Cloudy with a Chance of Misconceptions: Exploring Users' Perceptions and Expectations of Security and Privacy in Cloud Office Suites

Dominik Wermke, Nicolas Huaman, Christian Stransky, Niklas Busch, Yasemin Acar, and Sascha Fahl, *Leibniz University Hannover*

https://www.usenix.org/conference/soups2020/presentation/wermke

This paper is included in the Proceedings of the Sixteenth Symposium on Usable Privacy and Security.

August 10–11, 2020

978-1-939133-16-8

Open access to the Proceedings of the Sixteenth Symposium on Usable Privacy and Security is sponsored by USENIX.

Cloudy with a Chance of Misconceptions: Exploring Users' Perceptions and Expectations of Security and Privacy in Cloud Office Suites

Dominik Wermke
Leibniz University Hannover

Nicolas Huaman
Leibniz University Hannover

Christian Stransky
Leibniz University Hannover

Niklas Busch
Leibniz University Hannover

Yasemin Acar
Leibniz University Hannover

Sascha Fahl
Leibniz University Hannover

Abstract

Cloud Office suites such as Google Docs or Microsoft Office 365 are widely used and introduce security and privacy risks to documents and sensitive user information. Users may not know how, where and by whom their documents are accessible and stored, and it is currently unclear how they understand and mitigate risks. We conduct surveys with 200 cloud office users from the U.S. and Germany to investigate their experiences and behaviours with cloud office suites. We explore their security and privacy perceptions and expectations, as well as their intuitions for how cloud office suites should ideally handle security and privacy. We find that our participants seem to be aware of basic general security implications, storage models, and access by others, although some of their threat models seem underdeveloped, often due to lacking technical knowledge. Our participants have strong opinions on how comfortable they are with the access of certain parties, but are somewhat unsure about who actually has access to their documents. Based on our findings, we distill recommendations for different groups associated with cloud office suites, which can help inform future standards, regulations, implementations, and configuration options.

1 Introduction

During the 1970s, office software began to emerge in the world of personal computing. Early word processors such as Electric Pencil for the MITS Altair in 1976, WordStar for the CP/M in 1978, and later dedicated spreadsheet applications such as VisCalc were considered "*killer applications*" for their respective systems. These dedicated office tools helped the adoption of personal computers over more dedicated or mechanical systems for word processing. In recent years, another major shift is happening in the world of office applications. With Microsoft Office 365, Google Drive, and projects like LibreOffice Online, most major office suites have moved to provide some sort of cloud platform that allows for collaboration between multiple editors, automatic real-time storage on cloud or internal network servers, and easy access through the browser without requiring the installation of software.

The major selling points for these cloud office platforms might as well be their most concerning (security & privacy) weaknesses: easy sharing of documents, cloud storage of data, and the high similarity in design and UI to previously prevalent offline office software hide a large array of potential privacy and security trapdoors from the average office user.

With the shift from offline to cloud, many cloud office providers also moved from a pay-once model to a subscription-based model with a trial period or even a model with completely free usage. This shift accompanied a questionable change in business model drive towards data collection and profiling: the processing and storing of documents in the cloud provides the possibility of large-scale privacy intrusion by the providers for both end users and businesses that utilize the cloud. Due to the similarity in design to offline office software, end users are unlikely to fully comprehend this major impact on their privacy. This impact on privacy is further amplified by governments and administrations updating their infrastructure to cloud-based solutions, potentially processing and uploading the data of citizens in the cloud without their explicit consent. In a recent example, the Department of Defense awarded a $7.6 billion contract to General Dynamics to provide the Pentagon with the cloud-based Microsoft Office 365 [27].

Another major selling point of cloud office applications is the ease of access, often from almost anywhere on earth with an internet connection, without requiring any additional installation of software. While the actual location of the underlying servers is rarely mentioned in cloud advertisements, it has

USENIX Symposium on Usable Privacy and Security (SOUPS) 2020.
August 9–11, 2020, Virtual Conference.

notable implications for privacy and security. In July 2019, the central German State of Hesse declared that schools may not legally use Microsoft Office 365 or similar cloud office platforms due to collected telemetry and the potential access to stored data on U.S. servers by U.S. officials [34, 36]:

> "*What is true for Microsoft is also true for the Google and Apple cloud solutions. The cloud solutions of these providers have so far not been transparent and comprehensibly set out. Therefore, it is also true that for schools the privacy-compliant use is currently not possible.*" - Hessian commissioner of Data Protection and Freedom of Information [34].

In August 2019, Microsoft announced that it will be able to provide cloud services from data centers in Germany in late 2019 "*to meet evolving customer needs*" and to being "*committed to making sure that the Microsoft Cloud complies with [the European General Data Protection Regulation] GDPR*" [9]. As of February 2020, Microsoft offers Office 365 and Dynamics 365 from new German data center regions [21].

In this work, we investigate privacy and security misconceptions by end users of cloud office applications in a user study including participants from both the U.S. and Germany. For this, we conducted two online surveys with 200 crowd workers from Amazon's Mechanical Turk and ClickWorker. With a combination of qualitative and quantitative methods, we modeled the two surveys to explore the following research questions:

RQ1: "*How and why do our participants interact with cloud office applications?*" Cloud office suites are compelling to use with features such as collaboration between multiple editors, automatic real-time storage, and easy online access without installation. We are interested why and how our participants interact with cloud office applications both in a home and organizational setting.

RQ2: "*What are end users' awareness, perceptions, and attitudes about privacy in cloud office applications?*" The switch from offline to a cloud environment in both home and organizational settings introduced abrupt changes for privacy and security assumptions regarding office suites. We examine our participants' security and privacy perceptions and expectations, as well as their intuitions for how cloud office suites should ideally handle security and privacy.

RQ3: "*What are participants' understandings and related mental models regarding protection and security of their cloud documents?*" The actual server location, access by providers or governments, and handling of deletions has an enormous impact on the privacy of cloud office applications. We survey the extent of our participants' understanding and their basic mental models regarding cloud office documents.

The remainder of this paper is structured as follows: after this introduction (Section 1) we provide a background to cloud office suites in Section 2. We describe the setup and structure of our two surveys in Section 3 and report our results in Section 4. We discuss related work in Section 5. Finally, we discuss findings and give recommendations in Section 6 and conclude this work in Section 7.

2 Cloud Office Suites

For this work, we define Cloud office suites as cloud-based office applications that allow view, edit and comment on documents, spreadsheets and presentations in the browser.

Table 1 provides an overview of the most popular cloud office suites and their features relevant for this work. Prominent providers of cloud office suites are Google (Google Drive) [12], Microsoft (Office 365) [25], Apple (iWork for iCloud) [15], The Document Foundation (Libre Office Online) [11], and Ascensio System SIA (OnlyOffice) [37]. In contrast to traditional office suites such as Microsoft Office, cloud office suites provide browser-based user interfaces. Users are no longer limited to work on desktop computers using native office applications, but can access their files using any device that provides a modern browser. Hence, modern cloud office suites support mobile devices such as smartphones and tablets and allow easy access to their cloud applications wherever users have access to the internet.

In contrast to traditional office suites, cloud office suites allow users to easily share documents with multiple collaborators and edit the same document simultaneously. Cloud office documents can be shared using e-mail addresses or direct links to a document. For better user experience, cloud office suites allow their users to recover deleted documents. In addition to online access to their documents, Google Drive provides an offline mode that stores documents in the local browser storage and makes them available for offline editing. Offline documents are pushed to the cloud as soon as users have Internet access.

The three major providers Microsoft, Google, and Apple only provide cloud-hosted solutions while Ascensio System also provides a self-hosted community edition which allows keeping the data under users' control. Every hosted cloud office solution provides storage capabilities in the cloud. The amount of storage included depends on the license purchased and can be upgraded at any time. LibreOffice Online by The Document Foundation supports no storage by itself and is dependent on the underlying software like OwnCloud or NextCloud to provide the storage and authentication.

While all cloud office suites provide rudimentary access control for sharing, only Google Drive and OnlyOffice provide an option to share documents with read-only access that still allows to comment on documents.

	Storage	Offline Mode	Versions available				Mobile Version		Sharing				Document Recovery
			Self Hosted	Free	Paid	Trial	Android	iOS	E-Mail	Link	Read Only	Read & Comment	
Office 365	●	○	○	◐[1]	●	●	●	●	●	●	●	○	●
Google Drive	●	●	○	●	●	●	●	●	●	●	●	●	●
iWork for iCloud	●	○	○	●	●	-	○	●	●	●	●	○	●
LibreOffice Online	○	○	●	●	◐[2]	-	◐[3]	○	◐[4]	◐[4]	◐[4]	○	◐[4]
OnlyOffice	●	○	●	●	●	●	●	●	●	●	●	●	●

● Feature available ○ Feature unavailable ◐ Feature partially available
[1] Students and teachers receive a free online only version. [2] Support is only provided by third party companies and not directly by The Document Foundation. [3] Only a viewer is available. [4] Depends on underlying software.

Table 1: Overview of the most common cloud office suites and their related features.

3 Methodology

In this section we provide details on the procedure and structure of the two surveys we conducted with crowd workers from Amazon's Mechanical Turk ($n = 105$) and ClickWorker ($n = 95$). We also detail the coding process for our qualitative questions as well as the statistical analysis approach for our quantitative data. Finally, we report on our data collections and ethical considerations, and discuss the limitations of our work.

Note that while our two surveys may include participants living in the U.S. or in Germany, Austria, or Switzerland respectively, we refer to them as "U.S." and "German(y)" for a more succinct reporting.

3.1 Study Procedure

Both the German-speaking participants from ClickWorker and the English-speaking participants from Mechanical Turk were administered an almost identical survey, with the German survey being a direct translation from the English version by multiple native German speakers.

Questionnaire Development. The questionnaire development was guided by our established research questions. We included pre-tested and evaluated survey questions from previous work where appropriate to allow for a greater comparability between studies. In addition, we performed 5 in-depth, free-form interviews with both experts and non-experts to establish additional areas of interest for our survey.

Pre-Testing. Before we conducted the surveys, we pre-tested our questionnaires following the principle of cognitive interviews [31]. This allowed us to glean insights into how survey respondents might interpret and answer questions. We asked participants to share their thoughts as they answered each survey question and used our findings to iteratively revise and rewrite the survey questions to minimize bias and maximize validity. This first pre-test was conducted internally in both German and English with members of the groups, students of our university, and friends. In addition, we refined the surveys in multiple pilots with participants on Mechanical Turk ($n = 9$) and ClickWorker ($n = 20$) until a satisfactory convergence was reached.

Recruitment and Inclusion Criteria. We recruited participants for our study from Amazon's Mechanical Turk and ClickWorker during September 2019. We did not mention security or privacy in the initial recruitment ad to avoid certain recruitment biases. We generally required participants to be age 18 or older and to have used cloud office software before. For Amazon's Mechanical Turk, we additionally required participants to be comfortable with participating in the study in English and to live within the United States. To ensure sufficient data quality, we also required them to have completed a minimum of 1,000 hits and to have a task approval rate of at least 95% [30]. For ClickWorker, we additionally required participants to be comfortable with participating in the survey in German and to live within Germany, Austria, or Switzerland.

A total of 229 people responded to our surveys. Of those, 22 did not finish and 7 were excluded due to low-quality answers or due to failing at least one of our quality checks, resulting in 200 final participants whose responses we consider.

3.2 Survey Structure

We outline the survey structure in Figure 1 and below. Both our surveys consisted of a total of 9 sections, ranging from general cloud office questions to personal beliefs about the responsibilities of cloud office providers. The two survey versions differed slightly due to localized answer options (e.g., localized names for government agencies) and changes to concepts that do not exist or have a different privacy implication in German-speaking countries (e.g., social security number).

1. Use of Office Tools: Our surveys open with questions in which we explore the general usage patterns of offline and cloud office applications by our participants in both private and organizational contexts. We report general demograph-

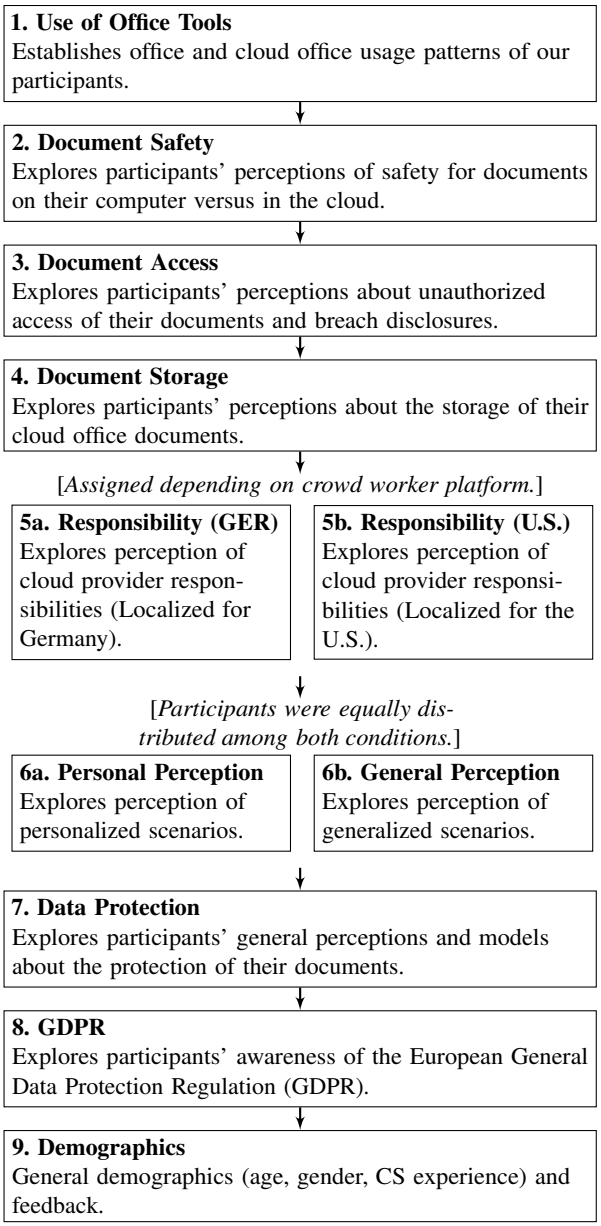


Figure 1: Illustration of the survey flow for both German and English surveys. Splits in the flow include a localized version of the "Responsibility" block for Germany and the U.S. and a split for generalized scenarios vs. personalized which were randomly assigned to participants.

ics and office-specific demographics of our participants in Section 4.1 and Table 3.

2. Document Safety: The "Document Safety" section explores how participants perceive the security of their documents in the cloud vs. locally on their computer and why. We report these results in Section 4.2.

3. Document Access: The "Document Access" section investigates participants' mental concepts and perceived risks related to the access of their documents. Questions related to which parties they think have access to their documents, who already might have accessed their documents without their authorization, and if the risk of unauthorized access by different parties is higher in the cloud or on their computer. Further, the section asks participants about who they think *would* inform them in case of an unauthorized access to their data and who they think *should* inform them and how. We report the results related to the access of cloud office documents in Section 4.3.

4. Document Storage: This section explores our participants' perception about the storage of their cloud office documents. We asked our participants about the number of copies they think exist of their documents and with whom they think copies remain after deleting their own versions. In addition, we asked who they think can delete their documents. We report the results for these questions in Section 4.4.

5a/b. Responsibility: In this section, we investigate our participants' perceptions about responsibilities of cloud office providers regarding access and protection of documents. The "Responsibility" section differs slightly between the German and English survey to allow for the localization of certain answer options such as law-enforcement agencies and government names. We report the results in Section 4.5.

6a/b. Perception: The "Perception" section contains questions to three different scenarios related to the processing of sensitive data in cloud document applications, either in a more personal or more generalized condition.

1. Data of children. The first scenario described the use of a cloud office application in an educational setting. We asked our participants to assess how much they felt at ease with using cloud office applications for handling data of children in schools, e.g., for storing grades or writing tasks.
2. Health data. The second scenario had a focus on health information. A general practitioner used a cloud office application to handle sensitive patient information including a patient's name, age, weight, diagnosis, and treatment plan. Again, we asked our participants to rate their level of comfort with the scenario.
3. Financial data. In the third scenario we illustrated a use case involving financial data. A financial advisor used a cloud office application to process client data. The processed documents include private information such as the client's name, social security number, and detailed financial information.

Participants of the study were equally distributed between both conditions and the order of scenarios was randomized for each participant. Results for the different scenarios are reported in Section 4.6.

7. Data Protection: The "Data Protection" section explores participants' mental models about the protection of their documents in the cloud. We asked our participants which data

they think is collected when they process documents in cloud office applications and how they think their data is protected. We report these results in Section 4.7.

8. General Data Protection Regulation: In the "GDPR" section we explored our participants' general knowledge about the European General Data Protection Regulation (GDPR) and what they know about the protections it offers. These questions link back to the "Responsibility" block, which asked participants about cloud office provider responsibilities directly implied by the GDPR. We report the general results for this block together with other demographics in Section 4.1 and combined it for our analysis of the responsibility section in Section 4.5.

9. Demographics: We administered demographic questions at the end of the questionnaire to prevent stereotype bias [22, 35]. Our demographic questions included age, gender (with free text), and previous experiences in CS education and CS jobs. Additionally we asked respondents for general feedback for the survey questionnaire. We report general demographics and office-specific demographics of our participants in Section 4.1 and Table 3.

3.3 Coding and Analysis

Our collected data includes both qualitative and quantitative data points.

Qualitative Coding. We analyzed all open-ended questions in an iterative open-coding process [7, 38]. Two researchers established an initial codebook [5], coded all open-ended questions together, and resolved emerging coding conflicts immediately in a consensus discussion or by introducing new codes. If new codes were introduced, all previous answers were revisited and re-coded if necessary. Due to the immediate resolving, reporting an intercoder agreement and reliability is uncommon for this approach [20]. The codebook remained stable once both researchers were satisfied that all important themes and concepts in the responses could be captured with the codes. Both surveys were coded with the same codebook and codes for the German survey were assigned by two native speakers.

Quantitative Analysis. We use the non-parametric Kruskal-Wallis H test (KW; non-parametric equivalent to the one-way ANOVA) to compare multiple independent groups. For multiple tests on paired groups, we use the Mann-Whitney U test (MWU) and control the results for multiple testing. We assume an alpha level of $\alpha = .05$ for significance in hypothesis tests. Where appropriate, we controlled our hypothesis tests for the multiple comparison problem with the conservative Bonferroni correction and report the "adjusted"/"adj." values. For certain tests, we map five-point Likert scale answers to numbers (-2, -1, 0, 1, 2).

We present the outcomes of our regressions in tables where each row contains a factor and the corresponding change of the analyzed outcome in relation to the baseline of the given factor. Linear regression models measure change from baseline factors with a coefficient (Coef.) of zero for the value of the outcome. For each factor of a model, we also list a 95% confidence interval (C.I.) and a p-value indicating statistical significance. We highlight p-values below a cut-off of .05 with a star ($*$).

As our regression analyses are intended to be exploratory, we consider a set of candidate models and select the final model based on the lowest Akaike Information Criterion (AIC) [4]. We consider candidate models consisting of the required factors "Country", "Condition", and "Scenario", as well as every possible combination of the optional variables. Required factors, optional factors, and corresponding baseline values are described in Table 2. In cases when we consider results on a per-scenario rather than a per-participant basis, we use a mixed linear model that adds a random intercept to account for multiple scenarios from the same participant.

3.4 Data Collection and Ethics

Our institutions did not require a formal IRB process for the studies conducted in this work. Nonetheless, we modeled our research plan and study procedures after an IRB-approved study, adhered to the strict German and U.S. data and privacy protection laws and the General Data Protection Regulation in the E.U., and structured our study following the ethical principals of the Menlo report for research involving information and communications technologies [10]. All participants approved to a consent form that informed them about the study purpose, the data we collected and stored, and included an e-mail address and phone number to contact the principal investigators in case of questions or concerns.

Recently, researchers faced issues with low data quality on Amazon MTurk [18]. Therefore, we included a number of filters to identify low-quality answers. During data cleaning and analysis, we identified 7 participants who did not pass our quality measures and excluded these invalid participants from further analysis.

We calibrated participants' compensations based on an average piloting time of 10 minutes and payed participants on Amazon's Mechanical Turk $1.70 and on ClickWorker €1.70 for an hourly wage of $10.20 and €10.20, respectively.

3.5 Limitations

As any study with online surveys, our work includes a number of limitations typical for this type of study and should be interpreted in context. In general, self-report studies may suffer from several biases, including over- and under-reporting, sample bias, and social-desirability bias. However, while we utilize self-report data, our central claims are not about the accuracy of respondents' answers to a given question, but

Factor	Description	Baseline
Required		
Country	Germany or U.S., participants assigned based on crowd working platform.	U.S.
Condition	General or Personal. Scenario condition, participants evenly distributed between both conditions.	General
Scenario	Child, Health, or Financial. Type of scenario, all 3 shown in randomized order.	Child
Participant	Random effect accounting for repeated measures (due to the 3 scenarios per participant).	n/a
Optional		
Office at work	True or False, uses office software at work, self-reported.	False
CS Education	True or False, has a CS education, self-reported.	False
CS Job	True or False, has a CS job, self-reported.	False
Age	Age in years, self-reported.	n/a

Table 2: Factors used in candidate regression models. Model candidates always included the required factors and covered all possible combinations of optional factors. Final models were selected based on lowest AIC. Categorical factors were individually compared to the baseline.

rather about the concepts and misconceptions conveyed by their answers.

Conducting user studies on crowd working platforms like Amazon's Mechanical Turk and ClickWorker is a commonly used and generally accepted procedure for human-computer interaction and usable security and privacy research [33]. While the quality of answers can suffer in a crowd worker context, we tried to ensure a high data quality by following best practices by limiting access to our surveys to high-reputation cloud workers [30] and by manually filtering low quality answers.

This study focuses on the responses of German and U.S. Internet users, and thus, we can offer no insight into the generalizability of results for international participants. We aimed to improve the internal validity of our study by providing localized answer options.

We explicitly ignored the implications of meta data collection and third party data of cloud office providers to allow participants to focus on their mental model of cloud document processing and access.

4 Results

In the following section we report and discuss results for all 200 valid participants of both the U.S. and German survey. Generally, participants were aware of certain security and privacy implications of writing their documents in cloud office applications, but were unaware or had severe misconceptions about others. Our reporting of results mostly follows the actual order of survey sections described in Section 3.2. After each subsection, we summarize our key findings.

4.1 Use of Office Tools

We report the general demographics of both surveys in Table 3. Overall, 127 participants responded to our survey on Amazon's Mechanical Turk (U.S.) and 102 on ClickWorker (German). Of those, 105 and 95 respectively completed the survey and were considered valid for a combined total of 200 participants for whom we report results.

Our participants identified predominantly as male (64.5%) with a median age of 33.0 years (mean $= 35.7$, $\sigma = 10.7$). Across both surveys, 28.0% of our participants classified themselves as having a CS education and 22.5% as having worked in a CS-related job. CS experiences are similar for both the U.S. and the German survey, with the exception of CS education (38.1% vs. 16.8%). We assume this discrepancy might be related to general differences in education systems, as the German school curriculum focuses less on IT education compared to the U.S. The majority of both our U.S. and German participants have a job that involves using office applications regularly with 80.0% and 77.8%, respectively.

The majority (97.1%) of our U.S. participants have used Google Drive (with its related cloud office tools such as Google Docs or Google Sheets) before, followed by Microsoft Office (Offline) (86.7%) and Microsoft Office 365 (Cloud) (70.5%). The majority of our German participants (87.4%) is more familiar with Microsoft Office (Offline), followed by Google Drive (80.0%) and Microsoft Office 365 (Cloud) (64.2%). We assume this difference is likely due to the extensive, almost exclusive usage of Microsoft Office products in German businesses and government[1]. These differences even out for office tools used in the last months where Google Drive prevails among both the U.S. and German participants (82.7%, 70.5%), followed by Microsoft Office (Offline) (50.5%, 65.3%) and Microsoft Office 365 (Cloud) (50.5%, 55.8%). The majority of our U.S. participants use office tools to process Spreadsheets (89.5%), Text (76.2%), and Emails (68.6%). As document types, the German participants process Text (90.5%), followed by Spreadsheets (82.1%) and Presentations (65.3%).

[1] E.g. the city of Munich decided to migrate to Windows 10 after it's 2003 decision to adopt Linux, partially due to incompatibility and communication problems with other organizations [13].

	U.S.	German	Combined
Participants			
Started	127	102	229
Finished	110	97	207
Valid ($n =$)	105	95	200
Gender			
Male	66.7%	62.1%	64.5%
Female	33.3%	37.9%	35.5%
Other (Free text)	0.0%	0.0%	0.0%
Age in years			
Mean	35.3	36.1	35.7
Std. dev. (σ)	9.9	11.5	10.7
Median	33.0	33.0	33.0
Computer Science			
CS Education	38.1%	16.8%	28.0%
CS Job	24.8%	20.0%	22.5%
Professional Usage			
Office software at work	80.0%	77.9%	79.0%
Office Usage*			
Google Drive	97.1%	80.0%	89.0%
Microsoft Office (Offline)	86.7%	87.4%	87.0%
Microsoft Office 365 (Cloud)	70.5%	64.2%	67.5%
LibreOffice Offline	18.1%	25.3%	21.5%
Apple's iWork Web (Offline)	9.5%	20.0%	14.5%
Apple's iWork Web (Cloud)	6.7%	17.9%	12.0%
LibreOffice Online	4.8%	9.5%	7.0%
Other	3.8%	5.3%	3.0%
OnlyOffice	1.0%	1.1%	2.5%
Document Usage*			
Spreadsheets	89.5%	82.1%	86.0%
Text	76.2%	90.5%	83.0%
Emails	68.6%	55.8%	62.5%
Presentations	49.5%	65.3%	57.0%
Calendar and Appointments	57.1%	50.5%	54.0%
Other	1.0%	2.1%	1.5%
Document Storage*			
Locally on my computer	73.3%	82.1%	77.5%
Google Drive	88.6%	52.6%	71.5%
Dropbox	33.3%	35.8%	34.5%
OneDrive	30.5%	29.5%	30.0%
iCloud	18.1%	24.2%	21.0%
Network Share	21.0%	16.8%	19.0%

* Multiple answers allowed, may not sum to 100%

Table 3: Demographics for all valid participants from the U.S. survey (Amazon's Mechanical Turk), German survey (ClickWorker), and combined.

Most of our U.S. participants prefer to store their documents in Google Drive (88.6%), followed by locally (73.3%), and Dropbox (33.3%). While the majority of German participants prefers local storage (82.1%), followed by Google Drive (52.6%), and Dropbox (35.8%). This mirrors the distribution of most used office tools for U.S. participants (Google Drive Office $\rightarrow$ Google Drive Storage) and German participants (Offline Microsoft Office $\rightarrow$ Local Storage).

Participants of both the U.S. and German survey agree on the top reasons why they (would) use cloud office applications over local office applications: easy remote access of documents (76.2%, 70.5%), ease of collaboration (58.1%, 59.0%), and free or cheap access (52.4%, 43.2%).

> Summary: Demographics. Somewhat unsurprisingly, participants prefer to store their documents on the platform they edit them with (e.g., locally for offline office). All of our participants agree on the benefits of cloud office applications: free access and easy collaboration for remote documents.

4.2 Document Security

In this question section we asked our participants to think about where they believe their documents are more secure from any unauthorized access, on their personal computer or in the cloud. Most participants reported that they feel their personal computer is more secure for storing their documents than the cloud (54.5% vs. 19.5%).

In addition to the quantitative questions, we asked our participants to explain their assessment. Most (94) of the participants who said they felt their documents would be more secure against unauthorized access on their personal computers mentioned that an attacker would require physical access to their machines to acquire access to documents, e.g., P30 said "*You would have to physically breach my computer to get to the documents, the drive is encrypted, no one can access it.*". Similarly P47 explained

> "*I think that documents are more secure on my computer because I'm the only one that can access them; and if there were any threats on my PC, I would use programs to get rid of them.*" - P47.

Some participants (21) who said files were more secure on their computer thought that it was easier to attack a cloud system than their personal computer, e.g., P27 mentioned that "*Because I know what security I have on my pc, but don't know about Google. Of course, I assume they've got top of the line security, but I don't actually know.*" (P27).

Participants who thought documents in the cloud were more secure (39; 19.5%) mostly mentioned two reasons. First, they believe that cloud office suite providers have more security expertise than they personally do. For example, P79 said

> "*The cloud is managed by big corporations. They probably take security more serious than individuals. They always have to worry about hackers so there [sic] security is likely very powerful.*" - P79.

Second, some participants assessed cloud office suites to be more "secure" than their personal computers because they have backups and losing data is less likely, e.g.,

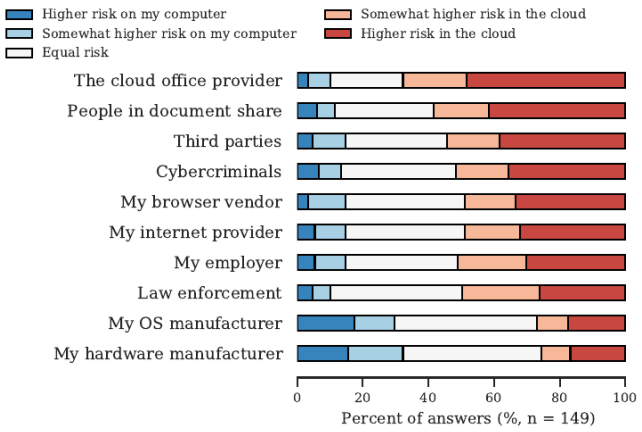


Figure 2: Likert scale for participants' associated risk of unauthorized access between their local computer and their cloud office documents for different parties.

> "*Local computers can be hacked and can crash. It happens. Too often, backups are not made regularly, so data can be lost in either case. With automatic backup to the cloud, documents are more secure in case of local computer issues.*" - P74.

Other reasons for believing in a secure cloud often seem to be based on insufficient technical knowledge, e.g., "*because I think it is not possible to hack the cloud.*" (P219). Few participants (3; 1.5%) mentioned the use of two-factor authentication and the application of encryption by cloud office suite providers as important security factors.

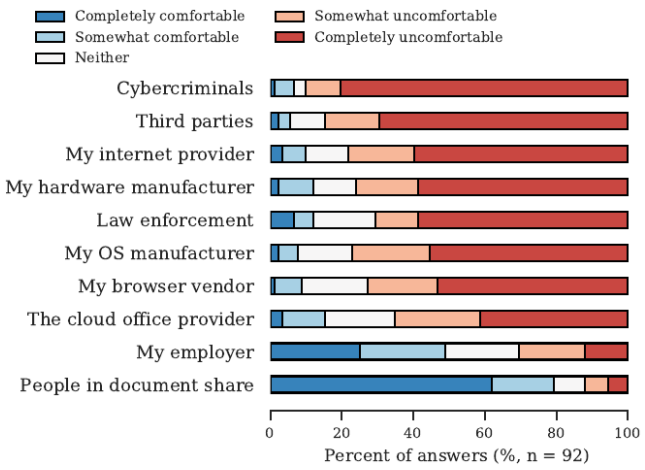
Summary: Document Security. Our participants seem to be aware of some general security implications of processing their documents in the cloud. They seem to prefer their local system in terms of security against unauthorized access, although some of their threat models appear to be underdeveloped.

4.3 Document Access

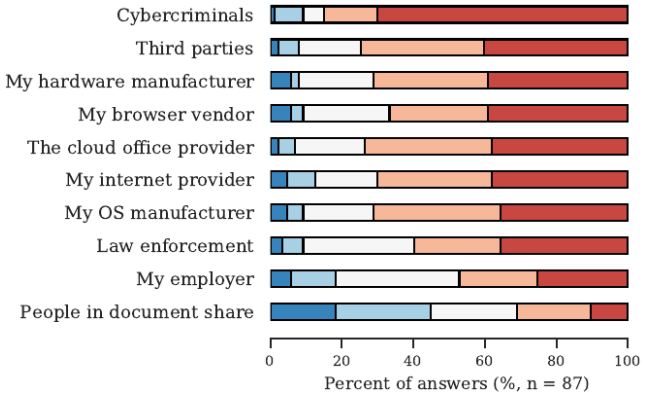
In this question section we explore our participants' perception, misconceptions, and mental models regarding the (unauthorized) access of specific parties to their potentially sensitive documents processed in cloud office applications.

We found a significant difference in the risk of different parties accessing the participants documents (KWH; $H = 102.33$; $p < 0.01$). This might indicate that participants seem to be aware of the changed attack surface for cloud office documents and associate a higher risk of unauthorized access by cybercriminals and third parties such as advertisers and plugin developers in the cloud (cf. Figure 2).

These answers coincide with parties of which participants thought that they already accessed their documents, although some participants have the misconception that their browser



(a) Survey participants from MTurk (U.S.).



(b) Survey participants from CrowdWorker (German).

Figure 3: How comfortable our participants are with different parties accessing their cloud documents. "*I don't know*" answers were omitted.

vendor and operating system provider also have accessed their cloud documents. Figure 3 shows the comfort level of our participants related to the access of different parties to their cloud office documents.

We also asked participants who *would* inform them if their cloud office documents are accessed by an unauthorized party and who *should* inform them. Participants' answers point at a responsible party here: While the German and U.S. participants are split on the cloud office provider (73; 69.5%) and nobody (39; 41.1%) as most common answer on who *would* inform them respectively, both groups agree that it is the cloud office provider that *should* inform them (153; 76.5%).

A large number of participants explicitly told us that they like to be informed about unauthorized access of their cloud office documents by email (119; 59.5%). In addition, some participants provided us with their wishes about the information they want to receive in case of such a data breach, e.g., P69 insisted that

"*[I] [n]eed to know basically everything that the person saw. When they saw it, what they saw, where they're from. I don't care who gives the analysis, just that its an accurate analysis and that they let me know.*" - P69.

Summary: Document Access. Overall, our participants seem to have a clear idea on by whom and how they should be informed about unauthorized access of their cloud documents: the cloud office provider via (secure) email. Our participants seem to have strong opinions on how comfortable they are with the access of certain parties, but are somewhat unsure about who actually has access to their documents.

4.4 Document Storage

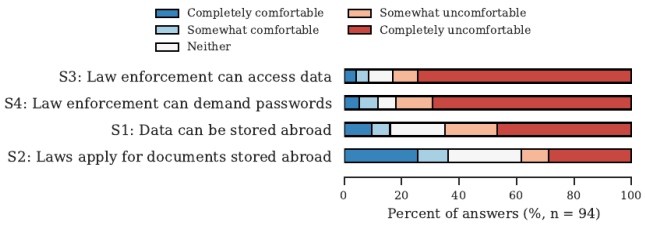
The majority of German participants believe that multiple copies of their cloud office documents exist (49; 51.6%), while most U.S. participants admit that they do not know (51; 48.6%). Of those that assume multiple copies exist (83; 41.5%), the majority thinks that only their copies are deleted if they delete a document (30 of 83; 36.1%), or they are unsure (21 of 83; 25.3%). Unsurprisingly the majority of our participants assume that their cloud provider can delete their documents (138; 69.0%), followed by people they shared the documents with for U.S. participants (43 of 105; 41.0%) and cybercriminals for German participants (46 of 95; 48.4%).

Some of our participants assume a rather basic mental model of why copies of their cloud documents might exist, e.g., P123 believes "*[. . .] that these copies exist just in case that [sic] the original documents get lost.*" (P123). Other participants had a less utilitarian view on the existence of potential copies, e.g., P79 had some rather dystopian thoughts about why copies of their documents are created: "*[T]o use against me when the time is right.*" (P79). For why not all of the copies are deleted, some participants had some very convincing arguments: "*[They are] used to train artificial intelligence or to make a profile of me for the future.*" (P79), "*possibly to sell to 3rd-party vendors for advertising*" (P96), and "*so they can be used for law enforcement.*" (P97).

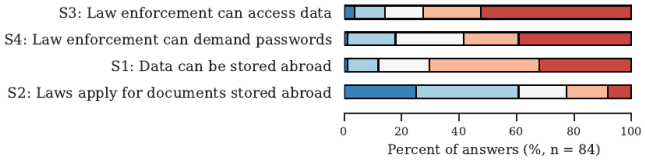
Summary: Document Storage. Overall, our participants seem to be rather unsure about the actual number of copies, access rights, and deletion procedures of their cloud documents. They appear pessimistic regarding the reasons of why additional copies are kept.

4.5 Document Responsibility

In this section we asked participants about which party they think is responsible for the protection of their documents. The majority of U.S. participants sees the cloud provider as responsible (83; 79.0%), while the majority of the German participants sees themselves as responsible (69; 72.6%),



(a) Survey participants from MTurk (U.S.).



(b) Survey participants from CrowdWorker (German).

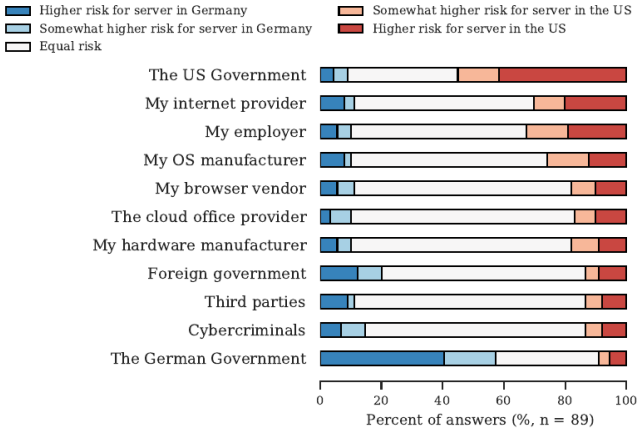
Figure 4: Participants' comfort with potential privacy violations by their government.

We also compared U.S. and German participants in their agreement regarding four scenarios exploring the responsibilities of cloud office providers:

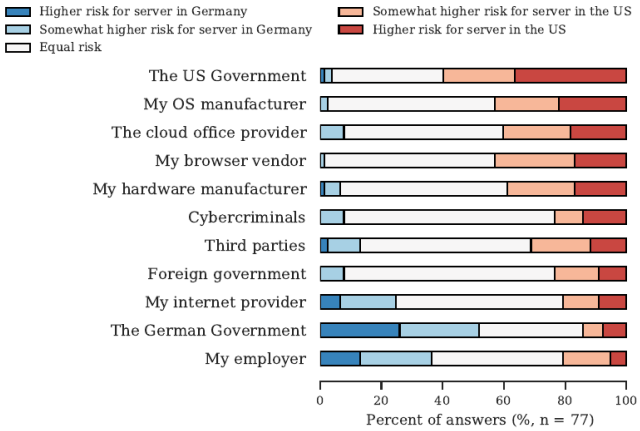
S1: "*Cloud office providers should offer adequate protection for cloud office documents.*"
(MWU; $U = 4445$; adj. p-value $= 1$)
S2: "*I should have the right to demand a full overview of my data collected by cloud office providers.*"
(MWU; $U = 4419$; adj. p-value $= 1$)
S3: "*Upon my request, cloud office providers should have to show what they do with my documents and who has or had access.*"
(MWU; $U = 4181$; adj. p-value $= 1$)
S4: "*Cloud office providers must be able to modify or delete any data they have on private individuals.*"
(MWU; $U = 4566$; adj. p-value $= 1$)

and found no significant differences between our U.S. and German participants. Similarly, we compared U.S. and German participants regarding their (dis)comfort with the following statements (Note that the statements were localized, e.g., an U.S. participant would be presented with "US regulation"):

S1: "*Cloud providers can store my documents on servers outside of the US/Germany without legal repercussions.*"
(MWU; $U = 4151$; adj. p-value $= 1$)
S2: "*US/German regulations and laws still apply if the documents are stored on servers outside of the US/Germany.*"
(MWU; $U = 4817$; adj. p-value $= 0.04$)
S3: "*US/German law enforcement can access my cloud documents without a court order.*"
(MWU; $U = 4768$; adj. p-value $= 0.02$)



(a) Survey participants from MTurk (U.S.).



(b) Survey participants from CrowdWorker (German).

Figure 5: Risk of unauthorized parties accessing participants' documents on servers in the U.S. vs. Germany.

S4: "*US/German law enforcement can force me to give up my cloud office password.*"
(MWU; $U = 5104$; adj. p-value < 0.001)

and found significant differences for S2, S3, and S4. These differences can be mostly attributed to U.S. participants being more uncomfortable with privacy violations by their government compared to the Germans (cf. Figure 4).

We further investigated differences between U.S. and German participants by asking them where they do think the risk is higher of different parties obtaining unauthorized access to their documents if they are either stored on a server in the U.S. or Germany (cf. Figure 5).

> Summary: Document Responsibility. While participants from the U.S. and Germany agree on the responsibilities of cloud providers, U.S. participants are comparably more uncomfortable regarding potential privacy violations by the government.

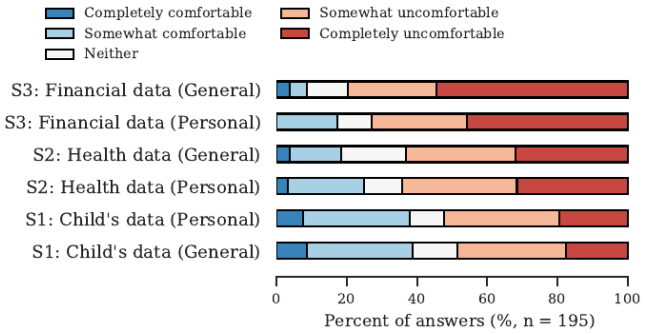


Figure 6: Participants' comfort with three different data scenarios (Financial, Health, and Children) and two different conditions (General and Personal perspective).

Factor	Coef.	C.I.	*p*-value
Scenario: Health	−0.50	[−0.66, −0.33]	$< 0.001*$
Scenario: Financial	−0.88	[−1.05, −0.72]	$< 0.001*$
Condition: Personal	0.03	[−0.25, 0.31]	0.843
Country: Germany	−0.11	[−0.39, 0.17]	0.431

Table 4: Final linear mixed regression model examining the perception of 3 different scenarios in 2 phrasing conditions. "I don't know" answers were omitted. See Section 3.3 and Table 2 for further details.

4.6 Scenario Perception

In this section, we wanted to explore the effect of different conditions and scenarios on how comfortable our participants are with processing documents in the cloud. For this, our participants were presented with three different types of private data stored in cloud documents: children data including names and grades, health data including names and diagnosis, and financial data including names and SSNs. As additional modifier, participants were equally distributed across two conditions: "General" with a more generalized phrasing and "Personal" with a more personalized phrasing (e.g., "a child" vs. "your child").

We explored participants' answers by selecting the best performing model from multiple linear regressions (cf. Table 4). We find that neither the country nor the condition has a significant coefficient in the regression. Both the "Health data" scenario and the "Financial data" scenario are significantly rated as less comfortable by our participants than the "Child data" baseline (cf. Figure 6).

> Summary: Scenario Perception. Our participants are uncomfortable the most with the scenario of processing financial documents in the cloud. Presenting a more personalized scenario nor their country did not significantly affect their comfort level.

4.7 Data Protection

In this section, we asked two free text questions to assess the amount of data our participants think cloud office suite providers collect when processing documents. Additionally, we asked our participants what security measures they think cloud providers deploy to protect their documents. Regarding data collection, most participants thought that cloud office suite providers collected the actual document content and metadata including the time and duration they used the cloud office application, IP addresses and filenames. A few participants were concerned that cloud providers would search their documents for keywords and report them to security agencies and law enforcement, e.g. P96 thinks that providers are "*searching for specific keywords, most notably for US security reasons*" (P96).

Most participants had very specific ideas of what security measures cloud office suite providers would deploy. The majority of our participants were convinced that providers would deploy encryption to protect their stored documents. For example, P74 believes that "*the cloud servers are supposed to be encrypted and follow industry-standard protocols [. . .]*" (P74). Similarly, participants mentioned access control and authentication, e.g., P79 hopes that "*Security is handled by the service provider of the cloud office applications. They probably use complicated passwords and 2 factor authentication.*" (P79). Finally, some participants mentioned firewalls and other network security measures. P111 hopes that "*they are protected by multiple firewalls [and] they are continuously monitored*" (P111).

> Summary: Data Protection. While our participants are aware that the content of their documents might be collected, only few were concerned that specific keywords might be reported to law enforcement. Our participants identified encryption as their preferred security measure their cloud office suite should employ.

5 Related Work

As we conduct surveys investigating end-user security and privacy perceptions, as well as expectations with cloud office suites, we discuss related work in the areas of security & privacy in the cloud and user studies within a context of cloud applications or cloud storage.

Security & Privacy in the Cloud. Past research in the cloud often investigated the privacy of data and sharing, a field still not fully solved judging by the overall unclear or pessimistic views of our participants. The backup and restore performance, liabilities, and problems with data privacy of four cloud storage providers was examined by Hu et al. in 2010 [14]. Also in 2010, Svantesson and Clarke reviewed the terms of use of Google Docs finding that cloud computing is associated with risks to privacy and consumer rights [39]. Johnson proposed in 2017 that the cloud providers should make changes to their terms of service to allow the users better control over their privacy [17], a proposal supported by our work. Similarly, Nestori et al. found in their 2018 paper, that Office 365 is not GDPR compliant [29]. MUBox introduced a meta-cloud storage application to help improve user collaboration on cloud storage services by introducing activity views and Nebeling et al. conducted a user study with 16 participants to examine accuracy and confidence with the activity views [28]. Massey et al. conducted a qualitative study with 27 participants and identified four different strategies that teams used in shared repositories and suggested ways to improve existing tools with new technologies [24].

A number of works concerns client-side encryption or hiding layers to prevent third parties including the cloud office provider from accessing the content of any document edited in the cloud [1, 8, 41], further underlining the need for native encryption, as identified by our participants.

User Studies of the Cloud. Often user surveys in the cloud context focus on the storage aspect: Tan et al. investigated the acceptance of SaaS collaboration tools like Google Docs in an organizational setting and found that their intention to continue using these tools is positively affected by the perceived usefulness and satisfaction [40], which corresponds to our findings regarding ease of use and sharing. Marshall et al. conducted a survey with 106 participants and 19 interviews to understand early user experiences and models of cloud storage systems, finding that users' misconceptions limit the ability to take full advantage of cloud features [23]. Burda et al. developed a technology acceptance model which incorporates users' perception of risk and trust and verified it in a study with 229 cloud storage users. They found evidence that trust in cloud archiving can be increased by a providers' reputation and user satisfaction [3].

Both Clark et al. and Khan et al. explored users' perception of file sharing status over time, finding a mismatch in user expectations and reality [6, 19]. Ramokapane et al. conducted a user study, finding that users struggle to delete their data from the cloud, as incomplete or inaccurate mental models based on a lack of information on deletion lead to a failure to remove the data properly [32]. Mijuskovic et al. conducted a qualitative user study with 28 participants and found that most users are aware of security and privacy risks in the cloud, but lack knowledge to describe potential risks in detail [26]. These previous studies agree with our findings of incomplete mental models, often due to lacking technical knowledge.

We consider the following works by Ion et al. and Arpaci et al. closest to our surveys. Ion et al. studied privacy attitudes and beliefs towards consumer cloud storage by conducting interviews and a survey with end-users in Switzerland and India, finding that requirements for consumer cloud storage differ from those of companies and that end-users prefer local offline storage for sensitive data [16]. Arpaci et al. conducted a study with 200 pre-service teachers to understand the effects

of security and privacy concerns of cloud computing in educational use and proposed a research model that indicates that security and privacy perception has a significant influence on students' attitudes towards cloud services [2].

Compared to these earlier studies consisting mostly of small-scale qualitative studies investigating the acceptance of cloud technology or larger studies focusing on cloud storage, our larger-scale study ($n = 200$) with both qualitative and quantitative parts investigates security and privacy explicitly in the context of cloud office suites.

6 Discussion

In this work, we explored the security and privacy perceptions and expectations of cloud office users, as well as their intuitions for how cloud office suites should ideally handle security and privacy. We performed two online surveys with 200 cloud office users from the U.S. and Germany to explore the following research questions:

RQ1: "*How and why do our participants interact with cloud office applications?*" The fairly recent shift from offline-only tools to cloud office suites includes immense changes of privacy and security implications, although the application design and end user experience remained mostly the same or even included new features. We find that a large majority of our participants regularly work on different document types in cloud office applications. The most common reason for using cloud office applications are the ease of sharing and the ease of use without requiring installation of additional software.

RQ2: "*What are end users' awareness, perceptions, and attitudes about privacy in cloud office applications?*" Users seem to be aware of some general security implications, storage models, and access by others, although some of their threat models seem underdeveloped (e.g., "*I think it is not possible to hack the cloud.*"), likely due to lacking technical knowledge.

RQ3: "*What are participants' mental models regarding protection and security of their cloud documents?*" We find that users' mental models regarding access and sharing are incomplete and their understanding of cloud office security and privacy is limited, likely caused by a lack of transparency of the services' operations.

Our findings suggest that the current state of cloud office suites leaves much to be desired in the eyes of end users. General misconceptions and the unclear responsibilities of cloud providers might result in additional challenges for end user adoption of cloud office suites.

6.1 Recommendations

Based on our findings, we offer recommendations for groups associated with cloud office suites.

For industry: Since our participants were somewhat unsure about who actually has access to their documents (Section 4.3), we recommend changes to user interfaces and sharing policies that will improve their awareness. In case of unauthorized access, we recommend notifications via email, as most of our participants prefer their provider to inform them this way (Section 4.5). Participants also identified encryption as their preferred security measure their cloud office suite should employ for improved security (Section 4.7).

For end users: A number of self-hosted alternatives to cloud office applications, such as Seafile or NextCloud, allow for most of the cloud conveniences while you retain full control of your data (Section 2).

For policy makers: Our participants are somewhat unsure about who actually has access to their documents and about how many copies actually exist on which servers (Section 4.3). Privacy-focused policies such as GDPR could serve as a first step for improving security and privacy considerations for end users and could enable more privacy-friendly applications. In addition, data-at-rest and responsible disclosure policies could help with user wishes such as prefer encryption measures and notifications by email in case of unauthorized access (Sections 4.5, 4.7).

7 Conclusion

This paper provides a comprehensive insight into the awareness, perceptions, and attitudes of cloud office users, their general usage, and basic mental models.

We find that participants commonly use cloud office suites, mainly for the convenience of free access and easy collaboration for remote documents. Compared to local offline office suites, they voice security and privacy concerns, mainly in terms of unauthorized access. They are somewhat uncomfortable with the security implications of processing their documents in the cloud, however, their threat models remain vague. Our participants have strong opinions on how comfortable they are with the access of certain parties, but are somewhat unsure about who actually has access to their documents. In cases of unauthorized access, participants clearly place the responsibility of informing them of the breach on the cloud office provider, preferably via email.

U.S. and German participants' perceptions, awareness and attitudes closely resembled one another, except that U.S. participants were more uncomfortable with government access to their cloud office data.

We think that, in light of the popularity and widespread use of cloud office suites, participants should be able to make informed decisions about their security and privacy.

We hope that our recommendations for different groups associated with cloud office suites, can help inform future standards, regulations, and implementations.

References

[1] Lilian Adkinson-Orellana, Daniel A Rodríguez-Silva, Felipe Gil-Castiñeira, and Juan C Burguillo-Rial. Privacy for Google Docs: Implementing a transparent encryption layer. In *Proc. 2nd Cloud Computing International Conference - CloudViews*, 2010.

[2] Ibrahim Arpaci, Kerem Kilicer, and Salih Bardakci. Effects of security and privacy concerns on educational use of cloud services. *Computers in Human Behavior*, 45:93–98, 2015.

[3] Daniel Burda and Frank Teuteberg. The role of trust and risk perceptions in cloud archiving—results from an empirical study. *The Journal of High Technology Management Research*, 25(2):172–187, 2014.

[4] K. P. Burnham. Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods & Research*, 33(2):261–304, 2004.

[5] Kathy Charmaz. *Constructing Grounded Theory*. SAGE Publications, 2014.

[6] Jason W Clark, Peter Snyder, Damon McCoy, and Chris Kanich. I saw images I didn't even know I had: Understanding user perceptions of cloud storage privacy. In *Proc. 33rd ACM Conference on Human Factors in Computing Systems (CHI'15)*, pages 1641–1644. ACM, 2015.

[7] Juliet Corbin and Anselm Strauss. Grounded theory research: Procedures, canons and evaluative criteria. *Zeitschrift für Soziologie*, 19(6):418–427, 1990.

[8] Gabriele D'Angelo, Fabio Vitali, and Stefano Zacchiroli. Content Cloaking: Preserving Privacy with Google Docs and Other Web Applications. In *Proc. 25th ACM Symposium on Applied Computing (SAC'10)*, pages 826–830. ACM, 2010.

[9] Esat Dedezade. Microsoft to deliver cloud services from new datacentres in Germany in 2019 to meet evolving customer needs. https://news.microsoft.com/europe/2018/08/31/microsoft-to-deliver-cloud-services-from-new-datacentres-in-germany-in-2019-to-meet-evolving-customer-needs/, August 2019.

[10] D. Dittrich and E. Kenneally. The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research. Technical report, U.S. Department of Homeland Security, August 2012.

[11] The Document Foundation. LibreOffice Online | LibreOffice - Free Office Suite. https://www.libreoffice.org/download/libreoffice-online/, June 2020.

[12] Google. Google Drive: Free Cloud Storage for Personal Use. https://www.google.com/intl/en/drive/, June 2020.

[13] Nick Heath. From linux to windows 10: Why did munich switch and why does it matter? https://www.techrepublic.com/article/linux-to-windows-10-why-did-munich-switch-and-why-does-it-matter/, November 2017.

[14] Wenjin Hu, Tao Yang, and Jeanna N Matthews. The good, the bad and the ugly of consumer cloud storage. *Operating Systems Review*, 44(3):110–115, 2010.

[15] Apple Inc. iWork - Apple. https://www.apple.com/iwork/, June 2020.

[16] Iulia Ion, Niharika Sachdeva, Ponnurangam Kumaraguru, and Srdjan Čapkun. Home is Safer Than the Cloud!: Privacy Concerns for Consumer Cloud Storage. In *Proc. 7th Symposium on Usable Privacy and Security (SOUPS'11)*. ACM, 2011.

[17] Eric Johnson. Lost in the cloud: Cloud storage, privacy, and suggestions for protecting users' data. *Stan. L. Rev.*, 69:867–909, 2017.

[18] Ryan Kennedy, Scott Clifford, Tyler Burleigh, Ryan Jewell, and Philip Waggoner. The Shape of and Solutions to the MTurk Quality Crisis. *Political Science Research and Methods*, pages 1–16, May 2018.

[19] Mohammad Taha Khan, Maria Hyun, Chris Kanich, and Blase Ur. Forgotten But Not Gone: Identifying the Need for Longitudinal Data Management in Cloud Storage. In *Proc. 36th ACM Conference on Human Factors in Computing Systems (CHI'18)*, pages 543:1–543:12. ACM, 2018.

[20] Klaus Krippendorff. *Content Analysis: An Introduction to Its Methodology (2nd ed.)*. SAGE Publications, 2004.

[21] Paul Lorimer. Microsoft office 365 and dynamics 365 now available from new german datacenter regions. https://www.microsoft.com/en-us/microsoft-365/blog/2020/02/20/microsoft-office-365-dynamics-365-now-available-from-new-german-datacenter-regions/, February 2020.

[22] Peter V Marsden and James D Wright. *Handbook of survey research*. Emerald Group Publishing, 2010.

[23] Cathy Marshall and John C Tang. That syncing feeling: early user experiences with the cloud. In *Proc. 9th ACM Conference on Designing Interactive Systems (DIS'12)*, pages 544–553. ACM, 2012.

[24] Charlotte Massey, Thomas Lennig, and Steve Whittaker. Cloudy forecast: an exploration of the factors underlying shared repository use. In *Proc. 32nd ACM Conference on Human Factors in Computing Systems (CHI'14)*, pages 2461–2470. ACM, 2014.

[25] Microsoft. Office 365 | Microsoft Office. https://www.office.com/, June 2020.

[26] Adriana Mijuskovic and Mexhid Ferati. User awareness of existing privacy and security risks when storing data in the cloud. In *Proc. International Conference on e-Learning and the Knowledge Society*, pages 268–273. European Commission, 2015.

[27] David Morris. General Dynamics wins $7.6 billion contract to supply Microsoft office software to the Pentagon. https://fortune.com/2019/08/29/general-dynamics-pentagon-contract/, August 2019.

[28] Michael Nebeling, Matthias Geel, Oleksiy Syrotkin, and Moira C Norrie. MUBox: Multi-user aware personal cloud storage. In *Proc. 33rd ACM Conference on Human Factors in Computing Systems (CHI'15)*, pages 1855–1864. ACM, 2015.

[29] Syynimaa Nestori and Viitanen Tessa. Is My Office 365 GDPR Compliant? - Security Issues in Authentication and Administration. In *Proc. 20th International Conference on Enterprise Information Systems, ICEIS'18*, pages 299–305, 2018.

[30] Eyal Peer, Joachim Vosgerau, and Alessandro Acquisti. Reputation as a sufficient condition for data quality on amazon mechanical turk. *Behavior Research Methods*, 46, December 2013.

[31] Stanley Presser, Mick P. Couper, Judith T. Lessler, Elizabeth Martin, Jean Martin, Jennifer M. Rothgeb, and Eleanor Singer. Methods for testing and evaluating survey questions. *Public Opinion Quarterly*, 68(1):109–130, March 2004.

[32] Kopo Marvin Ramokapane, Awais Rashid, and Jose Miguel Such. "I feel stupid I can't delete ...": A Study of Users' Cloud Deletion Practices and Coping Strategies. In *Proc. 13th Symposium on Usable Privacy and Security (SOUPS'17)*, pages 241–256, 2017.

[33] Elissa M. Redmiles, Sean Kross, and Michelle L. Mazurek. How I Learned to Be Secure: A Census-Representative Survey of Security Advice Sources and Behavior. In *Proc. 23nd ACM Conference on Computer and Communication Security (CCS'16)*. ACM, 2016.

[34] Jim Salter. Office 365 declared illegal in german schools due to privacy risks. https://arstechnica.com/information-technology/2019/07/germany-threatens-to-break-up-with-microsoft-office-again/, July 2019.

[35] Nora Cate Schaeffer and Stanley Presser. The science of asking questions. *Annual Review of Sociology*, 29(1):65–88, 2003.

[36] Cathrin Schaer. Microsoft Office 365: Banned in German schools over privacy fears. https://www.zdnet.com/article/microsoft-office-365-banned-in-german-schools-over-privacy-fears/, Jul 2019.

[37] Ascensio System SIA. Cloud online office suite for business management | ONLYOFFICE. https://www.onlyoffice.com/cloud-office.aspx, June 2020.

[38] Anselm Strauss and Juliet M Corbin. *Grounded theory in practice*. SAGE Publications, 1997.

[39] Dan Svantesson and Roger Clarke. Privacy and consumer risks in cloud computing. *Computer Law & Security Review*, 26(4):391–397, 2010.

[40] Xin Tan and Yongbeom Kim. User acceptance of SaaS-based collaboration tools: a case of Google Docs. *Journal of Enterprise Information Management*, 28(3):423–442, 2015.

[41] Marten van Dijk and Ari Juels. On the Impossibility of Cryptography Alone for Privacy-Preserving Cloud Computing. In *Proc. 6th USENIX Workshop on Hot Topics in Security (HotSec'11)*, pages 1–8. USENIX Association, 2010.

A Survey

The following survey is the English version of the survey, the German version followed the same structure with nearly identical questions. Differences in questions included localization changes, e.g., for country-specific agencies and institutions. Question numbers were not displayed to the participants and order of answer options was generally randomized.

Consent Form

[Consent Form with contact information.]

Please indicate, in the box below, that you are at least 18 years old, have read and understood this consent form, and you agree to participate in this online research study.

- □ I am age 18 or older.
- □ I have read this consent form or had it read to me.
- □ I am comfortable using the English language to participate in this study.
- □ I have used cloud office software before (e.g., Google Drive or Microsoft Office 365).
- □ I agree to participate in this research and I want to continue with the study.

Office demographics

For this survey, we are interested in your experience with and use of Cloud Office Suites and applications.
Cloud Office Application or Online Office Application are software that can be used to create office documents in a web browser, without requiring the installation of a dedicated software.
Examples for Cloud Office Suites are Google Docs/Sheets/Slides, Microsoft Office 365, and LibreOffice Online.

Q1.1: Which office suites have you used before?
(Please select all that apply)

- □ Microsoft Office (Offline; Word, Excel, Powerpoint, ...)
- □ Microsoft Office 365 (Cloud-based; Word, Excel, Powerpoint, ...)
- □ LibreOffice (Offline; Writer, Calc, ...)
- □ LibreOffice Online (Cloud-based; Writer, Calc, ...)
- □ Google Drive (Cloud-based; Docs, Sheets, Slides, ...)
- □ Apple's iWork App (Offline; Pages, Numbers. Keynote...)
- □ Apple's iWork Web (Cloud-based; Pages, Numbers. Keynote...)
- □ OnlyOffice
- □ Other (please specify): _____

Q1.2: Which office suites have you used this month?
(Please select all that apply)

- □ Microsoft Office (Offline; Word, Excel, Powerpoint, ...)
- □ Microsoft Office 365 (Cloud-based; Word, Excel, Powerpoint, ...)
- □ LibreOffice (Offline; Writer, Calc, ...)
- □ LibreOffice Online (Cloud-based; Writer, Calc, ...)
- □ Google Drive (Cloud-based; Docs, Sheets, Slides, ...)
- □ Apple's iWork App (Offline; Pages, Numbers. Keynote...)
- □ Apple's iWork Web (Cloud-based; Pages, Numbers. Keynote...)
- □ OnlyOffice
- □ Other (please specify): _____

Q1.3: Does your job involve using office applications on a regular basis?

- ○ Yes
- ○ No
- ○ I don't know
- ○ I'd prefer not to answer

Q1.4: Which types of documents do you process with office suites?
For this question, please give answers both for your job and your personal life.
(Please select all that apply)

- □ Text (Reports, Letters, etc.)
- □ Spreadsheets (Numbers, Dates, etc.)
- □ Presentations
- □ Calendar and Appointments
- □ Emails
- □ Other (please specify): _____

Q1.5: How do you store your documents?
For this question, please give answers for any documents you might store, including personal and work documents, including but not limited to documents that you edit with office applications.
(Please select all that apply)

- □ Locally on my computer
- □ My office suite saves them online automatically.
- □ Dropbox
- □ Google Drive
- □ Network Share
- □ Self-hosted cloud service
- □ OneDrive
- □ iCloud
- □ Other (please specify): _____

Q1.6: Why do you use **cloud** office applications (compared to local office applications)?
(Please select all that apply)

- □ Provided or required by work
- □ Easy remote access (e.g., from multiple devices)
- □ Ease of collaboration
- □ No installation required
- □ Built-in backup of documents
- □ Free / cheap access
- □ Other (please specify): _____

Document Safety

Q2.1: Where do you think your documents are more secure from any unauthorized access?
[Matrix question, the scale for answers is:]

- More secure on my computer
- Somewhat more secure on my computer
- Equally secure
- Somewhat more secure in the cloud
- More secure in the cloud

- I don't know

[The questions are:]

- Word documents
- Presentations
- Spreadsheets
- E-Mails
- Calendar and Appointments

Q2.2: Why (if at all) do you think your documents may be more secure **on your computer**?
[Free text field]

Q2.3: Why (if at all) do you think your documents may be more secure **in the cloud**?
[Free text field]

Document Access

Q3.1: Who else besides yourself might be able to access the documents you edit in cloud office applications?
(Please select all that apply)

- □ People I share the documents with
- □ My employer
- □ My internet provider
- □ The cloud office provider (e.g., Google or Microsoft)
- □ My browser vendor (e.g., Google or Mozilla)
- □ My operating system manufacturer (e.g., Apple or Microsoft)
- □ Cybercriminals (e.g., hackers or organized crime)
- □ Law enforcement or intelligence agencies (e.g., police, FBI or NSA)
- □ Third parties (e.g., online advertisers or plugin developers)
- □ The manufacturer of my computer hardware (e.g., Intel, AMD, Apple, or Lenovo)
- □ Other (please specify): _____

[The following 3 questions are matrix questions with the following options:]

- People I share the documents with
- My employer
- My internet provider
- The cloud office provider (e.g., Google or Microsoft)
- My browser vendor (e.g., Google or Mozilla)
- My operating system manufacturer (e.g., Apple or Microsoft)
- Cybercriminals (e.g., hackers or organized crime)
- Law enforcement or intelligence agencies (e.g., police, FBI or NSA)
- Third parties (e.g., online advertisers or plugin developers)
- The manufacturer of my computer hardware (e.g., Intel, AMD, Apple, or Lenovo)

Q3.2: Where do you think the risk is higher that the following parties can obtain unauthorized access to your cloud office documents?

- Higher risk on my computer
- Somewhat higher risk on my computer
- Equal risk
- Somewhat higher risk in the cloud
- Higher risk in the cloud
- I don't know

Q3.3: Do you think that any of these parties have already accessed your documents?

- Yes
- No
- I don't know

Q3.4: Please rate your level of (dis)comfort with the potential access of these parties to your cloud office documents.

- Completely comfortable
- Somewhat comfortable
- Neither
- Somewhat uncomfortable
- Completely uncomfortable
- I don't know

Q3.5: Who do you think **would** inform you if an unauthorized party or person accessed you documents?
(Please select all that apply)

- □ People I share the documents with
- □ My employer
- □ My internet provider
- □ The cloud office provider (e.g., Google or Microsoft)
- □ My browser vendor (e.g., Google or Mozilla)
- □ My operating system manufacturer (e.g., Apple or Microsoft)
- □ Law enforcement or intelligence agencies (e.g., police, FBI or NSA)
- □ Third parties (e.g., online advertisers or plugin developers)
- □ The manufacturer of my computer hardware (e.g., Intel, AMD, Apple, or Lenovo)
- □ The news
- □ Scientists
- □ Nobody would inform me
- □ Other (please specify): _____

Q3.6: Who do you think **should be responsible** for informing you if an unauthorized party or person accessed your documents?
(Please select all that apply)

- □ People I share the documents with
- □ My employer

- □ My internet provider
- □ The cloud office provider (e.g., Google or Microsoft)
- □ My browser vendor (e.g., Google or Mozilla)
- □ My operating system manufacturer (e.g., Apple or Microsoft)
- □ Law enforcement or intelligence agencies (e.g., police, FBI or NSA)
- □ Third parties (e.g., online advertisers or plugin developers)
- □ The manufacturer of my computer hardware (e.g., Intel, AMD, Apple, or Lenovo)
- □ The news
- □ Scientists
- □ Nobody would inform me
- □ Other (please specify): _____

Q3.7: **How** would you like to be informed if an unauthorized party or person accessed your cloud office documents?
[Free text field]

Document Storage

Q4.1: Do you think that multiple copies of your cloud office documents exist?
These can be documents that are shared with others or private documents.

- ○ Yes
- ○ No
- ○ I don't know
- ○ I'd prefer not to answer

Q4.2: [only shown if Q4.1 = Yes] For which purpose do you think these copies might exist?
[Free text field]

Q4.3: [only shown if Q4.1 = Yes] In which geographic locations do you think your cloud office documents and copies of these are stored?
[Free text field]

Q4.4: [only shown if Q4.1 = Yes] Which of the copies do you think are actually removed if you delete a cloud office document?

- ○ All
- ○ Mine and my collaborators'
- ○ Only mine
- ○ Only my collaborators'
- ○ None
- ○ I don't know
- ○ I'd prefer not to answer
- ○ Other (please specify):

Q4.5: [only shown if Q4.1 = Yes and Q4.4 != All] **Where or with whom** do you think copies remain?
[Free text field]

Q4.6: [only shown if Q4.1 = Yes and Q4.4 != All] For which purpose do you think that the copies remain?
[Free text field]

Q4.7: Who do you think can delete your documents?
(Please select all that apply)

- □ People I share the documents with
- □ My employer
- □ My internet provider
- □ The cloud office provider (e.g., Google or Microsoft)
- □ My browser vendor (e.g., Google or Mozilla)
- □ My operating system manufacturer (e.g., Apple or Microsoft)
- □ Cybercriminals (e.g., hackers or organized crime)
- □ Law enforcement or intelligence agencies (e.g., police, FBI or NSA)
- □ Third parties (e.g., online advertisers or plugin developers)
- □ The manufacturer of my computer hardware (e.g., Intel, AMD, Apple, or Lenovo)
- □ Other (please specify): _____

Q4.8: Who do you think is responsible for protecting your data?
(Please select all that apply)

- □ People I share the documents with
- □ My employer
- □ My internet provider
- □ The cloud office provider (e.g., Google or Microsoft)
- □ My browser vendor (e.g., Google or Mozilla)
- □ My operating system manufacturer (e.g., Apple or Microsoft)
- □ Cybercriminals (e.g., hackers or organized crime)
- □ Law enforcement or intelligence agencies (e.g., police, FBI or NSA)
- □ Third parties (e.g., online advertisers or plugin developers)
- □ The manufacturer of my computer hardware (e.g., Intel, AMD, Apple, or Lenovo)
- □ Myself
- □ The US-Government
- □ Other (please specify): _____

Responsibility

Q5.1: Please indicate your agreement with the following statements:
[5 point-likert scale from Strongly agree to Strongly disagree + I don't know option]

- Cloud office providers should offer adequate protection for cloud office documents (e.g., by encryption and well implemented security practices)
- I should have the right to demand a full overview of my data collected by cloud office providers.

- Upon my request, cloud office providers should have to show what they do with my documents and who has or had access.
- Cloud office providers must be able to modify or delete any data they have on private individuals.

Q5.2: Please indicate your (dis)comfort with the following statements:
[5 point-likert scale from Completely comfortable to Completely uncomfortable + I don't know option]

- Cloud providers can store my documents on servers outside of the US without legal repercussions.
- US regulations and laws still apply if the documents are stored on servers outside of the US.
- US law enforcement can access my cloud documents without a court order.
- US law enforcement can force me to give up my cloud office password.

Q5.3: Where do you think the risk is higher of somebody **obtaining unauthorized access** to your documents if they are either stored on a server in Germany or the US?
[5 point-likert scale from "Higher risk for server in Germany" to "Higher risk for server in the US" + I don't know option]

- My employer
- My internet provider
- The cloud office provider (e.g., Google or Microsoft)
- My browser vendor (e.g., Google or Mozilla)
- My operating system manufacturer (e.g., Apple or Microsoft)
- Cybercriminals (e.g., hackers or organized crime)
- Third parties (e.g., online advertisers or plugin developers)
- The manufacturer of my computer hardware (e.g., Intel, AMD, Apple, or Lenovo)
- US government
- German governments
- Foreign government (neither US nor German)

Personal Perception - Scenario A - Personalized Scenario

[Only scenario block A or B was randomly shown to the participants]
[Question order was randomized]
Below are listed three different scenarios. How comfortable do you feel with each approach?

Q6.A.1: Your child is required by the school to use a cloud office suite for tasks. The processed documents include private information such as your child's name and grades.

- ○ Completely comfortable
- ○ Somewhat comfortable
- ○ Neither
- ○ Somewhat uncomfortable
- ○ Completely uncomfortable
- ○ I don't know

Q6.A.2: Your general practitioner uses a cloud office suite to process patient data. The processed documents include private information such as your name, age, weight, diagnosis, and treatment plan.

- ○ Completely comfortable
- ○ Somewhat comfortable
- ○ Neither
- ○ Somewhat uncomfortable
- ○ Completely uncomfortable
- ○ I don't know

Q6.A.3: Your financial advisor uses a cloud office suite to process client data. The processed documents include private information such as your name, SSN, and financial information.

- ○ Completely comfortable
- ○ Somewhat comfortable
- ○ Neither
- ○ Somewhat uncomfortable
- ○ Completely uncomfortable
- ○ I don't know

Personal Perception - Scenario B - Generalized Scenario

[Only scenario block A or B was randomly shown to the participants]
[Question order was randomized]
Below are listed three different scenarios. How comfortable do you feel with each approach?

Q6.B.1: A school requires children to use a cloud office suite for tasks. The processed documents include private information such as children names and grades.

- ○ Completely comfortable
- ○ Somewhat comfortable
- ○ Neither
- ○ Somewhat uncomfortable
- ○ Completely uncomfortable
- ○ I don't know

Q6.B.2: A doctor's office uses a cloud office suite to process patient data. The processed documents include private information such as name, age, weight, diagnosis, and treatment plans.

- ○ Completely comfortable
- ○ Somewhat comfortable
- ○ Neither
- ○ Somewhat uncomfortable
- ○ Completely uncomfortable
- ○ I don't know

Q6.B.3: A financial advisor's office uses a cloud office suite to process client data. The processed documents include private information such as name, SSN, and financial information.

- ○ Completely comfortable
- ○ Somewhat comfortable
- ○ Neither
- ○ Somewhat uncomfortable
- ○ Completely uncomfortable
- ○ I don't know

Data Protection

Q7.1: What do you think — what data does the cloud office application collect when you process documents with it?
[Free text field]

Q7.2: How do you think documents processed by cloud office applications are protected?
[Free text field]

GDPR

Q8.1: Do you know what the GDPR is?

- ○ A data protection regulation in EU law
- ○ A plugin for Google Drive
- ○ A cloud office provider
- ○ A counter terrorism act in US law
- ○ I don't know
- ○ I'd prefer not to answer

Q8.2: [Only shown if Q8.1 = A data protection regulation in EU law] What do you think does the GDPR protect?
[Free text field]

Demographics

[We administered demographic questions at the end of the questionnaire to prevent stereotype bias.]

Q9.1: How old are you? (in years, e.g. 42. Optional)
[Free text field]

Q9.2: As which gender do you identify?

- ○ Male
- ○ Female
- ○ [Free text field]
- ○ I'd prefer not to answer

Q9.3: Do you have formal education (Bachelor's degree or higher) in computer science, information technology, or a related field?

- ○ Yes
- ○ No
- ○ I'd prefer not to answer

Q9.4: Have you held a job in computer science, information technology, or a related field?

- ○ Yes
- ○ No
- ○ I'd prefer not to answer

Q9.5: Do you have any feedback or additional comments for us? (completely optional)
[Free text field]