Learning About Work Tasks to Inform Intelligent Assistant Design

Johanne R. Trippas
Damiano Spina
Falk Scholer
johanne.trippas@rmit.edu.au
damiano.spina@rmit.edu.au
falk.scholer@rmit.edu.au
RMIT University
Melbourne, Australia

Ahmed Hassan Awadallah
Peter Bailey
Paul N. Bennett
Ryen W. White
hassanam@microsoft.com
pbailey@microsoft.com
pauben@microsoft.com
ryenw@microsoft.com
Microsoft
Redmond, USA

Jonathan Liono
Yongli Ren
Flora D. Salim
Mark Sanderson
jonathan.liono@rmit.edu.au
yongli.ren@rmit.edu.au
flora.salim@rmit.edu.au
mark.sanderson@rmit.edu.au
RMIT University
Melbourne, Australia

ABSTRACT

Intelligent assistants can serve many purposes, including entertainment (e.g. playing music), home automation, and task management (e.g. timers, reminders). The role of these assistants is evolving to also support people engaged in work tasks, in workplaces and beyond. To design truly useful intelligent assistants for work, it is important to better understand the work tasks that people are performing. Based on a survey of 401 respondents' daily tasks and activities in a work setting, we present a classification of work-related tasks, and analyze their key characteristics, including the frequency of their self-reported tasks, the environment in which they undertake the tasks, and which, if any, electronic devices are used. We also investigate the cyber, physical, and social aspects of tasks. Finally, we reflect on how intelligent assistants could influence and help people in a work environment to complete their tasks, and synthesize our findings to provide insight on the future of intelligent assistants in support of amplifying personal productivity.

ACM Reference Format:

Johanne R. Trippas, Damiano Spina, Falk Scholer, Ahmed Hassan Awadallah, Peter Bailey, Paul N. Bennett, Ryen W. White, Jonathan Liono, Yongli Ren, Flora D. Salim, and Mark Sanderson. 2019. Learning About Work Tasks to Inform Intelligent Assistant Design. In *2019 Conference on Human Information Interaction and Retrieval (CHIIR '19), March 10–14, 2019, Glasgow, United Kingdom.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3295750.3298934

1 INTRODUCTION

Intelligent assistants such as Amazon Alexa or Microsoft Cortana provide a useful way for people to manage many of their tasks, including personal and work-related activities. In recent years, there has been growing interest in applications of these assistants in workplaces to empower employees, through offerings such as Alexa for Business[1] and Cortana Skills Kit for Enterprise.[2] Despite the potential for these assistants to help people complete their work tasks (at work, at home, or on-the-go), penetration of these assistants in workplaces is limited [19], and task support is restricted to low-level tasks such as controlling devices, seeking information, or entertainment [26]. In work settings in particular, intelligent assistants are mostly used for basic tasks such as voice dictation, calendar management, and customer/employee support [19]. To increase the uptake of intelligent assistants for work tasks, a better understanding of the tasks that people perform, and how next-generation intelligent assistants could support them, is needed.

There are many ways to understand tasks and activities, including diary studies [17], naturalistic field studies [32], and lifelog analysis [13]. A commonly used way to characterize tasks and activities is through statistical time-use surveys [6], enabling an understanding of which activities people are engaging in and for how long. Examples include the average number of hours that individuals spend on activities including traveling, working in paid and unpaid jobs, social life, and personal care. However, these surveys do not provide a fine-grained understanding of how these tasks and activities are completed. Understanding how people complete activities—including important aspects such as the devices used and the involvement of other individuals—can provide insights into how to support these activities with digital assistants. Other studies have examined daily information needs [8], tasks in communications [4], or information-seeking [16, 22]. Recently, the influences of cyber, physical, and social behaviors were used to analyze information-seeking activities indoors [29, 30]. However, these studies do not focus on work tasks, including the impact of task properties (cyber, physical, or social factors), or the support that intelligent assistants can offer to help workers complete their tasks.

To inform the design of intelligent assistants to support work tasks, we developed a survey to better understand people's everyday tasks and task performance. We focus on information workers, since their activities frequently require digital support. We distributed the survey broadly and received around 400 responses. In this paper,

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
CHIIR '19, March 10–14, 2019, Glasgow, United Kingdom
© 2019 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-6025-8/19/03…$15.00
https://doi.org/10.1145/3295750.3298934

[1] https://aws.amazon.com/alexaforbusiness/
[2] https://blogs.microsoft.com/ai/cortana-for-enterprise/

we focus on tasks described by survey respondents as *work-related*. We apply thematic and quantitative analysis to provide deep insight and understanding into the work tasks that people perform. The main contributions of this paper are:

- A comprehensive classification of work-related tasks.
- An analysis of the key characteristics of tasks, focusing primarily on temporal aspects such as task duration.
- An investigation of the cyber, physical, and social (CPS) properties of tasks, including device usage.
- Design recommendations for how intelligent assistants could support work tasks, both on demand and proactively, drawing on insights from the survey.

The remainder of the paper is structured as follows. In Section 2, we describe related work, including task taxonomies, intelligent assistants, and tasks and activities in email and search. Section 3 describes the survey methodology, including the pilot survey. Section 4 describes the results of the analysis. We discuss the results and their implications in Section 5 and conclude in Section 6 with a summary of our findings and potential avenues for future work.

2 RELATED WORK

In this research, we characterize tasks performed by information workers and examine opportunities for intelligent assistants to support them.

2.1 Task Taxonomies

Previous work on understanding human activities has attempted to build taxonomies of tasks and human activities. Kahneman et al. [17] used a Day Reconstruction Method (DRM) to assess how people spend their time and how they experience various activities in their lives. They asked study participants to construct a short diary by segmenting their day into episodes and assigning a name, a start time, and an end time for each episode. They used this information to construct an activity taxonomy (e.g. socializing, eating, working, commuting, etc.) and interaction partners (e.g. friends, relatives, co-workers, etc.). This taxonomy was used in defining activities for activity identification in lifelog data [13].

Time-use surveys have been carried out at different levels for many decades [6]. A time-use survey is a statistical survey that aims to report data on how people spend their time. They typically characterize time spent in different categories such as paid work, unpaid work (e.g. domestic work, voluntary work, etc.), social life and leisure (e.g. sports, mass media, etc.), and personal care (e.g. sleeping, eating, etc.) [6]. The objective is to identify, classify and quantify the main types of activity that people engage in which has implications on a wide range of policy concerns [38].

Other research has focused on characterizing tasks carried out in team settings [40]. The study noted the increased use of team-based work structure and suggested that studying work tasks should take team characteristics into account. Different types of tasks such as managing others, advising others, problem solving, and social interaction were considered. Analysis showed that different types of teams engage in multiple types of tasks to accomplish their mission.

Other studies have focused on specific types of tasks. For example, Church et al. [8] studied the daily information needs of people through a large-scale, in-depth, quantitative investigation. They conducted a study spanning a three-month period and involving more than 100 users. The study gathered in situ insights into the types of needs that occur from day to day, how those needs are addressed, and how contextual, technological, and demographic factors impact on those needs. They showed that there are recurring daily information needs such as finding availability and information about people. They also showed that people employ two main modalities for satisfying daily information needs (Internet access and asking others for help). Similar work has focused on understanding mobile information needs [9, 34] and found that the majority of such needs happen where people are away from their familiar contexts and that geographical needs were very popular.

Our research also aims to characterize tasks, but we focus on understanding everyday work-related tasks and task performance with the objective of informing the design of intelligent assistants to support work tasks. We also examine cyber, physical, and social aspects of these tasks. Finally, we examine the tools currently used by information workers to accomplish these tasks and identify opportunities for intelligent assistants to provide support.

2.2 Intelligent Assistants

Intelligent digital assistants on mobile devices, personal computers, and smart speakers have recently gained considerable attention as increasing number of users interact with them to fulfill their information needs and complete their tasks. This has motivated work on user interaction studies to characterize the usage of these assistants. Mehrotra et al. [26] investigated the different use cases of a desktop-based digital assistant. The analysis reveals that, given the conversational nature of user interactions, longer sessions (i.e. sessions with a large number of queries) are more common than in the traditional web search paradigm. Exploring the different use cases, they observed that users go beyond general search and use a digital assistant to accomplish several tasks such as time management (setting reminders and meetings), entertainment (e.g. playing music), communication (e.g. sending text messages), etc.

The interest in new form factors for intelligent assistants, such as smart speakers, motivated new studies of the most common use cases for intelligent assistants on these devices. A recent report [19] focused on studying how people use smart speakers. The findings showed that the living room has become the most common location choice at 46% for smart speaker owners, closely followed by the kitchen (41%) and bedrooms (37%). Trailing far behind is the work office at 2.7%, indicating that such assistants still have a long way to go to penetrate the workspace. They report that listening to music is the top use case, followed by checking the weather, asking questions, and setting timers and alarms. Other interesting use cases include calling and messaging, and smart home control.

Understanding how intelligent assistants are used in a workplace setting is less studied and not very well understood. A recent report [37] focused on this angle by conducting a survey of over 500 people from different industries and organization sizes. The study showed that the penetration of intelligent assistants in the workplace is still limited, and that they are mostly used for basic tasks such as voice dictation or calendar management.

These studies highlight two important aspect that motivate this work. The first is that intelligent assistant usage for work-related tasks is rather limited, and developers of intelligent assistants have a considerable challenge in order to penetrate the workspace. The second is that the current tasks being supported by intelligent assistants are limited in scope, with basic question answering, device control, and calendar management being the most frequent use cases. This highlights the need for a detailed study of tasks performed by information workers, and the potential for intelligent assistants to support users with these tasks.

2.3 Tasks in Email and Web Search

Web search and email have been extensively used to understand what tasks users are engaged with—even though both give only a partial view of tasks and activities. Prior research has treated emails as a communication tool for workplace activities or a task management resource [4]. Kushmerick and Lau [21] formalized e-commerce activities as finite-state automata, where transitions among states represent messages sent between participants. Dredze and Wallach [12] used user-generated activity labels and classified emails into activities using overlapping participants and content similarity. Shen et al. [33] predicted tasks associated with an incoming email by leveraging email sender, recipients and distinct subject words. Dredze and Wallach [12] and Qadir et al. [28] used approaches inspired by topic modeling to use content, people and thread information to identify different activities in email.

Tasks and activities have also been extensively studied in the context of Web search. Kelly and Belkin [18] conducted a longitudinal study to elicit tasks that were of current interest, or that were expected to be of interest, to subjects doing web search. Subjects were asked to think about their online information-seeking activities in terms of tasks and to create personal labels for each task. They used this information to produce the following nine task groupings: *Academic Research*, *Entertainment*, *News and Weather*, *Personal Communication*, *Shopping and Selling*, *Teaching*, *Hobbies and Personal Interests*, *Travel*, and *Jobs/Career*. Using that data, White and Kelly [39] developed different implicit feedback models for each task grouping. Other work focused on extracting tasks from search logs [16, 24], identifying cross-session search tasks [1, 20] or supporting complex search tasks that are composed of several smaller sub-tasks [14]. Research on developing taxonomies of information needs in web search [5], understanding searchers' information goals [31], and understanding tasks via action-topic pairs over both search and browsing logs [3] is also relevant in this context.

While we consider communication and information seeking, our objective is to come up with a broader characterization of tasks that covers several additional task types. Additionally, we do not consider how to improve task support in email clients or search engines but rather focus on characterizing the opportunities for intelligent assistants to provide more comprehensive task support to their users.

3 METHODOLOGY

We conducted an online survey[3] using Qualtrics[4] to gather information about tasks which are carried out in a work context. The survey consisted of both open-ended and multiple-choice questions, as well as screener questions such as age, location, employment status, and highest completed education.

The aim of the survey was to understand task behaviour in daily (work) life by capturing the range of tasks and sub-tasks that people perform, and their motivation, task duration, or task regularity. We then analyzed which tasks are completed most often, interruptions while performing the task, CPS (Cyber-Physical-Social) activities involved, and which planning tools are used.

Participation was voluntary and anonymous. Our call for participation informed participants[5] that the survey was titled "Exploration of Human Tasks in Daily Life".

3.1 Survey

An initial pilot survey was conducted with a panel constrained to participants located in Australia, using a sampling approach that followed national demographics based on participant age. Further restrictions were based on gender and population distribution according to states and territories in Australia. People in the Qualtrics database meeting these criteria were randomly selected and emailed an invitation for participating in the online survey.

Pilot data was collected from 106 participants; analysis of the responses helped to estimate the quality of the data collected, and to revise and extend the set of questions. For instance, we revised the survey to include further questions about tasks such as their regularity, interruptions that occur while performing the task, and which planning tools are used.

Moreover, the analysis of the initial pilot showed that respondents were often not employed or retired, and often had a high school degree as the highest level of education. Since respondents were allowed to provide tasks from their everyday life, it meant that many of the tasks were short and household-related. As a result, participation restrictions were developed for the main survey, requiring that respondents:

- Were between 25–59 years old;
- Were employed or self-employed;
- Had completed a bachelor's degree or higher.

For the main survey, the panel was constrained to participants located in Australia, the United Kingdom, and the United States. People in the Qualtrics database meeting the above criteria were randomly selected and emailed an invitation to participate in the online survey. The survey was conducted from May 17 to July 2 2018, and sent out to 833 individuals. 410 respondents completed the survey: a 49% response rate. The completed responses were manually checked, and 9 were identified as invalid, leaving a final number of 401 respondents, whose data is analyzed in this paper.

Overall, participants were required to supply information about four tasks, each of which took a different duration (less than 30 minutes; 30 minutes to two hours; two hours to a day; more than a day). The survey questions are attached in Appendix A. In the online setup, the same questions were displayed for each of the four tasks that participants supplied, and the word "task" was automatically replaced with their corresponding answers for Question 1.

While the definition of *information worker* is not agreed upon in the literature, by sampling from professionals with a degree, we aim

[3]Survey reviewed and approved by the Human Research Ethics Committee at RMIT University (SEHAPP 11-18).
[4]https://www.qualtrics.com/

[5]We use *respondents* and *participants* interchangeably.

to characterize the tasks of people whose work commonly requires digital information support.

3.2 Categorization of Free Text Results with a Thematic Analysis Approach

Thematic analysis involves identifying, analyzing, and reporting patterns (themes) within the qualitative data [10, 11]. This method allows for analyzing qualitative data in an accessible and theoretically flexible manner and is often seen as a fundamental technique for analysing qualitative data [10]. We adopted the six-step process as outlined by Clarke and Braun [11]: (Step 1) familiarizing self with data; (Step 2) generating initial codes; (Step 3) searching for themes; (Step 4) reviewing themes; (Step 5) defining and naming themes; and (Step 6) producing the report.

The first pass of steps 1–3 were completed by one author, leading to an initial list of detailed low-level categories and a set of candidate themes. Steps 4 and 5 were then iterated over by a group of authors. The potential merging of categories, and alternative categories, were proposed and discussed; if agreement was reached, the proposed changes were introduced. The final set of higher level categories are the result of an iterative process and in-depth discussions among the assessors. All categories were developed only from the specific task description that respondents entered in response to Survey Question 1 (“List a task that you carried out in a work setting”). A task was allocated to a single category; if a multi-faceted task was specified (e.g. “write an email to advance a project”) then only the first facet was used to determine the category (in this case, *Communication*). The final categorization scheme, consisting of 14 task categories, is presented in detail in Section 4.2.

4 RESULTS

In this section we report the tasks that respondents performed in a work context based on our categorization scheme. We further analyze the tasks based on time-use categories; duration, continuity and regularity; and Cyber-Physical-Social aspects.

4.1 Demographics

We collected a total of 401 individual responses to the survey, originating from people living in Australia (139 respondents), the United Kingdom (133 respondents), and the United States (129 respondents). A total of 221 respondents identified as being female, 179 respondents identified as being male, and 1 respondent preferred not to answer this question. Respondents cover the entire spectrum of ages we aimed to target (25–59 years old), balanced across the three countries, with a small skew towards younger age brackets: 25–29 years old (72 respondents); 30–34 (77); 35–39 (66); 40–44 (61); 45–49 (63); 50–54 (41); and 55–59 (21). According to the time-tracking system embedded in Qualtrics, participants spent a median of 16 minutes to complete the survey.

We used occupation categories taken from the Australian Bureau of Statistics [2] for respondents to identify their occupation. All eleven occupation categories were covered at least by 4 respondents. The majority of respondents (166) identified as being a *qualified professional*, followed by *clerks, skilled office, sales and service staff* (54); *senior management in large business organizations* (51); and *associate professionals* (47).

The respondents provide a broad view of professionals that are currently working, of an age consistent with primary career years, and having at least a bachelor’s degree. These professionals are likely to perform substantial (and diverse) digital information work.

4.2 Characterizing Information Workers’ Tasks

All respondents were asked “*List a task you carried out in a work setting*”. Respondents had to describe four different tasks they carried out depending on the length of time it took them to complete the task (less than 30 minutes, longer than 30 minutes, longer than two hours, and more than a day). We classified all free text input for the question according to the method described in Section 3.2, to obtain the following categories:

Communication. This category consists of tasks related to individuals exchanging information directly, either synchronously or asynchronously, e.g. calling (phone or online), emailing, face-to-face meetings (direct or via an online meeting platform). The emailing section of the communication category consists of many subgroups which can be divided into the action people are taking, such as replying, reading, sending, checking, and deleting.

Documentation. This category consists of tasks related to creating or refining written documents. Many respondents described compiling or writing different types of reports: annual, weekly, etc. Many other respondents described actions such as typed, designed, or edited without specifying a document type. Some responses included in this category indicated they made “notes”. Thus, this category is focused on documents. We do not include emails in this category since the primary purpose of email message creation is generally *communication.*

Admin and Management. This is a broad category covering tasks to operate a workplace: recruitment and staff inductions, management of people, management of files including printing, copying or scanning a document, and general paperwork.

Planning. This category consists of tasks which are related to preparing, engineering, planning, or booking items such as arranging events, meeting preparation, booking travel, etc. Some of these tasks include immediate and precise planning such as “arrange a meeting with an international client” or “created plan and agenda for an upcoming meeting with participants in person and via skype”, as well as broader future planning such as “monthly planning”.

Education. This category consists of two major subgroups: teaching and learning. Common examples of the first subgroup include “teaching a class” or “training new employees”. Examples of the second subgroup include “undertook a training course” or “attending training”. A third minor subgroup is marking work which is mostly undertaken in an educational setting such as grading assessments, papers, or exams.

IT. This category consists of software- or hardware-related tasks. Software-related tasks include maintaining websites, writing and designing software, or setting up databases. The hardware-related tasks include fixing computers or printers.

Finance. This category includes monitoring and recording expenditures; bookkeeping tasks such as tax, bills, and payroll; as well as budgeting and invoicing.

Physical. This category consists of three subgroups: cleaning, sporting activities, and physical labour. All three require physical activity in order to complete the task.

Problem solving. This category is associated with respondents who carry out research: e.g. conduct, collate, or analyze input, which includes searching and researching.

Low-level. This category consists of data entry tasks: gathering, collection, or importing. Other responses which are also related to this low-level grouping are tasks related to spreadsheets such as creating, filling out, or updating a spreadsheet.

Project. This category consists of responses which explicitly mentioned making progress or working on a project such as collaborating, completing, or finalizing a project.

Customer care. This category groups responses which relate to helping clients or customers (patients are considered customers in this setting). Responses include dealing with, serving, or taking care of customers.

Meals and breaks. This category has two major subgroups: one related to sustenance and one describing the act of pausing. Examples of the first subgroup include: breakfast, eating lunch, or making a meal. Such responses were often seen as a break in a work setting. Therefore we grouped these tasks with the second subgroup, "having a break".

Travel. This category includes tasks related to transit such as flying to a client, driving to an interview, or transporting goods, and are all categorized as travel.

Invalid. As a result of the annotation process, a total of 1,035 responses have been classified to one of the task categories described above. The remaining 569 (35.5% of original) have been categorized as *Invalid*. These include responses explicitly indicating that no task was completed for a particular time length; generic responses (e.g. "work in progress"); responses that were not clear (e.g. "desk job work") or not applicable (e.g. "all the details").

4.3 Tasks According to Time-use Survey

In order to understand the different kind of tasks respondents completed, participants categorized their own tasks according to the time-use survey categories [38].[6] This survey is designed to understand how people behave in a 24-hour period and provide a starting point to an existing classification schema since no existing categorization schema exists specifically for tasks.

The three most commonly reported categories on which time was spent were *Working and work-related activities* (540 tasks, 52.2%), *Educational activities* (152 tasks, 14.7%), and *Professional and personal care services* (110 tasks, 10.6%). Together, these three covered 77.5% of the annotated tasks (802 out of 1,035 tasks).

[6] https://unstats.un.org/unsd/gender/timeuse/

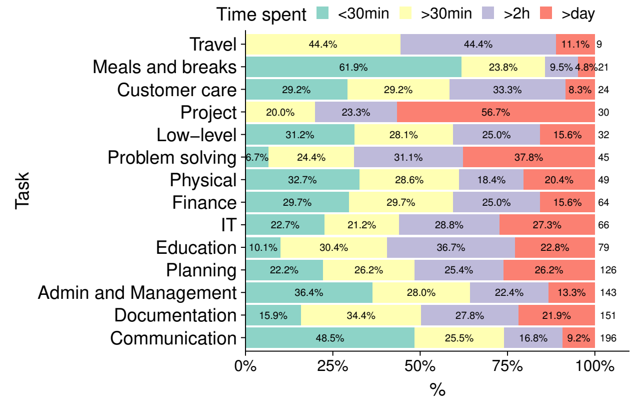

Figure 1: Tasks reported by respondents, sorted by frequency. For consistency, we maintain this sort order for the figures shown in the remainder of the paper.

4.4 Task Duration, Continuity and Regularity

The amount of time that it takes to complete a task is likely to have implications for the kind of support that can be offered, and a fundamental consideration of the survey design was therefore to gather information about tasks of different durations. For example, longer tasks may provide an opportunity for task support such as: assisting in remembering the state for incomplete tasks, assisting in breaking the task into smaller portions, resuming the task on another digital device, or leveraging idle time to perform other work if the task does not require full attention. Likewise, tasks which are interrupted are opportunities for assisting resuming the task. Therefore we ask:

- What is the distribution of task duration among task types? To answer this, we characterize task types in terms of those where more than one third of the responses indicated the task took at least two hours. We then give examples of particular responses where digital support is more likely and those where the benefit of digital support seems less obvious.
- What types of tasks are interrupted most frequently? To answer this, we characterize interruptions by quantifying the blocks of time needed to complete the task and again report example responses where digital support is more or less likely.
- Are tasks performed regularly? To answer this, we look at the responses of Questions 14 and 15 of the survey (see Appendix A).

Figure 1 shows the duration of different type of tasks. The graph shows that all the task types besides *Meals and breaks*, and perhaps *Communication*, include instances that take a considerable amount of time. *Project* tasks are typically long: 56.7% take more than a day and no responses indicated less than 30 min. Digital support would be useful in tasks such as "collaborating on a project" or "working on a new business project". Travel tasks reported in the survey also take at least more than half an hour. An intelligent assistant is not necessarily useful in this particular type of task (e.g. "transporting goods"), unless it is part of a larger task (e.g. "travelled to client"). *Problem solving* tasks are also typically long (37.8% more than a day; 31.1% more than 2 hours). In tasks such as "searching for a resolution to a problem", "data analysis" or "inventory analysis", an intelligent assistant may be useful by providing access to search, recommendation and summarization tools.

Figure 2 (left) shows how many blocks of time the participants spent to complete the task. For all categories, the majority of tasks were reported as completed in a single block of time, without interruptions. In particular *Breaks and meals* and *Travel* were all completed with no interruptions. Responses also suggest that it is not straightforward to quantify blocks of time: the "More than 5 blocks of time" option is typically preferred to indicate interruptions. *Planning*, *Problem solving* and *Low-level* are the tasks that have more interruptions (e.g. "preparing financial budget", "planning for teaching", "research", or "data collection"). *Planning* and *Problem solving* are typically long duration tasks. *Education* tasks are also long, but often taken in a continuous block of time (e.g. "lecturing"). *Low-level* tasks have interruptions, but are not necessarily long (e.g. "data entry").

Figure 2 (right) shows the regularity of tasks that are atomic or part of a larger tasks or goals (*subtasks*). Most of the tasks are performed regularly (86.7% for subtasks; 69.9% for atomic tasks; 79.6% in total). This suggests that an intelligent assistant may *see* multiple instances of a task, and therefore has more chances to identify patterns and respond proactively (e.g. suggest to perform a task, see Section 5). The figure also shows that tasks that are part of a larger task or goal (subtasks) are very likely to be tasks carried out regularly (86.7%).

4.5 Cyber-Physical-Social Aspects of Tasks

Cyber (Electronic Devices and Online Activities to Perform Tasks). Of the 1,035 tasks categorized in the taxonomy described in Section 4.2, in 71.9% (744) involve the use of electronic devices to complete a task, while 29.1% (291) of the time no electronic devices are used to complete a task. The leftmost bar chart in Figure 3 shows the number of devices used for each task category. The chart shows that devices are used less when performing *Physical* and *Meals and breaks* tasks. On the other hand, *Problem solving*, *Customer care*, and *Planning* tasks involve the use of multiple devices at least 25% of the time.

Respondents had a free text field to indicate which electronic devices they used to complete the task. We categorized these devices into a set of categories including *Desktop*, *Laptop*, *Phone*, *Tablet*, *Specialized equipment*, *Copier/Printer/Scanner* and *Other*. From the tasks that identified the use of electronic devices, personal computers are used most of the time (92.7%), with desktops (54.0%) being more widely used than laptops (37.7%). However, no physical tasks are performed with desktops and laptops. People also tend not to do complex tasks with their phones and tablets, e.g. tasks associating with other devices, such as *Documentation*, *IT*, *Finance*, and *Project* tasks. The *Specialized equipment* category includes devices such as carpet shampoo machines, dryers, electronic tills, or soldering irons. Categories which received less than 1% of the responses include instances such as "computer databases", "cloud computing", or "servers".

If the participants indicated the use of a device, we also asked whether they include any online activities to complete the task. We asked them to choose any combination of: Web browsing, Web searching, performing online transactions, using a specific app, online chatting, e-mailing, using a digital assistant, or other.

Figure 3 (right) shows the number of online activities for the different tasks. Most of the tasks involved some type of online activity, being *Physical*, *Meals and breaks*, and *Travel* the ones that involved the lowest. *Project* tasks involve multiple online activities: Web browsing (12), Web searching (14), e-mailing (12), and online chatting (10). Web browsing, Web searching and e-mailing are also the main online activities involved in other tasks, especially in *Documentation*, *Communication*, *Planning*, and *Problem solving*.

For *Low-level* and *Finance* tasks, "other" and "using a specific app" were the most reported online activities, respectively.

Physical. Respondents answered a question "*Did you engage in any physical activities in order to complete the task?*". If they answered negatively, they skipped ahead to the next questions. However, the respondents who answered affirmatively (18.1%) were asked a follow-up question about which kind of physical activity they performed.

As seen in Figure 4 (left) the categorization provides us with an overview of which tasks required physical activity in order to be completed. As may be expected, the *Physical* tasks category has the highest percentage of all categories which involved physical activity (69.4%); the key subcategories inside this grouping include cleaning and other physical work.

The *Physical* category is followed by *Meals and breaks* (61.9%) and *Travel* (55.6%). The rest of the categories only involve physical activities a third of the times or less. If we compare the low proportion of physical activity with the high proportions we have seen in the cyber dimension, we can envisage opportunities to provide support with intelligent assistants.

Social. The social aspect of the self-reported tasks was asked in the question "*Did you engage in any social interactions in order to complete the task?*". If respondents answered affirmatively, they were also asked what kind of social interactions they engaged in.

Figure 4 (right) shows the percentage of tasks that involve social interactions, for each category. Overall, around 40% of the tasks reported involve social interactions. *Customer care* tasks involve social interactions 75% of the time, substantially more than the others. Around half of the tasks categorized as *Project*, *Problem solving*, *Planning*, *Communication*, and *Meals and breaks* involve social interaction. The categories that are less likely to involve social interactions are *Low-level* and *Finance* (both around 19%).

4.6 Supporting Information Workers' Tasks

Currently used tools. Responses for Question 19 show that the most common tool is a digital calendar, used by nearly half of our respondents; this is followed in popularity by non-digital tools such as paper lists, post-it notes, and a paper calendar, which together correspond to 48% of the tools reported by the participants. Other digital tools that can assist in task planning, completion and tracking—such as digital lists, alarm clocks or productivity platforms—were less commonly used.

Opportunities for intelligent assistants. We also asked participants to describe the features or capabilities that they would want

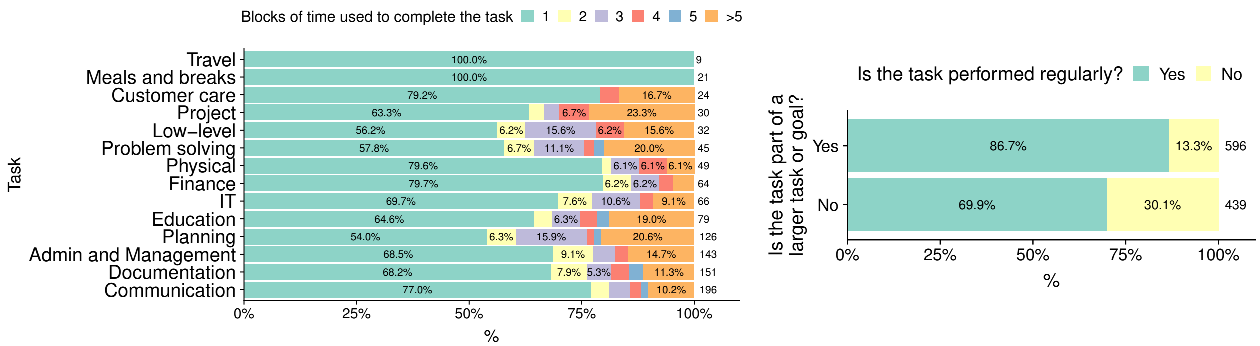

Figure 2: Task continuity (left) and task regularity (right).

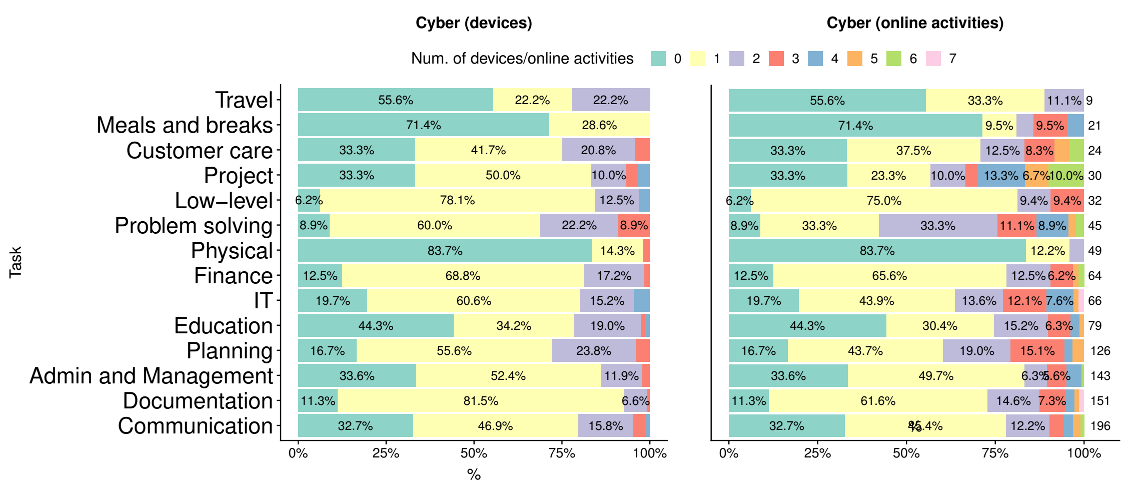

Figure 3: Devices and online activities (cyber). Respondents could indicate both multiple devices and online activities used to complete a task.

to have in a hypothetical new piece of technology. Figure 5 shows the features reported.

Many participants (46.9%) do not report a particular feature (or report they do not need any). Most wanted features related to more effective task management: reminders, calendars or task scheduling. Other features are related to: assisting with the organization of information (automatic sorting, listing, etc.); the need for technology that allows access to information from multiple devices and synchronizes the content in a common place (*ubiquitous access*); and tracking of task progression and completion. The long tail of features includes: voice-enabled interaction; capabilities to adapt to the context of the user; assistance with collaboration; taking notes; searching while the user is performing a task.

5 DISCUSSION AND IMPLICATIONS

The analysis of the survey responses shows that information workers perform a broad range of tasks. A substantial number of tasks take more than two hours to complete. We found that tasks categorized as *Planning*, *Problem solving*, and *Low-level* are more likely to be interrupted. We also found that most of the tasks are performed regularly. Our analysis of CPS aspects showed that multiple devices are used simultaneously to complete tasks to perform online activities such as Web browsing, search and, e-mailing. Tasks that involve physical activities typically involve less device usage. In terms of the social aspect, 40% of the tasks involve social interactions. We also found that, although non-digital tools such as paper lists, post-it notes, and paper calendars are heavily used, participants desired features for more effective task management.

The analysis allows us to better understand how people perform tasks in work settings, and reveals that there are several ways in which intelligent assistants can support information workers:

Task identification. The characteristics of tasks observed in our analysis can be used to automatically identify the current task; user activity can also help [35]. For instance, if the user is performing simultaneous online activities using multiple devices, without social interactions, and for a long period of time, it is likely that the user is performing a *Planning* or *Problem solving* task. Intelligent assistants may engage with users (e.g. asking questions via voice [36]) to help manage uncertainty regarding any task inferences.

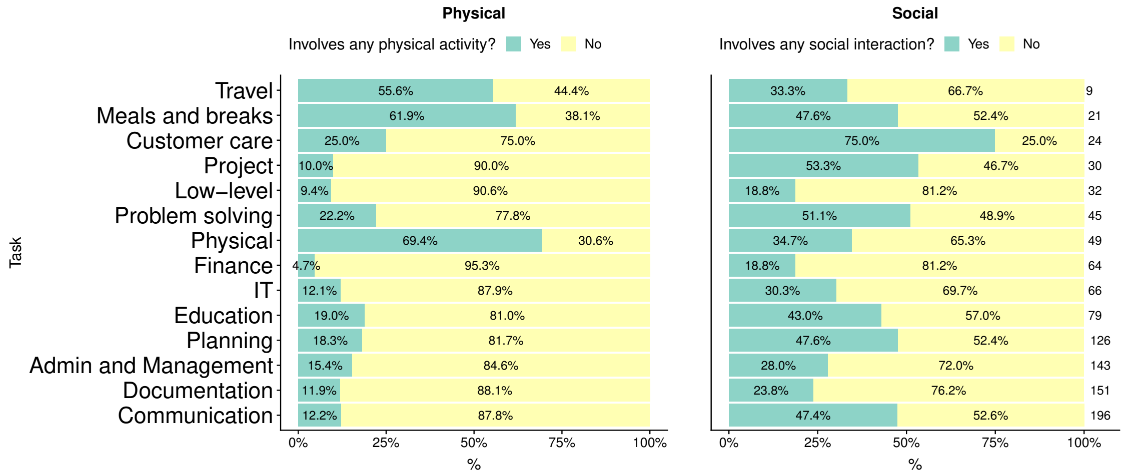

Figure 4: Physical and social activities.

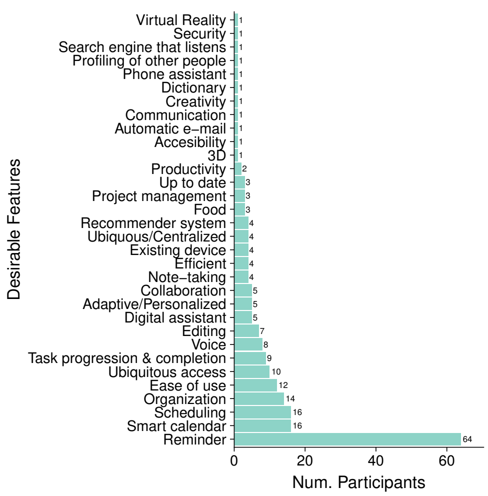

Figure 5: Desirable features for an imaginary new piece of technology as reported by participants.

Task progression and completion. Similarly, task characteristics such as duration or interruptions can be used to track progress and completion of tasks. We found that the majority of tasks are performed regularly and (as Figure 5 shows) reminders were the most requested feature from respondents. It should be feasible to apply data mining and machine learning methods to identify patterns in the behaviours pertaining to task activities and their progression over time, as shown in other domains, e.g. medicine [27].

Task recommendation and automatic scheduling. CPS signals may help generate contextual recommendations about when and where users can complete tasks. Support for organization, scheduling, and calendering were also among the most requested features (Figure 5). The assistant may find time (from duration estimates) to complete pending tasks, set agendas (from time, TO-DOs, and task dependencies), or divide tasks into microtasks [7].

Task resumption. Our analysis showed that many tasks spanned blocks of time and multiple devices. This increases the need for tools to support task resumption [15, 25]. Keeping track of interruptions may allow an intelligent assistant to proactively indicate that it is timely to resume a specific task (e.g. continuing the task on a different device or in a different context).

Multi-step tasks. Given that tasks are likely to be part of a larger task or goal (58% of tasks per Figure 2), an intelligent assistant may support multi-step tasks by recommending content, resources, or actions relevant to the current step with knowledge of already-completed steps (and where possible, what steps are still to come).

The analysis presented in this paper has some limitations. Although a broad number of tasks have been identified, our analysis is based on a survey of a specific cohort (currently employed, older than 25 years old, and from a set of English-speaking countries). There are aspects of the responses that have also not been analyzed herein (e.g. locations or task difficulty) and will be in future work.

6 CONCLUSIONS AND FUTURE WORK

This work aimed to better understand the completion of tasks in work settings. We analyzed the responses of around 400 participants to a survey that capture the range of tasks and sub-tasks people perform. Our thematic and quantitative analysis derived a characterization of work tasks. An investigation of task duration, continuity, and regularity, as well as a detailed analysis of cyber-physical-social aspects, allowed us to provide design recommendations for how intelligent assistants could support work tasks, both on-demand and proactively. Future plans include running user studies to collect cyber-physical-social aspects from mobile devices in-situ during information work [23]. Sensors in these devices will record human activities and routines, to be associated with task progression. Consequently, we will explore how these data could be used for task identification, tracking, and recommendation.

ACKNOWLEDGMENTS

This research was partially supported by Microsoft Research. Any opinions, findings and conclusions expressed in this paper are those of the authors and do not necessarily reflect those of the sponsor.

REFERENCES

[1] Eugene Agichtein, Ryen W. White, Susan T. Dumais, and Paul N. Bennett. 2012. Search, Interrupted: Understanding and Predicting Search Task Continuation. In *Proc. SIGIR*. ACM, 315–324.

[2] Australian Bureau of Statistics. 2013. Directories, Classifications and Manuals by Topic. http://www.abs.gov.au/ausstats/abs@.nsf/ViewContent?readform&view=DirClassManualsbyTopic&Action=Expand&Num=6.1.3 Accessed on 15-01-2019.

[3] Peter Bailey, Liwei Chen, Scott Grosenick, Li Jiang, Yan Li, Paul Reinholdtsen, Charles Salada, Haidong Wang, and Sandy Wong. 2012. User Task Understanding: A Web Search Engine Perspective. In *NII Shonan Meeting on Whole-Session Evaluation of Interactive Information Retrieval Systems*.

[4] Victoria Bellotti, Nicolas Ducheneaut, Mark Howard, and Ian Smith. 2003. Taking Email to Task: The Design and Evaluation of a Task Management Centered Email Tool. In *Proc. CHI*. ACM, 345–352.

[5] Andrei Broder. 2002. A Taxonomy of Web Search. *ACM SIGIR Forum* 36, 2 (2002), 3–10.

[6] Jacques Charmes. 2015. Time Use across the World: Findings of a World Compilation of Time-Use Surveys. *Background paper for Human Development Report* (2015).

[7] Justin Cheng, Jaime Teevan, Shamsi T. Iqbal, and Michael S. Bernstein. 2015. Break It Down: A Comparison of Macro- and Microtasks. In *Proc. CHI*. ACM, 4061–4064.

[8] Karen Church, Mauro Cherubini, and Nuria Oliver. 2014. A Large-Scale Study of Daily Information Needs Captured In Situ. *ACM Trans. on Computer-Human Interaction* 21, 2, Article 10 (Feb. 2014), 46 pages.

[9] Karen Church and Barry Smyth. 2009. Understanding the Intent Behind Mobile Information Needs. In *Proc. IUI*. ACM, 247–256.

[10] Victoria Clarke and Virginia Braun. 2006. Using Thematic Analysis in Psychology. *Qual. Res. Psychol.* 3, 2 (2006), 77–101.

[11] Victoria Clarke and Virginia Braun. 2013. *Successful Qualitative Research: A Practical Guide for Beginners*. Sage.

[12] Mark Dredze and Hanna Wallach. 2008. User Models for Email Activity Management. In *IUI Workshop on Ubiquitous User Modeling*.

[13] Cathal Gurrin, Hideo Joho, Frank Hopfgartner, Liting Zhou, Rashmi Gupta, Rami Albatal, Dang Nguyen, and Duc Tien. 2017. Overview of NTCIR-13 Lifelog-2 Task. In *Proc. NTCIR*.

[14] Ahmed Hassan Awadallah, Ryen W. White, Patrick Pantel, Susan T. Dumais, and Yi-Min Wang. 2014. Supporting Complex Search Tasks. In *Proc. CIKM*. ACM, 829–838.

[15] Shamsi T. Iqbal and Eric Horvitz. 2007. Disruption and Recovery of Computing Tasks: Field study, Analysis, and Directions. In *Proc. CHI*. ACM, 677–686.

[16] Rosie Jones and Kristina Lisa Klinkner. 2008. Beyond the Session Timeout: Automatic Hierarchical Segmentation of Search Topics in Query Logs. In *Proc. CIKM*. ACM, 699–708.

[17] Daniel Kahneman, Alan B. Krueger, David A. Schkade, Norbert Schwarz, and Arthur A. Stone. 2004. A Survey Method for Characterizing Daily Life Experience: The Day Reconstruction Method. *Science* 306, 5702 (2004), 1776–1780.

[18] Diane Kelly and Nicholas J. Belkin. 2004. Display Time as Implicit Feedback: Understanding Task Effects. In *Proc. SIGIR*. ACM, 377–384.

[19] Bret Kinsella and Ava Mutchler. 2018. *Smart Speaker Consumer Adoption Report*. https://voicebot.ai/wp-content/uploads/2018/03/smart_speaker_consumer_adoption_report_2018.pdf Accessed on 15-01-2019.

[20] Alexander Kotov, Paul N. Bennett, Ryen W. White, Susan T. Dumais, and Jaime Teevan. 2011. Modeling and Analysis of Cross-Session Search Tasks. In *Proc. SIGIR*. ACM, 5–14.

[21] Nicholas Kushmerick and Tessa Lau. 2005. Automated email activity management: an unsupervised learning approach. In *Proc. IUI*. ACM, 67–74.

[22] Yuelin Li and Nicholas J. Belkin. 2008. A Faceted Approach to Conceptualizing Tasks in Information Seeking. *Inf. Proc. & Man.* 44, 6 (2008), 1822–1837.

[23] Jonathan Liono, Johanne R. Trippas, Damiano Spina, Mohammad S. Rahaman, Yongli Ren, Flora D. Salim, Mark Sanderson, Falk Scholer, and Ryen W. White. 2019. Building a Benchmark for Task Progress in Digital Assistants. In *Proc. WSDM'19 Task Intelligence Workshop (TI@WSDM19)*.

[24] Claudio Lucchese, Salvatore Orlando, Raffaele Perego, Fabrizio Silvestri, and Gabriele Tolomei. 2011. Identifying Task-Based Sessions in Search Engine Query Logs. In *Proc. WSDM*. ACM, 277–286.

[25] Gloria Mark, Victor M. Gonzalez, and Justin Harris. 2005. No Task Left Behind?: Examining the Nature of Fragmented Work. In *Proc. CHI*. ACM, 321–330.

[26] Rishabh Mehrotra, Ahmed Hassan Awadallah, Ahmed E. Kholy, and Imed Zitouni. 2017. Hey Cortana! Exploring the Use Cases of a Desktop Based Digital Assistant. In *Proc. 1st International Workshop on Conversational Approaches to Information Retrieval (CAIR'17)*.

[27] Michael J. Paul, Ryen W. White, and Eric Horvitz. 2016. Search and Breast Cancer: On Episodic Shifts of Attention over Life Histories of an Illness. *ACM Transactions on the Web (TWEB)* 10, 2 (2016), 13.

[28] Ashequl Qadir, Michael Gamon, Patrick Pantel, and Ahmed Hassan Awadallah. 2016. Activity Modeling in Email. In *Proc. NAACL*. 1452–1462.

[29] Yongli Ren, Martin Tomko, Flora D. Salim, Jeffrey Chan, Charles L.A. Clarke, and Mark Sanderson. 2018. A Location-Query-Browse Graph for Contextual Recommendation. *IEEE Trans. Know. Data Eng.* 30, 2 (Feb 2018), 204–218.

[30] Yongli Ren, Martin Tomko, Flora D. Salim, Kevin Ong, and Mark Sanderson. 2017. Analyzing Web Behavior in Indoor Retail Spaces. *JASIST* 68, 1 (2017), 62–76.

[31] Daniel E. Rose and Danny Levinson. 2004. Understanding User Goals in Web Search. In *Proc. WWW*. ACM, 13–19.

[32] Miamaria Saastamoinen and Kalervo Järvelin. 2017. Search Task Features in Work Tasks of Varying Types and Complexity. *JASIST* 68, 5 (2017), 1111–1123.

[33] Jianqiang Shen, Lida Li, Thomas G. Dietterich, and Jonathan L. Herlocker. 2006. A Hybrid Learning System for Recognizing User Tasks from Desktop Activities and Email Messages. In *Proc. IUI*. ACM, 86–92.

[34] Timothy Sohn, Kevin A. Li, William G. Griswold, and James D. Hollan. 2008. A Diary Study of Mobile Information Needs. In *Proc. CHI*. ACM, 433–442.

[35] Simone Stumpf, Xinlong Bao, Anton Dragunov, Thomas G. Dietterich, Jon Herlocker, Kevin Johnsrude, Lida Li, and JianQiang Shen. 2005. Predicting User Tasks: I Know What You're Doing. In *AAAI'05 Workshop on Human Comprehensible Machine Learning*.

[36] Johanne R. Trippas, Damiano Spina, Lawrence Cavedon, Hideo Joho, and Mark Sanderson. 2018. Informing the Design of Spoken Conversational Search. In *Proc. CHIIR*. ACM, 32–41.

[37] Peter Tsai. 2018. *AI Chatbots and Intelligent Assistants in the Workplace*. https://community.spiceworks.com/blog/2964-data-snapshot-ai-chatbots_and-intelligent-assistants-in-the-workplace Accessed on 15-01-2019.

[38] United Nations. Statistical Division. 2007. *Guide to Producing Statistics on Time Use: Measuring Paid and Unpaid Work*. United Nations Publications.

[39] Ryen W. White and Diane Kelly. 2006. A study on the effects of personalization and task information on implicit feedback performance. In *Proc. CIKM*. ACM, 297–306.

[40] Jessica L. Wildman, Amanda L. Thayer, Michael A. Rosen, Eduardo Salas, John E. Mathieu, and Sara R. Rayne. 2012. Task Types and Team-Level Attributes: Synthesis of Team Classification Literature. *Human Resource Development Review* 11, 1 (2012), 97–129.

A QUESTIONNAIRE

(1) List a task that you carried out in a work setting and
   (a) took less than 30 minutes.
   (b) took longer than 30 minutes.
   (c) longer than 2 hours.
   (d) more than a day.
(2) Where did you carry out the task?
(3) When did you carry out the task?
(4) Was the task completed in one continuous block of time? (how many?)
(5) What tools, mechanisms or processes did you use to remind yourself to continue the task?
(6) In which context did you compete the task? (Time-used categories)
(7) Did you use electronic devices in order to complete the task? (which?)
(8) Did you do any of the following online activities in order to complete the task? (web browsing, web searching, performing online transactions, using a specific app, online chatting, emailing, using a digital assistant, other)
(9) Did you engage in any physical activities? (which one?)
(10) Did you engage in any social interactions in order to complete the task? (which one?)
(11) I wanted to do the task.
(12) The task was easy.
(13) Are there triggers that lead you to work on the task? (Which?)
(14) Is it a regular task?
(15) Is the task part of a larger goal? (which?)
(16) List the steps to complete the task
(17) Was there anything you did before initiating the task?
(18) Was there anything you did after initiating the task?
(19) Which tools do you currently use to help you plan, complete or keep track of tasks?
(20) Imagine there is a new piece of technology, what features or capabilities would you want it to have?