Understanding How Task Dimensions Impact Automation Preferences with a Conversational Task Assistant

Jessica He[1,*], David Piorkowski[2], Michael Muller[3], Kristina Brimijoin[2], Stephanie Houde[3] and Justin D. Weisz[2]

[1]IBM Research, Seattle WA US
[2]IBM Research, Yorktown Heights NY US
[3]IBM Research, Cambridge MA US

Abstract
Organizations have recently begun to deploy conversational task assistants that collaborate with business users to partially automate their work tasks. These assistants are becoming more intelligent: users initiate automated task support through natural language, and the system can dynamically orchestrate new task sequences accordingly. As these tools become more intelligent and automated, they sometimes shift control away from users to increase process efficiency at the cost of consequences for users' preferences and productivity. Particularly in high stakes work environments, this shift raises questions of when automation is suitable or unsuitable and how to delegate agency such that users feel sufficiently in control of their tasks. We explored these questions through two studies comprised of interviews and co-design activities with business users and identified four task dimensions along which their automation and interaction preferences vary: process consequence, social consequence, task familiarity, and task complexity. These dimensions are useful for understanding when, why, and how to delegate control between users and conversational task assistants.

Keywords
Task automation, Conversational user interface (CUI), Multi-agent AI, Mixed-initiative human-AI, Human-centered AI (HCAI)

1. Introduction

Recent innovations in AI automation have enabled the adoption of AI assistants in partially automating business workflows for knowledge workers. Partial automation in high stakes work environments raises questions of when automation is suitable and unsuitable, and accordingly, when users should retain task control and when they can delegate it to the AI. In this paper, we explore these questions in the context of *conversational task assistants* that interact with business users through a natural language interface and execute business processes in back-end systems [1, 2, 3]. One example of such a system is Watson Orchestrate, which automates repetitive tasks for business users in a variety of domains [4].

Such systems are built on multi-agent *orchestration* technologies, where a front-end dialogue-manager transforms natural language *utterances* (what the user says or types) into an executable sequence of actions, powered by an underlying set of AI *skills* that connect to back-end services [1, 2, 3]. Multi-agent systems have many agents, each specialized for a particular skill, such as collecting data from a data store or filling out a form [5]. An orchestration layer sits above the agents, selecting the appropriate agent for given inputs and returning outputs from that agent back to the user. The diverse skill set of these systems allows conversational task assistants to support a variety of workflows.

Wiberg and Bergqvist [6] posit that automation poses a tension with designing for user control. They build on earlier work [7, 8] in *allocation of function* to humans and AIs through their Engaging Interaction through Automation scale, which outlines a spectrum from "no automation of interaction" to "full automation of interaction." Other prior workshop papers have discussed the notion that within a workflow, only certain steps or sub-tasks may be suitable for automation [9, 10].

Given the wide variety of tasks that conversational task assistants can support and the broad spectrum of users interacting with them, we sought to explore factors that inform a suitable division of automation. In line with principles of human-centered AI, we explored this from the perspective of human preferences. Through two studies comprised of user interviews and co-design activities, we identify four dimensions of tasks that can help designers of conversational task assistants determine when to automate interactions vs. when to provide user control. These dimensions are: consequences of task errors

AutomationXP23: Intervening, Teaming, Delegating - Creating Engaging Automation Experiences, April 23rd, Hamburg, Germany
*Corresponding author.
jessicahe@ibm.com (J. He); djp@ibm.com (D. Piorkowski); michael_muller@us.ibm.com (M. Muller); kbrimij@us.ibm.com (K. Brimijoin); stephanie.houde@ibm.com (S. Houde); jweisz@us.ibm.com (J. D. Weisz)
0000-0003-2368-0099 (J. He); 0000-0002-6740-4902 (D. Piorkowski); 0000-0001-7860-163X (M. Muller); 0000-0001-5616-9567 (K. Brimijoin); 0000-0002-0246-2183 (S. Houde); 0000-0003-2228-2398 (J. D. Weisz)

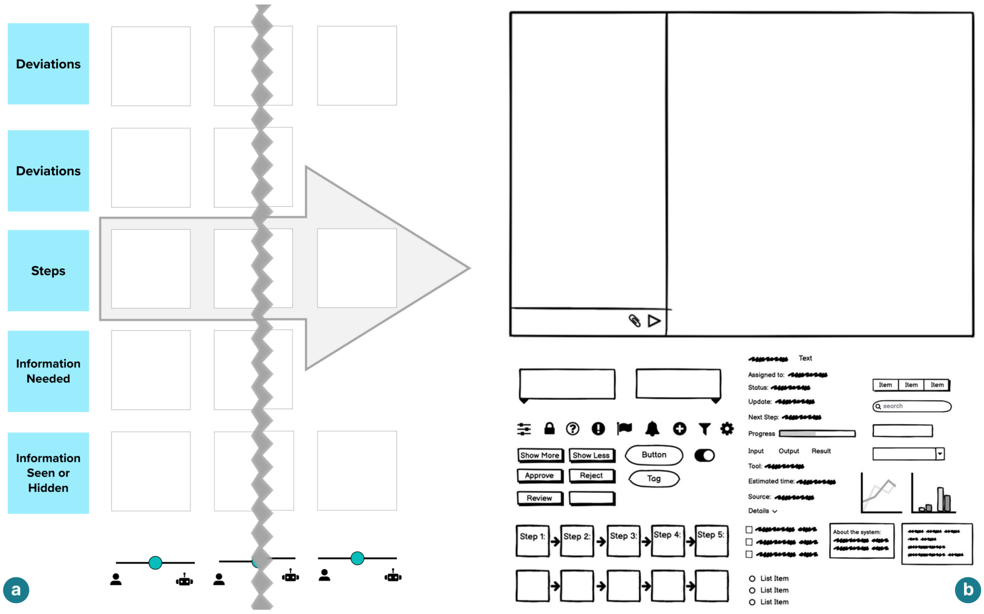

Figure 1: Participatory activity templates. 1a: Template used in Study 1 for participants to map out steps and automation preferences for a task. 1b: Blank interface and UI components provided for the co-design activities in Study 2.

(referred to as process consequence), social consequence, users' familiarity with a task, and task complexity.

Researchers have previously characterized task support in terms of multiple task dimensions based on classifications and trade-off dimensions. Tasks have been analyzed along dimensions of retrospective vs. prospective [11], informative vs. actionable [12], reminding vs. being-reminded [13], visible vs. invisible [14, 15, 16], content-oriented vs. relationship-oriented [17, 13, 18], and holistic [19] vs. itemized [20, 21, 22, 23].

In this paper, we build on this work through two studies that help define requirements for designing an AI task assistant. We show the importance of four new task dimensions based on human-AI interaction that are useful for understanding when, why, and how to delegate control between users and conversational task assistants.

2. Methods

We conducted two interview studies. Adapting a "bifocal" approach [24], the first study was a high-level task analysis, and the second study was a more detailed participatory design based on the task analysis. Participants for both studies were recruited from several company-wide Slack channels for business users of a large international technology company. Through a brief survey that asked them to describe their day-to-day job, we screened for people who had experience with workflows suitable for automation. 13 participants with varying roles (including developers, researchers, designers, managers, and salespeople) took part in Study 1. 15 new participants, also with varied job functions (including design, research, and business), took part in Study 2. Both studies were comprised of one-hour sessions with one researcher and one participant. Participants were compensated the equivalent of $25 USD. We refer to participants from Study 1 as P1xx, and participants from Study 2 as P2xx.

In the first study, sessions were comprised of an interview and a co-design analysis aimed at understanding current task practices and eliciting considerations relevant to the design of a conversational task assistant. For the co-design activity, we used a more structured version of the CARD method [25], adapted for remote participation via Mural (Figure 1a). Participants were asked to identify a work task suitable for partial automation (with input from the researcher). They mapped out their current process and pain points for this task in Mural, then discussed the level of automation they would be comfortable with for each step and *why*, along with any information and tooling needs they envisioned.

In the second study, we extended our inquiry into *how* automation preferences should be incorporated into the design of a conversational task assistant by conducting a participatory design study using an online version of paper-prototyping [26, 27]. We began each session by introducing the conversational task assistant's capabilities and limitations and showing a brief video demo. Participants were provided with a blank, low-fidelity version of the assistant's user interface, along with UI components that they could drag and drop and tools to design their own (Figure 1b). Using these components, they were asked to co-design a series of interactions to support the partial automation of two tasks: scheduling a meeting and a task of their choice. This activity served as the basis for discussions on participants' automation needs.

We collected 28 hours of video interview data. Using participants' responses and video transcriptions, we conducted a thematic analysis [28] to analyze the data for factors that affected participants' automation preferences. We describe these factors as characteristics of tasks, which formed the basis of what we termed *task dimensions.* We then analyzed how users preferred to work with the assistant in the context of these dimensions to understand design implications for task assistants. The following section presents results across both studies.

3. Task dimensions

We identified four dimensions of tasks along which automation preferences varied:

- **Process consequence:** The user's perceived cost of failure when the assistant makes a mistake.
- **Social consequence:** The user's perceived risks in allowing an assistant to represent them to others.
- **Familiarity:** The user's knowledge of the system and/or the task.
- **Complexity:** The overall difficulty or effort required for a user to complete a task.

Process consequence. When asked for tasks to automate, participants described a mix of tasks with minimal and significant consequences of error. Our studies revealed several instances where significant process consequences made users hesitant to automate a task or step. For example, P109 described an onboarding task that was high-consequence due to the sensitive information requested of them. They wanted to know how the system would use their information: *"I would want to know if that information is secure, if it's going anywhere after that, or if they just delete everything."* Thus, we find that the *process consequences* associated with a task is a dimension that affects users' information and agency needs.

These findings imply that only low-consequence tasks would be considered for automation, but further probing revealed that additional user oversight and control may help participants automate work *without* substantial risks. Among these controls were abilities to preview to-be-automated actions, verify outputs, and see consequential steps of an upcoming task. These granular insights would increase user comfort and control of the automation (similar to Park et al. [29]). Such interactions can draw attention to the risks within a task to help users understand the system's behavior and give them the control to mediate potential risks.

Social consequence. Some participants chose tasks involving others, such as teammates or clients, as candidates for automation. For example, P212 wanted the assistant to *"email everyone on [a]... project,"* and P104 wanted to schedule customer calls. Regarding automation preferences, P104 emphasized that initial contact with a new customer should be handled by a human *"to establish the... relationship with the customer."* Similarly, P111 felt that customer follow-ups should be human-driven, and P106 wanted *"to be able to control how widely communication [goes]."* These observations indicate that *social consequence* is another task dimension along which information needs and desired control vary.

Where interpersonal skills and emotional sensitivity are required, participants felt that the task assistant lacked the emotional intelligence to handle the task independently, an insight consistent with Goffman's emphasis that impression-management is important to people in their organizational lives [30]. These findings suggest that automation involving other people should largely remain under user control. Some participants did consider certain tasks less consequential despite involving other people (e.g. creating an HR ticket). Such tasks may be suitable for automation but should first request user approval given the diversity of concerns.

Familiarity. Participants also varied by familiarity, either with a task or with the system. Both types were determinants of how much automation, transparency, and guidance participants wanted from the system.

Non-experts of a task asked for transparency into system actions and guidance on the task process. Task familiarity is dynamic—people's inactivity in a particular domain may transform them from expert to non-expert, and unfamiliar tasks can also become more familiar over time. Participants expected the system to recognize the latter and reduce its support accordingly. After P212's first few expense approvals, they said they would no longer require step-by-step guidance on required inputs and instead would initiate the conversation with *"Submit an expense report for <event>. Here are the dates and locations. Here's a folder of the receipts."*

Similarly, transparency can orient non-experts of a system to its capabilities. Several participants expressed an initial distrust of the system. For instance, P107 spoke of submitting an expense report—a task that could have financial ramifications if done incorrectly: *"if I hadn't gotten that trust yet, then I'd probably ask the system, 'prepare the expense report for my review' rather than letting the system submit on its own."* P105 also commented that to automate work with the assistant, they wanted to *"watch it first"* to calibrate trust (similar to [31, 32]). A review step that offers users insight into the automation and allows them to verify outputs supports non-experts in understanding a system's capabilities.

Complexity. In the co-design activities of Study 2, participants adapted their designs to reflect the *complexity* of tasks and steps. Participants showed that they preferred chat interactions for simple workflows, but they preferred traditional graphical user interfaces (GUIs) for more complex, information-rich workflows. One reason for this difference was the perceived richness of these modalities. Namely, natural language-based interactions were more efficient when participants had small amounts of information to convey to the assistant. In contrast, for tasks that they considered more complex, the richer interactions of GUIs afforded participants with more control (e.g. direct manipulation of a list of files).

Participants suggested that the additional details and user control provided by a richer interface could make automation of complex tasks more desirable. For a meeting scheduling task that was more complex than usual, P203 said, *"If the meeting is with more people... probably I would prefer another kind of interaction, maybe a traditional one where I can see the schedule of the people."* Hence, we saw that completing complex work with a task assistant requires interactions beyond a chat interface for participants to feel comfortable with automation.

4. Discussion

Although some models propose that humans and AI assistants may be equally capable of doing certain tasks [33, 34, 35], current task assistant architectures are based on more asymmetric principles in which the AI assistant retains execution capability for many operations, as described in the allocation models [36, 7, 8]. Participants' comments and designs encourage us to re-examine *whether and when* the assistant should retain control. The four task dimensions provide guidance on *when* to delegate control to whom in automation, and the features described by participants in the context of these dimensions are examples of *how* to delegate it.

Participants' concerns about possible process and social consequences, along with concerns about working on complex tasks using natural language, made them wary of automation. Low familiarity with the assistant also raised concerns about how much they could trust the assistant to automate consequential tasks. These task dimensions can provide insight on where along Wiberg and Bergqvist's [6] Engaging Interaction through Automation scale to design, and address questions of what kinds of tasks or steps are suitable for automation [9, 10].

Despite these concerns, participants described several affordances across dimensions that could provide them with more control and hence help them feel more comfortable with automation: transparency into the task assistant's capabilities and step-by-step process, review and confirmation steps that provide users with decision-making authority, and richer interaction modalities. These affordances support user efficiency and comfort by transforming rigid workflows into human-AI collaborative processes with users in control. Thus, far from the rigid and predetermined control structures of the allocation models [36, 7, 8], we propose that **participants positioned themselves as co-creators of their task automation experience**.

Beyond these affordances, participants also expected task assistant interactions to be adaptive to the their background and needs and highly personalized to their unique ways of working. Becoming more efficient in the parts of their jobs that they personally cared about was the primary motivator that participants cited for working with a conversational task assistant. Identifying such priorities through user studies with specific workgroups can help prioritize development of task assistance that focuses on functionality that matters most.

Personalization plays a larger role as AI assistants are adopted in more contexts with broader user groups. We saw that users expected to control how they interact with the assistant on a per-task basis both in terms of the interaction and how the task was represented. We expect that as such assistants become more intelligent and widespread, the need for personalization will similarly increase and the design of these assistants should take this into account.

5. Conclusion

We conducted two studies to understand business users' automation preferences and needs for working comfortably and efficiently with a conversational task assistant.

This work identified important user considerations in the context of task dimensions for the design of conversational task assistants:

- We identified four task dimensions along which automation preferences varied: **process consequence, social consequence, familiarity, and complexity**. These task dimensions provide a human-centered perspective into **when, why, and how to delegate control** between system and user.
- Along these dimensions, we elicited several affordances to put users more in control: transparency into system actions, decision-making authority on if and how a step or task gets done, and richer interaction modalities.
- The dynamic nature of these dimensions means that users' automation preferences will change over time. Therefore, conversational task assistants should allow users to situationally adapt their task controls for a comfortable and efficient automation experience.

References

[1] T. Chakraborti, S. Agarwal, Y. Khazaeni, Y. Rizk, V. Isahagian, D3ba: a tool for optimizing business processes using non-deterministic planning, in: International Conference on Business Process Management, Springer, 2020, pp. 181–193.

[2] T. Chakraborti, V. Isahagian, R. Khalaf, Y. Khazaeni, V. Muthusamy, Y. Rizk, M. Unuvar, From robotic process automation to intelligent process automation, in: International Conference on Business Process Management, Springer, 2020, pp. 215–228.

[3] Y. Rizk, A. Bhandwalder, S. Boag, T. Chakraborti, V. Isahagian, Y. Khazaeni, F. Pollock, M. Unuvar, A unified conversational assistant framework for business process automation, arXiv preprint arXiv:2001.03543 (2020).

[4] IBM Corporation, IBM Watson Orchstrate, 2022. URL: https://www.ibm.com/products/watson-orchestrate.

[5] R. G. Farrell, J. Lenchner, J. O. Kephjart, A. M. Webb, M. J. Muller, T. D. Erikson, D. O. Melville, R. K. Bellamy, D. M. Gruen, J. H. Connell, et al., Symbiotic cognitive computing, AI Magazine 37 (2016) 81–93.

[6] M. Wiberg, E. S. Bergqvist, User experience (ux) meets artificial intelligence (ai): Designing engaging user experiences through'automation of interaction', in: 2022 Workshop on Engaging with Automation-Understanding and Designing for Operation, Appropriation, and Behaviour Change, AutomationXP 2022, New Orleans, LA, USA, April 30, 2022, CEUR-WS, 2022.

[7] R. Parasuraman, T. B. Sheridan, C. D. Wickens, A model for types and levels of human interaction with automation, IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans 30 (2000) 286–297.

[8] T. B. Sheridan, W. L. Verplank, Human and computer control of undersea teleoperators, Technical Report, Massachusetts Inst of Tech Cambridge Man-Machine Systems Lab, 1978.

[9] C. Martinie, P. Palanque, Task models based gameful design as a mean to increase engagement with automation, in: Workshop on Engaging with Automation (AutomationXP 2022) co-located CHI 2022, volume 3154, CEUR-WS. org, 2022.

[10] T. H. Soe, Automation in video editing: Assisted workflows in video editing., in: AutomationXP@ CHI, 2021.

[11] J. Gwizdka, Timely reminders: a case study of temporal guidance in pim and email tools usage, in: Chi'00 extended abstracts on human factors in computing systems, 2000, pp. 163–164.

[12] N. Singh, M. Tomitsch, M. L. Maher, A time and place for preparatory methods in email, in: Proceedings of the 14th Annual ACM SIGCHI_NZ conference on Computer-Human Interaction, 2013, pp. 1–8.

[13] M. Muller, C. Dugan, M. Brenndoerfer, M. Monroe, W. Geyer, What did i ask you to do, by when, and for whom? passion and compassion in request management, in: Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, 2017, pp. 1009–1023.

[14] J. M. Corbin, A. L. Strauss, The articulation of work through interaction, The sociological quarterly 34 (1993) 71–83.

[15] S. L. Star, A. Strauss, Layers of silence, arenas of voice: The ecology of visible and invisible work, Computer supported cooperative work (CSCW) 8 (1999) 9–30.

[16] K. Schmidt, Taking cscw seriously: Supporting articulation work (1992), in: Cooperative Work and Coordinative Practices, Springer, 2008, pp. 45–71.

[17] D. Chasanidou, B. Elvesæter, A.-J. Berre, Enabling team collaboration with task management tools, in: Proceedings of the 12th International Symposium on Open Collaboration, 2016, pp. 1–9.

[18] S. Whittaker, Supporting collaborative task management in e-mail, Human–Computer Interaction 20 (2005) 49–88.

[19] J. E. Bardram, Activity-based computing for medical work in hospitals, ACM Transactions on Computer-Human Interaction (TOCHI) 16 (2009) 1–36.

[20] V. Bellotti, N. Ducheneaut, M. Howard, I. Smith, Taskmaster: recasting email as task management, PARC, CSCW 2 (2002).

[21] N. Ducheneaut, V. Bellotti, E-mail as habitat: an exploration of embedded personal information management, interactions 8 (2001) 30–38.

[22] M. Czerwinski, E. Horvitz, S. Wilhite, A diary study of task switching and interruptions, in: Proceedings of the SIGCHI conference on Human factors in computing systems, 2004, pp. 175–182.

[23] J. Gwizdka, Email land–an exploration of email user interfaces supporting pending tasks, in: Conference Companion of UIST 2002, Citeseer, 2002, pp. 9–10.

[24] M. J. Muller, L. G. Tudor, D. M. Wildman, E. A. White, R. W. Root, T. Dayton, R. Carr, B. Diekmann, E. Dykstra-Erickson, Bifocal tools for scenarios and representations in participatory activities with users, in: Scenario-based design: envisioning work and technology in system development, 1995, pp. 135–163.

[25] M. J. Muller, Layered participatory analysis: New developments in the card technique, in: Proceedings of the SIGCHI conference on Human factors in computing systems, 2001, pp. 90–97.

[26] M. J. Muller, Retrospective on a year of participatory design using the pictive technique, in: Proceed-

ings of the SIGCHI conference on Human factors in computing systems, 1992, pp. 455–462.

[27] B. Still, J. Morris, The blank-page technique: Reinvigorating paper prototyping in usability testing, IEEE Transactions on Professional Communication 53 (2010) 144–157.

[28] G. Terry, N. Hayfield, V. Clarke, V. Braun, Thematic analysis, The SAGE handbook of qualitative research in psychology 2 (2017) 17–37.

[29] S. Park, A. X. Zhang, L. S. Murray, D. R. Karger, Opportunities for automating email processing: A need-finding study, in: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 2019, pp. 1–12.

[30] E. Goffman, et al., The presentation of self in everyday life. 1959, Garden City, NY 259 (1959/2002).

[31] J. D. Weisz, M. Muller, S. Houde, J. Richards, S. I. Ross, F. Martinez, M. Agarwal, K. Talamadupula, Perfection not required? human-ai partnerships in code translation, in: 26th International Conference on Intelligent User Interfaces, 2021, pp. 402–412.

[32] J. Drozdal, J. Weisz, D. Wang, G. Dass, B. Yao, C. Zhao, M. Muller, L. Ju, H. Su, Trust in automl: exploring information needs for establishing trust in automated machine learning systems, in: Proceedings of the 25th International Conference on Intelligent User Interfaces, 2020, pp. 297–307.

[33] I. Grabe, M. González-Duque, S. Risi, J. Zhu, Towards a framework for human-ai interaction patterns in co-creative gan applications, HAIGEN 2022: 3rd Workshop on Human-AI Co-Creation with Generative Models (2022).

[34] M. Muller, J. D. Weisz, W. Geyer, Mixed initiative generative ai interfaces: An analytic framework for generative ai applications, 2020. URL: https://computationalcreativity.net/workshops/cocreative-iccc20/papers/Future_of_co-creative_systems_185.pdf.

[35] A. Spoto, N. Oleynik, Library of mixed-initiative creative interfaces, 2017. URL: http://mici.codingconduct.cc/.

[36] P. M. Fitts, M. Viteles, N. Barr, D. Brimhall, G. Finch, E. Gardner, W. Grether, W. Kellum, S. Stevens, Human engineering for an effective air-navigation and traffic-control system, and appendixes 1 thru 3, Technical Report, Ohio State Univ Research Foundation Columbus, 1951.