized as a repository to address multi-LLM usage or interoperability.
2. Users require granular control over the application of context to address the challenge of handling large amounts of data.
3. Finally, designers must minimize workflow disruptions when allowing for the injection or working with context as well as storing context to support the iterative refinement for more complex tasks.

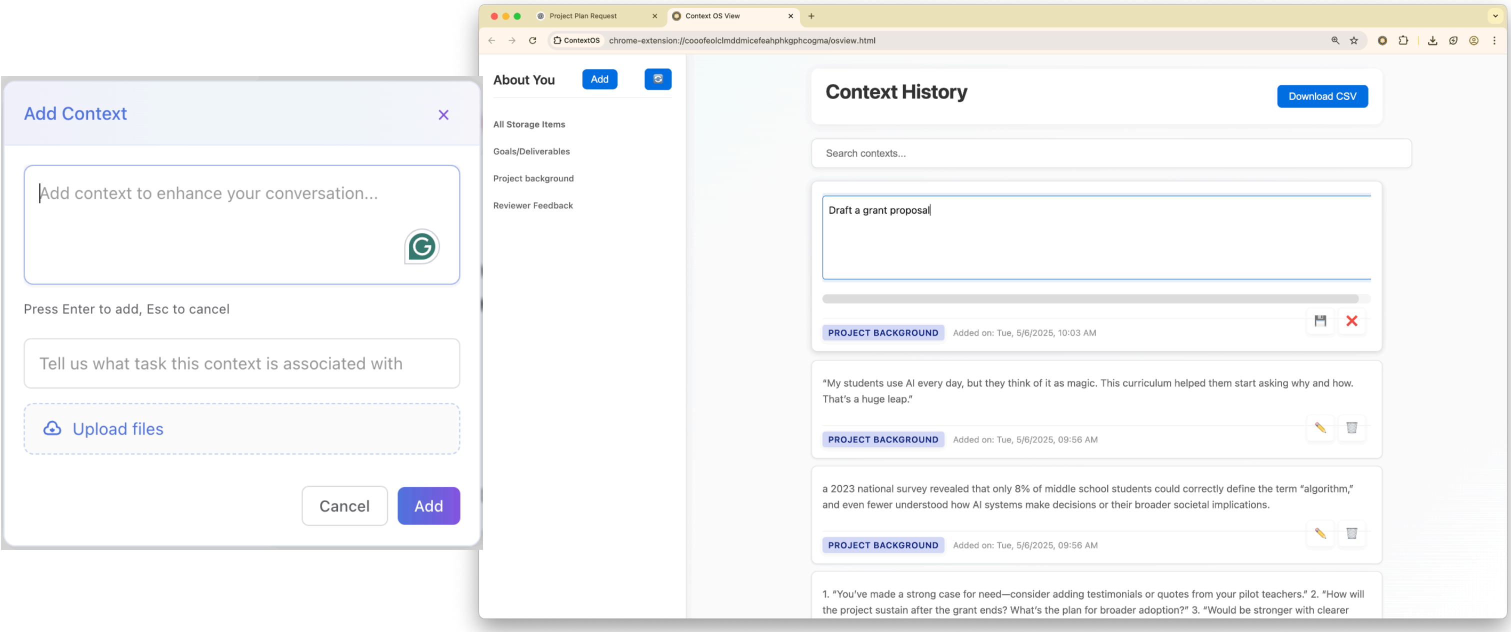
We’ve created a conceptual prototype to best exemplify the execution of these three design recommendations. We call this conceptual prototype ContextOS. ContextOS is made up of a set of key features that meet these three design requirements.

As it pertains to context centralization and interoperability:



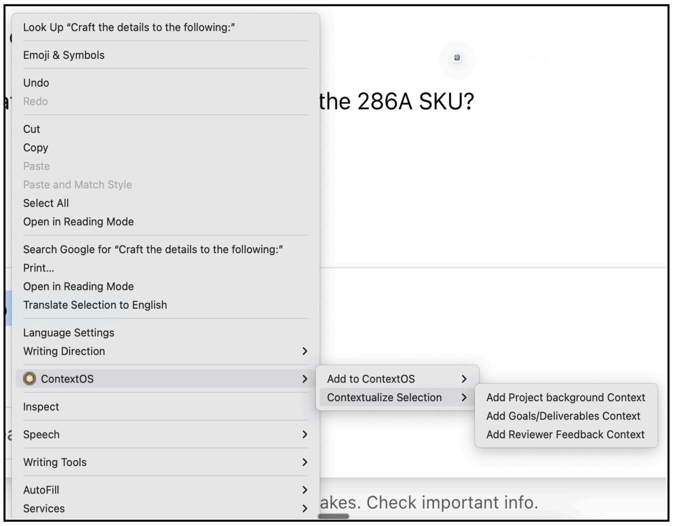
ContextOS consists of a set of features to meet the interoperability design requirement. Context OS relies on browser-based storage, meaning data remains on the device. Rather than remaining locked into specific LLM providers. Furthermore, ContextOS is enabled on all popular LLMs and affords users the ability to create user-generated categories to organize their context depending on the use case, task, or whatever they feel is necessary that best organizes their context.

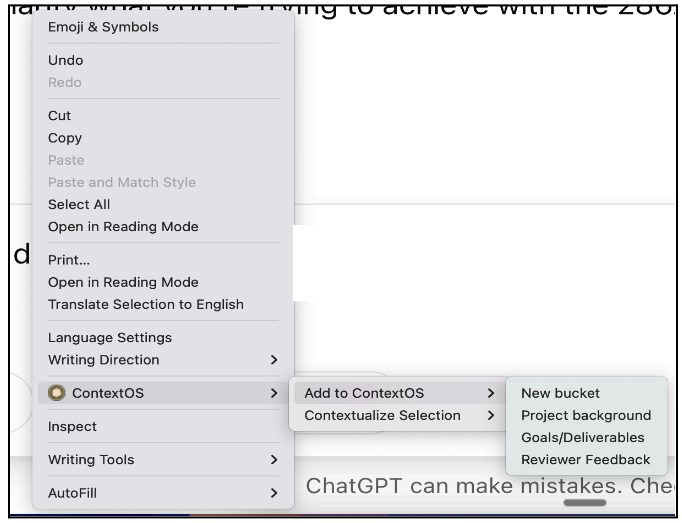
As it pertains to The second design requirement requires granular control over the application of context.



ContextOS is designed to be flexible, allowing users to create categories that fit their mental model for organizing context, depending on how they use LLMs. ContextOS is designed to allow the insertion of specific context that is relevant to the job at hand. When context is inserted, ContextOS uses an LLM in the background to read and structure the initial prompt based on the required task and relevant context.

Finally, pertaining to the third design goal of minimizing workflow disruptions.





Context OS allows for the injection of content via a right-click interaction design, using contextual menus baked into the browser. Users can highlight text, right-click, and insert context. Users can also highlight context that they've already sent to an LLM and add it to their browser storage in an ad-hoc fashion to accommodate for the fact that context is being added constantly. Finally, we allow for keyboard shortcuts to optimize for speed and minimize workflow disruptions.

References

1. Amershi, Saleema, et al. “Guidelines for Human-AI Interaction.” Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, ACM, 2019, pp. 1–13.
2. Bansal, Gagan, et al. “Does the Whole Exceed its Parts? The Effect of AI Explanations on Complementary Team Performance.” Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, ACM, 2021, pp. 1–16.
3. Cai, Carrie J., et al. “Onboarding Productivity Assistants: Emerging Tensions in AI-Augmented Work.” Proceedings of the ACM on Human-Computer Interaction, vol. 3, no. CSCW, 2019, pp. 1–30.
4. Dang, Tu N., et al. “AI Chains: Transparent and Controllable Human-AI Interaction by Chaining Large Language Model Prompts.” Proceedings of the ACM on Human-Computer Interaction, vol. 6, no. CSCW1, 2022, pp. 1–22.
5. Google Research. AI Assistance in Software Engineering: Productivity Gains and Challenges. Google Research Blog, 2024. https://research.googleblog.com/2024/03/ai-assistance-productivity.html.
6. Hancock, Braden, et al. “AI-Mediated Communication: How the Perception That Profile Text Was Written by AI Affects Trustworthiness and Hireability.” Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, ACM, 2022, pp. 1–15.
7. Lee, Jina, et al. “Understanding User Concerns in Using Productivity Applications Enhanced by Artificial Intelligence.” Proceedings of the ACM on Human-Computer Interaction, vol. 5, no. CSCW2, 2021, pp. 1–25.
8. Microsoft Research. Addressing Interoperability and Lock-in Challenges in AI Platforms: A Framework for Shared Context. Microsoft AI Whitepaper, 2025. https://www.microsoft.com/en-us/research/publication/addressing-interoperability-ai-platforms/.
9. OpenAI. The Model Context Protocol (MCP): Toward Cross-Platform Contextual Continuity in AI Assistants. OpenAI Technical Report, 2023, updated 2025. https://openai.com/research/model-context-protocol.
10. Purohit, Hemant, et al. “Socially Responsible AI Design: Guidelines, Practice, and Reflections from a Participatory Design Approach.” Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, ACM, 2023, pp. 1–14.
11. Brynjolfsson, Erik, Danielle Li, and Lindsey R. Raymond. “Generative AI at Work.” NBER Working Paper Series, no. 31161, National Bureau of Economic Research, Apr. 2023, revised Nov. 2023, https://www.nber.org/system/files/working_papers/w31161/w31161.pdf.
12. Qian, Crystal, and James Wexler. “Take It, Leave It, or Fix It: Measuring Productivity and Trust in Human-AI Collaboration.” arXiv, 1 Apr. 2024, https://arxiv.org/abs/2402.18498.

13. Cranefield, Jocelyn et al. “Partnering with AI: the case of digital productivity assistants.” *Journal of the Royal Society of New Zealand* vol. 53,1 95-118. 4 Sep. 2022, doi:10.1080/03036758.2022.2114507
14. Jahanlou, Amir, et al. “Task-Centric Application Switching: How and Why Knowledge Workers Switch Software Applications for a Single Task.” *Graphics Interface Conference 2023*, May 2023, https://www.research.autodesk.com/app/uploads/2024/02/Task-Centric-Application-Switching_GI23.pdf.
15. Mahmood, Amama, Junxiang Wang, Bingsheng Yao, Dakuo Wang, and Chien-Ming Huang. “User Interaction Patterns and Breakdowns in Conversing with LLM-Powered Voice Assistants.” *arXiv*, 28 Nov. 2024, https://arxiv.org/abs/2309.13879.
16. Spitzer, Philipp, et al. “Human Delegation Behavior in Human-AI Collaboration: The Effect of Contextual Information.” *arXiv*, version 3, 9 Jan. 2025, https://arxiv.org/abs/2401.04729.
17. Smith, Tom. “Model Context Protocol: The New Standard for AI Interoperability.” *Techstrong.ai*, 28 Mar. 2025, https://techstrong.ai/aiops/model-context-protocol-the-new-standard-for-ai-interoperability/.
18. Sarkar, Debanjan. “Model Context Protocol (MCP): The Future of AI Collaboration.” *LinkedIn*, 23 Mar. 2025, https://www.linkedin.com/pulse/model-context-protocol-mcp-future-ai-collaboration-debanjan-sarkar-3acmc/.
19. Amershi, Saleema, et al. “Guidelines for Human-AI Interaction.” *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ACM, 2019, pp. 1–13. https://www.microsoft.com/en-us/research/publication/guidelines-for-human-ai-interaction/.
20. Subramonyam, Hari, et al. “Prototyping with Prompts: Emerging Approaches and Challenges in Generative AI Design for Collaborative Software Teams.” *arXiv*, 30 Mar. 2025, https://arxiv.org/abs/2402.17721.
21. Saxena, Devansh, et al. “Emerging Practices in Participatory AI Design in Public Sector Innovation.” *arXiv*, 25 Feb. 2025, https://arxiv.org/abs/2502.18689.