

Intelligent Futures in Task Assistance

Ryen W. White, Microsoft Research, ryenw@microsoft.com

Forthcoming in Communications of the ACM, Accepted January 2022

Tasks are the primary unit of personal and professional productivity. They describe activity toward an objective. Tasks can be explicitly specified, e.g., in a to-do list, or inferred from user behavior and context, e.g., during search engine interaction. Task management software, including to-do applications such as Todoist, Google Tasks, and Microsoft To Do, is a multibillion-dollar industry, with large gains forecast over the next five years. We can improve individual and team task productivity through a mixture of education on best practices and intelligent task assistance. There are many strategies for the former, e.g., Getting Things Done (GTD) [1], and there are plentiful opportunities for the computer science community to spearhead the latter. For example, by investing more in task-related research and engineering in areas such as natural language processing (NLP), machine learning (ML), and human-computer interaction (HCI). Among other things, intelligent task assistance can help people prioritize their task backlogs, plan their daily agendas, decompose their complex tasks into manageable steps, and ensure that their commitments to others are met.

Helping people make progress toward task completion is the primary objective of task assistance. Few intelligent systems offer end-to-end completion support, with some exceptions, e.g., task-oriented dialog systems for, say, ticket reservations. Search engines provide result lists that serve as starting points for post-query navigation. Digital assistants such as Alexa or Google Assistant provide situational reminders to perform tasks at the correct time or in the correct location. Dedicated task management applications provide a means of storing and organizing tasks but rely on users to drive activities such as scheduling and completion. These challenges are not restricted to individuals. Teams may also track tasks carefully, via dedicated project management tools, e.g., Trello, Asana, and monday.com. However, task assistance requires significant human intervention, with limited system intelligence. Experienced professionals, using established project management methods (e.g., Kanban boards, Gantt charts), are often required to manage team tasks.

Task Intelligence

Putting the task management burden on humans alone is problematic considering challenges that they may face, e.g., cognitive overload and cognitive biases such as over-optimism. Only a fraction of people use task management software. Others prefer alternatives (e.g., sticky notes, emails to self), and some do not explicitly track their tasks at all. Artificial intelligence (AI) can assist with many aspects of the task lifecycle, including: 1) capturing task intentions from communications (e.g., email, meeting transcripts), content (e.g., photos, notes), and activities (e.g., searching, browsing); 2) organizing tasks to find time to complete them given a schedule and task attributes such as priority; and 3) assisting with task completion activities, e.g., contextual reminders, task automation. There is a significant opportunity to enhance people's personal and professional productivity and give them more time for other things in their lives, like having fun, via more focus on task intelligence. The need for intelligent task support may be amplified in the global pandemic and its aftermath, where home and work tasks are more intertwined, and people may require even more task assistance from AI systems.

Examples of Task Intelligence: There has been prior work on intelligent task assistance, although more is required to make these research investments production ready and assist users across the complete task lifecycle. At Microsoft Research, we have been developing a task intelligence

platform called Titanium (Ti), comprising ML models spanning several applications and many task scenarios. Examples of recent task-related research, grouped by some of the GTD themes, include:

Capture: This involves collecting, extracting, and understanding tasks from a variety of digital and analog sources. Studies have shown that many people track tasks explicitly via to-do lists¹ that can assume many forms (notepads, sticky notes, email, etc.) [2]. To help everyone, intelligent systems must reliably extract tasks from both analog and digital sources, e.g., via email intent understanding (e.g., identifying commitments made to others), observing application usage, converting handwritten task lists to digital form, and mining “tasklets” (i.e., interface automation scripts) from websites [7]. Semantic representations, including task embeddings [8], can enrich system understanding of tasks and enable better matching for applications such as task ranking and task recommendation.

Organize: This involves arranging and scheduling task execution, which can be challenging for systems given many constraints on users’ time (some latent) and uncertainty about priorities. Working in tandem with users, systems can offer support including task duration estimation [11], to help schedule time; task completion detection [12], to minimize redundant notifications; task prioritization [13], to help people determine their next actions, and complex task decomposition [14], dividing tasks into more manageable steps (microtasks), to help people make progress on their complex tasks over time and sometimes assign them to crowd workers to get help in doing so [4].

Engage: This involves supporting the completion of tasks via system recommendations acted on by users or systems completing tasks automatically on users’ behalf, with consent. Progress on robotic process automation has enabled systems to extract action graphs from websites to automatically execute multi-step tasks, such as food delivery or restaurant reservations, on users’ behalf [10]. Mixed initiative task assistance capitalizes on complementary human and machine capabilities, e.g., assisting with task completion by suggesting resources for review. Task-based recommendation systems suggest the most appropriate tasks for specific situations, personalized to the individual’s task habits [6] or based on population data about situations where people typically perform tasks [3].

To be effective in practice, and build and maintain user trust, intelligent task assistance must seek consent to access and use task and observational data, and work together with users, offering them agency over system support while providing explanations for their task assistance.

Evaluating Task Intelligence: Beyond designing and developing task intelligence technologies, there are also considerable challenges in evaluating systems that offer task assistance. These systems can be highly complex, containing many components chained together, making the attribution of good or bad performance to specific system components challenging. For example, the architecture for the Project Execution Assistant (PExA) for task-time management [9] lists at least eight different system components (managers, coordinators, predictors, explainers, etc.). We need both holistic metrics (e.g., satisfaction, success) and per component metrics (e.g., relevance, efficiency) to fully understand system performance, and even then, experimenters might face challenges with metric overload (i.e., which metric(s) to focus on?). Integrated metrics may help capture multiple facets simultaneously, but the challenge is in how to interpret these combined metrics and compile a story of system performance around them. For many tasks-related applications, tasks data are highly confidential and are not accessible for third-party human review, making it difficult to debug task intelligence in operation or to develop new AI-powered features. A lack of interpretable benchmarks, and the strong effects on situational and individual factors, mean that it is difficult to amass

¹ <https://news.microsoft.com/2008/01/14/survey-shows-increasing-worldwide-reliance-on-to-do-lists/>

sizeable and comprehensive tasks datasets, making it challenging to measure progress in research and development over time.

Lessons So Far

We have learned much from our experiences in developing and fielding task support across different applications, in Microsoft products such as Cortana, Outlook, and To Do. While users can appreciate the additional assistance, how they accomplish tasks has been shown to be highly personal [4]. We therefore need to offer help selectively and be prepared to adapt to individual user styles. Beyond the challenges with evaluating task intelligence surfaced earlier, we have also learned several additional lessons, including: 1) lightweight task capture is essential, since many people do not use task management software and some do not explicitly track their tasks at all; 2) integration with existing applications is important, since dedicated task management application usage is rare; 3) information on the task completion process is often limited, e.g., quite often we might not observe completion events or actions in the data used to train ML models, making support for task completion difficult to develop; 4) tasks span application, device, and session boundaries, making it challenging to reliably track task progress through a limited lens on user activity – other mechanisms such as in-situ surveys might be required; 5) along similar lines, multiple devices may be needed to perform tasks most effectively, e.g., a tablet plus a smart speaker for recipe preparation; 6) many factors contribute to complex task attributes such as priority (comprising both urgency and importance) and these must be accurately modeled for systems to provide optimal task assistance; 7) tasks can be highly context dependent, e.g., relevant to particular times and/or locations, with plenty of other situational aspects (e.g., the application context) that need to be understood; 8) task feasibility is an important dimension in offering intelligent support that has not been a focus to date, including whether users have the tools, resources, and data required to complete the current task; 9) task repetition is common, both within specific users and in general in the population – intelligent systems could do more to assist with this repeat behavior (e.g., by surfacing previous solutions, decompositions, etc.), and; 10) tasks can reveal sensitive information explicitly (e.g., passwords in to-do lists) or implicitly (e.g., n-grams mined from activity streams).

The Road Ahead

Despite the large market opportunity and significant need for this type of task support, there is still much to do in task intelligence. Computer science researchers and practitioners are well placed to develop and apply advances in technology to help people focus on what matters to them and get more done with less effort. In what follows, we offer eight examples of future directions for task intelligence research (grouped by GTD theme, plus three foundational directions that span all themes), each with general computer science research topics that would enable progress, and tasks-specific research topics to accompany them (see Table 1).

1. Better Understand Tasks to Enable More Intelligent Assistance (Capture): Tasks are expressed in natural language in many forms, e.g., as commitments and requests in interpersonal communications or as short (few-word) task entries in to-do lists. To be actionable by intelligent systems, we need consistent task representations (e.g., a “task2vec” mapping). Research on language understanding can help here, but more focus on tasks is needed given the differences in meanings and user intents from words and phrases appearing standard prose. Research on semantic understanding for extremely short texts such as search queries or social media postings is also relevant. In addition, more research is needed on intent understanding and task understanding from observational activity data, including application usage and screen understanding (e.g., a “pixel2vec” representation for tasks and activities).

2. Enhance Task Planning and Support Goal Attainment (Organize): More research on task and time management and methods to model high-level user goals (e.g., planning a successful birthday party) would help intelligent systems better reason over activities and how they can be optimally organized with respect to goals. We must also explore more sophisticated tasks-specific inferences, e.g., task feasibility determination (i.e., can this task be done in this context?), more granular task progress estimation (not just binary prediction [12]), task repetition (i.e., has this task been done before and if so, how?) and task decomposition (breaking down not only complex tasks [14], but goals too).

3. Improve Task Outcomes via Human-AI Cooperation (Organize and Engage): Additional research on human-AI teaming in general and in the context of intelligent task assistance could help improve task outcomes (both ongoing and future). We can develop systems that can coach and guide users on best practices (e.g., hints and tips for more effective task management), nudges toward positive task outcomes (e.g., task reminders), improve user understanding and build user trust through system explanations (e.g., clarifying why tasks were recommended and prioritized), and help users action their tasks by cooperating with them (e.g., via conversations in task-oriented dialog systems).

4. Boost Task Productivity for Individuals and Teams (Engage): The new future of work emerging from the global pandemic will create an even more pressing need for intelligent task assistance across work and life boundaries. As part of this, understanding hybrid work and work life balance is ever more important, as are gig work mechanisms such as task marketplaces (e.g., TaskRabbit²), where tasks are performed on behalf of users, and task automation, where systems complete tasks automatically. Further research on “microproductivity” [4] will help people make incremental progress on their tasks by first identifying microtasks and then matching them with short snippets (5-10 minutes) of available time. More research on teamwork and specifically, project management: learning from roles, task delegation, task dependencies, etc. will enable intelligent systems to one day improving team productivity via automatic task assignment, task sequencing, and cross-team load balancing.

5. Provide Holistic Support to Facilitate Task Completion (Engage): Intelligent support should be integrated within existing applications to foster uptake and span different applications and/or devices, e.g., as part of a task AI fabric such as Titanium (mentioned earlier), to support more comprehensive task understanding and end-to-end task completion. Tasks can persist over time, may require multiple applications (e.g., connecting a to-do list with a web search engine for actionable, search-friendly tasks) or need multiple devices (e.g., connecting a smartphone or tablet with a smart speaker for complex, multi-step tasks), hence cross-session, cross-application, and cross-device workflows become important. More research on intelligent assistance in general, irrespective of tasks, would also help build algorithms and experiences that can still assist users with tasks and inform further task-related research and development.

6. Refine Assistance Over Time via Feedback Loops (Foundations): The first of several key future research directions that span all stages of the task lifecycle is feedback loops, i.e., collecting implicit and explicit feedback from users and refining/adapting intelligent systems accordingly. General research on personalization and contextualization is relevant here, as are tasks-specific variants, e.g., personal task prioritization and contextual task recommendations. Beyond the individual and their current situation, more research is needed on cohort modeling, i.e., learning from others

² <http://www.taskrabbit.com>

performing similar tasks about, for example, which resources are typically used for the current task and in what situations tasks are typically feasible.

7. Expedite Progress via New Datasets and Evaluation Methods (Foundations): One reason that there has not been more interest in task intelligence in general is a lack of large-scale tasks data to train and test ML models. As is often the case with nascent ML application domains, labeled datasets are challenging to obtain. It requires some creative thinking to make progress on building such datasets: in some of our research on task duration estimation, we had to collect our own large dataset of task titles and associated task durations directly from crowd workers. Shared tasks datasets, task-related competitions (e.g., the Alexa TaskBot challenge³), benchmarking studies, and living laboratories, would all help drive further advances in task intelligence. Deeper study of evaluation metrics and community agreement on primary metrics would also expedite progress.

8. Build and Maintain User Trust (Foundations): Trust must be a cornerstone of anything we do involving user data. User consent is critical and advances in areas such as privacy-preserving machine learning, FATE (fairness, accountability, transparency, and ethics), and client-side learning are clearly relevant and will make tasks applications more robust. Within these well-explored areas there are tasks-specific research challenges, such as removing and securing sensitive information (passwords, bank details, medicine, etc.) from to-do tasks datasets, reminding users of the risks of including such sensitive information in their to-do lists, de-biasing team task auto-assignments (which may be learned from skewed historic data), and performing private and secure (client-side) inference on observational data.

Overall, given the existing capabilities and the many avenues for additional research, I believe that the future is bright for task intelligence. There are considerable commercial and scientific opportunities in this area and the computer science community should do more to lead the way.

³ <https://developer.amazon.com/alexaprize/challenges/current-challenge/taskbot/faq>

Table 1. Future directions in task intelligence, with examples of research topics, both general topics in computer science (CS) (especially ML, NLP, and HCI) and topics focused on task management.

	Future Direction	Examples of general CS research topics that can contribute	Examples of tasks-specific research topics
Capture	Better Understand Tasks to Enable More Intelligent Assistance	Language understanding (especially short texts), Intent understanding, Context modeling	Task understanding and modeling from language and activities (task2vec), Screen to task understanding (pixel2vec), Mining action graphs from websites
	Enhance Task Planning and Support Goal Attainment	Time management, Goal understanding and modeling, Situational intelligence	Task feasibility, Task progress estimation, Task repetition, Complex task (or goal) decomposition into microtasks
Organize	Improve Task Outcomes via Human-AI Cooperation	Nudges, Coaching, Explanations, Human-AI teaming, Conversational AI	Task reminders, Task suggestions (hints, tips, etc.), Task-oriented dialog systems
	Boost Task Productivity for Individuals and Teams	Hybrid work, Work-life balance, Gig economy, Robotic process automation, Teamwork (roles, delegation, dependencies, etc.)	Microproductivity, Task marketplaces, Task automation, Project (team task) management (assignment, sequencing, load balancing, etc.)
	Provide Holistic Support to Facilitate Task Completion	Intelligent assistance, Longitudinal user histories and analysis, Cross- and multi-device experiences	Cross-session task continuation, Cross-app task workflows (e.g., to-do applications to search engines), Cross- and multi-device task completion
Engage	Refine Assistance over Time via Feedback Loops	Personalization, Contextualization, User modeling, Cohort modeling	Task prioritization, Task recommendations
	Expedite Progress via New Datasets and Evaluation Methods	Datasets, Experimentation	Shared tasks datasets, Tasks competitions (e.g., Alexa TaskBot challenge), Tasks-based evaluation metrics
	Build and Maintain User Trust	Privacy preserving machine learning, FATE, Client-side inference	Removing and securing sensitive information in to-do task data (passwords, etc.), Private inference on observational task activity data, De-biasing team task management assignments
Foundations			

References

1. Allen, D. (2015). *Getting Things Done: The Art of Stress-free Productivity*. Penguin.
2. Bellotti, V., Dalal, B., Good, N., Flynn, P., Bobrow, D.G., and Ducheneaut, N. (2004). What a to-do: Studies of task management towards the design of a personal task list manager. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems* (pp. 735-742).
3. Benetka, J.R., Krumm, J., and Bennett, P.N. (2019). Understanding context for tasks and activities. In *Proceedings of the 2019 ACM SIGIR Conference on Human Information Interaction and Retrieval* (pp. 133-142).
4. Cheng, J., Teevan, J., Iqbal, S.T., and Bernstein, M.S. (2015). Break it down: A comparison of macro- and microtasks. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 4061-4064).
5. Haraty, M., McGrenere, J., and Tang, C. (2016). How personal task management differs across individuals. *International Journal of Human-Computer Studies*, 88, 13-37.
6. Kessell, A. and Chan, C. (2006). Castaway: A context-aware task management system. In *Proceedings of the ACM SIGCHI Extended Abstracts on Human Factors in Computing Systems* (pp. 941-946).
7. Li, Y., and Riva, O. (2021). Glider: A reinforcement learning approach to extract UI scripts from websites. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 1420-1430).
8. Mehrotra, R. and Yilmaz, E. (2017). Task embeddings: Learning query embeddings using task context. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management* (pp. 2199-2202).
9. Myers, K., Berry, P., Blythe, J., Conley, K., Gervasio, M., McGuinness, D.L., Morley, D., Pfeffer, A., Pollack, M., and Tambe, M. (2007). An intelligent personal assistant for task and time management. *AI Magazine*. 28(2):47-61.
10. Riva, O. and Kace, J. (2021). Etna: Harvesting action graphs from websites. In *Proceedings of the 34th Annual ACM Symposium on User Interface Software and Technology* (pp. 312-331).
11. White, R.W. and Hassan Awadallah, A. (2019). Task duration estimation. In *Proceedings of the 12th ACM International Conference on Web Search and Data Mining* (pp. 636-644).
12. White, R.W., Hassan Awadallah, A., and Sim, R. (2019). Task completion detection: A study in the context of intelligent systems. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 405-414).
13. Zhang, C., Kiseleva, J., Jauhar, S.K., and White, R.W. (2021). Grounded task prioritization with context-aware sequential ranking. *ACM Transactions on Information Systems*, 40(4):68.
14. Zhang, Y., Jauhar, S.K., Kiseleva, J., White, R.W., and Roth, D. (2021). Learning to decompose and organize complex tasks. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics* (pp. 2726-2735).