

“Make Them Change it Every Week!”: A Qualitative Exploration of Online Developer Advice on Usable and Secure Authentication*

Jan H. Klemmer
Leibniz University Hannover
Hannover, Germany
klemmer@sec.uni-hannover.de

Marco Gutfleisch
Ruhr University Bochum
Bochum, Germany
marco.gutfleisch@rub.de

Christian Stransky
CISPA Helmholtz Center for
Information Security
Hannover, Germany
stransky@sec.uni-hannover.de

Yasemin Acar
Paderborn University
Paderborn, Germany
yasemin.acar@uni-paderborn.de

M. Angela Sasse
Ruhr University Bochum
Bochum, Germany
martina.sasse@rub.de

Sascha Fahl
CISPA Helmholtz Center for
Information Security
Hannover, Germany
sascha.fahl@cispa.de

ABSTRACT

Usable and secure authentication on the web and beyond is mission-critical. While password-based authentication is still widespread, users have trouble dealing with potentially hundreds of online accounts and their passwords. Alternatives or extensions such as multi-factor authentication have their own challenges and find only limited adoption. Finding the right balance between security and usability is challenging for developers. Previous work found that developers use online resources to inform security decisions when writing code. Similar to other areas, lots of authentication advice for developers is available online, including blog posts, discussions on Stack Overflow, research papers, or guidelines by institutions like OWASP or NIST.

We are the first to explore developer advice on authentication that affects usable security for end-users. Based on a survey with 18 professional web developers, we obtained 406 documents and qualitatively analyzed 272 contained pieces of advice in depth. We aim to understand the accessibility and quality of online advice and provide insights into how online advice might contribute to (in)secure and (un)usable authentication. We find that advice is scattered and that finding recommendable, consistent advice is a challenge for developers, among others. The most common advice is for password-based authentication, but little for more modern alternatives. Unfortunately, many pieces of advice are debatable (e.g., complex password policies), outdated (e.g., enforcing regular password changes), or contradicting and might lead to unusable or insecure authentication. Based on our findings, we make recommendations for developers, advice providers, official institutions, and academia on how to improve online advice for developers.

*This paper has an extended version: <https://arxiv.org/abs/2309.00744>.



This work is licensed under a Creative Commons Attribution International 4.0 License.

CCS CONCEPTS

• **Security and privacy** → **Usability in security and privacy**;
Software security engineering; Authentication.

KEYWORDS

Usable Security; Advice; Software Development; Authentication

ACM Reference Format:

Jan H. Klemmer, Marco Gutfleisch, Christian Stransky, Yasemin Acar, M. Angela Sasse, and Sascha Fahl. 2023. “Make Them Change it Every Week!”: A Qualitative Exploration of Online Developer Advice on Usable and Secure Authentication. In *Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security (CCS '23)*, November 26–30, 2023, Copenhagen, Denmark. ACM, New York, NY, USA, 15 pages. <https://doi.org/10.1145/3576915.3623072>

1 INTRODUCTION

Password-based authentication is the status quo on the web and beyond [6, 12, 13]—despite its many known problems. This includes challenges around memorizing passwords, password reuse [31, 72, 91], easily guessable and insecurely stored passwords [46, 91], being forced to follow complicated password policies and change passwords regularly [46, 51, 87, 88], or phishing attacks [48]. Eventually, passwords are a prime example of the importance of *usable* security: The underlying problem is not the technical security of passwords but the overwhelming challenges they present to users that negatively impact authentication security. Replacing passwords with potentially more usable and secure alternatives has been a popular idea for many years [12]. However, alternatives such as two-factor authentication (2FA) and FIDO2 both lack adoption [7, 16, 21, 29, 53, 73], and passwords still remain the *de facto* standard.

Overall, companies, development teams, individual developers, and other involved stakeholders significantly impact the usable security of software systems in general and authentication in particular [36]. For example, they make decisions on password composition policies or the deployment of 2FA. To inform security-related software development decisions, including decisions on authentication deployments, developers draw heavily on online security advice, e.g., from Stack Overflow (SO) [2, 3, 5, 15, 24, 25, 26, 60, 98]. In the ACM CCS 2022 keynote, Michelle Mazurek concluded

that security advice is a “*disaster*” and affects developers and other stakeholders [56]. As developers draw on security online advice and online advice on authentication is common, the question of what developers can learn from such advice to make authentication deployments usable and secure was the main driver of this work. Based on a qualitative exploration of the *status quo* of online security advice on authentication deployments for developers, we aimed to develop further ideas for improving future advice. Previous work on general security advice for end-users [14, 20, 38, 43, 75, 77, 79] and for developers [2, 3, 5, 9, 10, 15, 24, 25, 26, 28, 34, 63, 78, 98] does not address online security advice for developers on implementing usable security for end-users. To the best of our knowledge, we are the first to explore this research gap.

Goals. In this paper, we aim to qualitatively explore developer security advice on implementing authentication that affects end-user usable security. We aim to understand the accessibility and quality of online advice and to provide insights into how online advice might contribute to (un)usable and (in)secure authentication. More specifically, we seek to explore which topics are covered and identify potential challenges developers might face when following advice on implementing authentication online. Based on our findings, we make recommendations for future advice to better support developers in implementing usable and secure authentication.

Approach. In our qualitative exploratory in-depth analysis, we applied the following multi-step process: First, we collected online documents that contained advice on authentication. Therefore, we recruited 18 professional web developers to collect a set of 264 online resources (Sect. 3) and their Google search queries. We used these queries to enrich the document corpus further. The overall document corpus consists of 406 documents on web authentication. Second, three researchers extracted all advice on web authentication from each document through qualitative open coding (Sect. 4.1). Finally, in an in-depth qualitative analysis, we explored advice distribution (Sect. 4.2) and topics (Sect. 4.3) and investigated characteristics like recommendable (Sect. 4.4), debatable (Sect. 4.5), and contradicting advice (Sect. 4.6).

Key Findings. The advice we found roughly reflected current (usable) security authentication challenges on the web, e.g., passwords are still ubiquitous, and password advice was most prevalent (51.6% of documents). For example, the advice covered password composition policies or password recovery. Moreover, only about 22.5% of advice was relevant to usable security of authentication. We classified 31.6% to be recommendable. However, we also found 18.0% to be debatable, e.g., being outdated, factually wrong, or counteracting usable security (e.g., enforcing regular password changes). We identified 15 groups of two or more pieces of advice that contradict each other. Overall, we found that advice from research results and official institutions was rarely integrated into the advice we analyzed. Furthermore, the advice was scattered across many online resources, making it difficult for developers to find.

Contributions. In this paper, we make the following contributions: (i) To the best of our knowledge, this is the first study investigating online advice for developers on authentication that affects security and usability for end-users. (ii) We present an overview of authentication advice topics and further explore recommendable, debatable, and contradicting advice. (iii) Based on our findings,

we make recommendations for future usable security advice for developers. Furthermore, we highlight open challenges and outline future research. (iv) Finally, we provide our document corpus, all pieces of advice, and additional supplementary material in a replication package (Appendix A).

2 RELATED WORK

We discuss related work in three key areas: (i) general security advice for end-users, (ii) (usable) security advice for developers, and (iii) usable and secure web authentication and its challenges. We highlight our novel contributions and put it into context.

General End-User Security Advice. A large portion of online advice on security addresses end-users [14, 20, 39, 43, 59, 75, 77], e.g., concerning topics like how to handle passwords or spam. Non-experts, like most end-users, heavily request such advice [38]. As lots of advice exists, especially online, the selection of advice sources has to be considered [75]. For example, factors like the source’s perceived trustworthiness [75] and convenience [20] influence advice adoption. One explanation for low advice adoption is that users rationally reject advice as it often offers a poor cost–benefit trade-off [39]. Overall, security advice can be challenging to adopt, as a mismatch between advice and behavior was perceived among both experts and non-experts [14, 43]. Even security experts were found to lack consensus for advice [79]. Similarly, users and experts (like administrators) struggle with advice prioritization [77].

While advice can be used to educate end-users, we argue that it is (at least equally) important to help software professionals design systems that people can use securely—ideally without users needing any advice. Therefore, we focus on advice for software developers.

(Usable) Security Advice for Developers. In recent years, researching human factors and security with a focus on developers has gained popularity [4, 35, 74, 93]. This is also true for research on security advice for developers. Several studies found that advice impacts security and that sources containing insecure code find their way into software [2, 3, 5]. For example, Acar et al. found that developers who used SO produced significantly less secure code and that only 17% of SO posts contained secure code snippets. Insecure code from SO is also prevalent in many Android apps [24] and contained in top Google search results [25]. However, nudging [26] and re-ranking search results [25] can create a positive security impact of SO. Overall, SO is a place for software professionals to discuss security-relevant topics [98], while secure and insecure answers were almost balanced [15]. Apart from SO, researchers also found challenges in other advice sources, e.g., not actionable IoT advice [9, 10].

Advice can also be given through other channels. This includes IDE plug-ins that aid developers in preventing security pitfalls. For example, Nguyen et al. presented and evaluated a plug-in that supports Android developers [63]. Similar tools exist for other use cases, e.g., secure password storage [28]. Besides such tools, Gorski et al. found that integrating security advice into application programming interfaces (APIs) can significantly improve code security without affecting API usability [34]. Other advice forms can be more general by outlining principles (e.g., Microsoft’s *NEAT* design principles for security warnings [78]) and abstract guidance (e.g., heuristics for accessibility and usable security on the web [61]).

Most similar to our work, Iacono, Gorski, et al. systematized security principles, guidelines, and patterns based on scientific literature [33, 52], but only “*concentrated on principles and patterns for end-users*” [33]. In contrast, we focus especially on advice for developers that affects end-users’ usable security. Moreover, we focus specifically on (non-scientific) online resources. To the best of our knowledge, no paper examined advice for developers that covers end-user usable security. We argue that end-user (usable) security, however, highly depends on decisions and implementation during software development. Our analysis addresses this gap.

Usable and Secure Web Authentication. Passwords, probably the oldest authentication approach [13] and a starting point for early usable security research [6], are the most used authentication approach despite many shortcomings. This includes but is not limited to insecure storage [46, 91], password reuse [31, 72, 91], and usability and (resulting) security problems through password policies [46, 51, 86, 87, 88]. To aid in creating secure passwords, password strength meters are a common tool and visualization technique [95]. However, their construction is difficult. For example, researchers found inconsistencies, weaknesses [19], and accuracy issues [30, 96]. Similarly, password managers (PWMs) should help users deal with their passwords, e.g., by creating secure (random) passwords [99] that fit websites’ password policies [27]. However, even PWMs have interaction problems with websites [40] or overwhelm their users [66], both limiting their usability. Due to passwords’ numerous problems, the idea of replacing them became popular. However, many approaches failed [12], and replacing passwords remains an unsolved challenge.

For example, replacing passwords could be achieved with the FIDO2 standard [23]. However, FIDO2 lacks adoption, which might be caused by overlooked benefits [21], concerns, and not trusting the new technology despite good usability [29]. Despite generally good usability, users struggle with hardware token setup [17, 82] or rate the usability of smartphones for FIDO2 as low [69]. As tokens can get lost and are associated with costs, electronic IDs can be used for FIDO2 [85]—albeit being hard to set up [44].

2FA, or generally multi-factor authentication (MFA), is an important concept for strengthening security—especially when using passwords—but could also negatively affect usability [1, 80, 81]. Users might have problems with the initial setup [80], and the 2FA user experience (UX) is inconsistent [53]. Overall, 2FA adoption is still relatively low and needs to be increased [7, 16, 73, 76]. Widely adopted approaches like SMS-based 2FA are known to be vulnerable to phishing [54].

To conclude, authentication has pitfalls that software developers need to consider. Interestingly, most outlined challenges are no technical security problems and boil down to a usability problem. For example, passwords would be more secure if used correctly. Therefore, we investigate how developer advice supports usable and secure authentication.

3 COLLECTING ONLINE DEVELOPER ADVICE

In this section, we describe our data collection of 406 online resources on authentication. While we aimed to create an ecologically valid dataset, our main goal was diversity due to the exploratory nature of our work. In a developer survey between August and

November 2021, 18 professional web developers searched and reported online resources on web authentication they considered for an imaginary web authentication development task. We extend the document corpus further by including search results for our participants’ search queries. As we had already reached saturation with this document corpus, we stopped recruiting more participants.

3.1 Developer Survey

Our goal was to create a comprehensive corpus of documents containing online advice regarding usable and secure web authentication for developers that impacts usable security for end-users. We asked professional software developers to search for online resources they consider relevant for implementing usable and secure web authentication. We gave them a scenario and task to search for relevant advice on the web.

3.1.1 Scenario. We created a task containing a web development scenario and framed it towards authentication and usable security. We aimed to collect document URLs, search queries, and other information from software developers.¹ Our scenario consisted of a hypothetical new web-based health platform *KeepYourHealth*, that should become the central pivot point for personal health, e.g., storing and sharing medical data or communicating with doctors. We chose the healthcare framing for three reasons: First, medical and personal health information is considered highly sensitive, hopefully shifting the participants’ attention to security. Second, to highlight the need for good usability, medical applications have a diverse user base that likely contains many non-tech-savvy users. Third, medical applications must follow certain compliance or data protection rules (e.g., HIPAA in the US, PIPEDA in Canada, GDPR in Europe). While this might skew the results slightly towards official documents, we also hoped to receive a more diverse and complete document sample.

As part of the *KeepYourHealth* project, we wanted participants to imagine that they are responsible for designing and implementing the authentication mechanism for both a mobile and a desktop version. To create the concept for *KeepYourHealth*’s authentication system (including registration and login), the participants had to perform online research and look for helpful resources. We designed the scenario to be very open to ensure diverse results, as participants can explore various topics in their search. Moreover, we instructed participants to freely consider any resource, authentication approach, technology, etc. Our only requirement was that the authentication system had to be usable and secure. We used words like *easy-to-use*, *user-friendly*, *secure*, *usability*, and *security* to emphasize the requirement for usable security.

3.1.2 Search Task. To collect online advice, we gave software professionals the task of searching for online resources that provide advice that potentially helps to build a usable and secure web authentication system for *KeepYourHealth*. We explained our research project’s purpose and that we wanted to collect and analyze advice from various online sources. We additionally explained that we, as researchers, are biased and probably would search differently than web development experts. We also wanted the software developers to report how they found a specific online source, i.e., the

¹See this paper’s extended version for the full scenario description and search task.

search query. We provided the participants with a table to report documents and search queries. The table consisted of the following columns: URL, how the document was found (e.g., bookmark, link within another site, search engine), and (if applicable) the search query that yielded the document. Each column's purpose was also explained. Our participants were free to choose any search approach they wanted, i.e., we did not constrain them to a specific search engine.

3.1.3 Questionnaire. We embedded the search task in an online questionnaire. After an initial consent form, the participants were introduced to the scenario and their tasks. We gave them 30 minutes for the task. For the search task, we displayed the above-mentioned table for the users to enter URLs and search queries. After completing the task, we asked the participants for their most used search engine and to rank different information resource types by their helpfulness. Next, we followed up with standard demographic questions, including age, gender, country of residence, highest level of education, and employment status. Additionally, we asked some questions specific to this study's context regarding participants' experience, skills, and security and usability education. We estimated the maximum time for the whole survey at 45 minutes. The replication package (Appendix A) includes the full questionnaire.

3.1.4 Instrument Development & Piloting. We developed and piloted the questionnaire, scenario, and task in multiple iterations. We asked nine persons to pilot the study and questionnaire, including researchers with experience conducting usable security studies and experiments. We made smaller adjustments regarding wording and description texts based on the feedback that we collected via think-aloud and feedback text boxes on each survey page. Finally, we tested the whole setup with three professional developers until no further adjustments were necessary. During piloting, we collected 48 documents and 38 search queries. We included those in our final document corpus to enhance its diversity further.

3.1.5 Ethics & Data Protection. We received ethical review board (ERB) approval for this study from one of the involved institutions, concluding that there were no ethical concerns. The remaining institutions had neither an institutional review board (IRB) nor an ERB that applied to our study. This study was designed to adhere to the ethical research principles of the *Menlo Report* [45]. We further adhered to the strict German and EU privacy laws. All participants consented to collecting, storing, and analyzing their data for this research. We clarified that we would only evaluate and publish de-identified and aggregated data. Furthermore, we informed all participants that they could terminate the study at any time.

3.1.6 Recruiting. We aimed to recruit a diverse sample of software developers with web development experience. Therefore, we used the freelancer platform Upwork. We created an Upwork job post describing our study and communicated the idea of our research. Participants were required to have sufficient proficiency in English. Besides that, we checked the Upworkers' profiles and job history to ensure they have web development experience. We also directly invited Upworkers who had indicated experience in web development in their profiles. The job was rewarded with \$45 (hourly wage of \$60), above Upwork's \$15–\$30 median hourly pay for software developers [94] to offer competitive compensation.

Table 1: Demographics of the 18 web developers.

Gender					
Male	16	88.9%	Female	2	11.1%
Age [years]					
Min.	19.0		Max.	45.0	
Mean (Std.)	28.4	±7.9	Median	25.5	
Highest Education Level					
< High school	1	5.6%	Bachelor's degree	7	38.9%
High school	4	22.2%	Attending grad. school	1	5.6%
College	2	11.1%	Master's degree	3	16.7%
Country of Residency					
United Kingdom	4	22.2%	Brazil	2	11.1%
United States	3	16.7%	Other ¹	7	38.9%
India	2	11.1%			
Industry Experience [years]					
Min.	1.0		Max.	22.0	
Mean (Std.)	6.1	±6.0	Median	3.8	
Employment Status[†]					
Employed full-time	10	55.6%	Student	2	11.1%
Self-employed/Freelancer	9	50.0%	Self-describe	1	5.6%
Employed part-time	2	11.1%			
Domain of Web Experience[†]					
Full-stack developer	15	83.3%	UI/UX	6	33.3%
Frontend developer	11	61.1%	API developer	6	33.3%
Backend developer	9	50.0%	Testing	4	22.2%
Web design	7	38.9%	Other	1	5.6%
Security/Usability as Part of Training/Studies					
Security	8	44.4%	Usability	5	27.8%
Security/Usability Training in Professional Activity					
Security	7	38.9%	Usability	7	38.9%

[†] Multiple answers allowed; may not sum to 100%. ¹ Each country occurring once.

3.1.7 Demographics. We recruited 18 web developers via Upwork whose demographics are comparable with those of professional developers from SO's latest developer survey [89]. An overview of participants' demographics can be found in Table 1. The participants were from 11 different countries, with the UK and the U.S. being the most common, including others such as Canada, Nepal, Romania, Germany, Nigeria, France, and Turkey. The majority had at least a bachelor's degree. Their professional software development experience varied between 1.0 and 22.0 years. Participants covered various roles, such as software engineers, web developers, architects, team leads, and CTOs. Most participants reported some experience with usable security. Some participants also received education in this area in their training, studies, or professional career.

3.1.8 Reported Documents and Search Queries. The participants overall found 264 distinct documents (284 with duplicates reported by multiple participants) across 171 different website domains. This already indicates that advice is contained in many resources from different authors. However, some domains were reported more frequently, as shown in Table 2a. This mainly includes authentication platforms like *Auth0*, *Okta*, *swoopnow.com*, but also blogs and tutorial sites like *medium.com* and *dzone.com*, or documentation like the one provided by *Google* or the library *next-auth*. The Open Web Application Security Project (OWASP)'s *Cheat Sheet Series* is included.

We collected 145 distinct search queries from our participants. All participants reported having used *Google* for their web searches. The most frequent terms and term bigrams are shown in Tables 2b and 2c. While *authentication* is the most frequent term, participants

searched specifically for security aspects, *best practices*, and usability aspects. This also indicates that our task worked as intended, as the terms reflect the goal of usable and secure authentication.

Summary: Initial Document Collection. The search task with 18 professional web developers resulted in 264 documents and 145 search queries. Both document domains and search queries show high diversity.

3.2 Search Expansion

We extended our document corpus with search results for the collected search queries that participants reported in addition to the directly reported documents. We performed this search expansion because participants might have missed relevant documents to counterbalance participants' personalized search results. Moreover, this increases overall diversity and helps to reach saturation in the best case, while adding no additional insights in the worst case. For each query, we obtained the corresponding search results to expand our document corpus. Therefore, we leveraged Google's *Custom Search API* [32] to execute a search for each query and extracted its top 10 results, i.e., its URL and some metadata. We chose this Google-based engine, as Google is the most used search engine [62, 90] and all participants reported that they use Google as one of their main search engines. We chose 10 results per query, as this corresponds to the first page of search results. Moreover, the click-through rate (CTR) drops below 10% after the fourth and quickly decreases further to almost 0% for lower-ranked results [70]. That said, almost no user looks at the second page of results [18]. A top-10 expansion, therefore, covers search results a user potentially clicks while also including a safety margin.

3.3 Final Document Collection

The search task with professional software developers resulted in 264 distinct documents and 145 distinct search queries. As our goal for this qualitative study is diversity, we decided also to include all 48 documents and 38 search queries from piloting. As described above, we extend our corpus with documents found using Google. We randomly sampled 100 search results from the search expansion and included those in the final document corpus. Combining all these documents, we obtained a final dataset of 406 documents overall. For the documents that indicate a publishing date, the median year for the last update was 2021 (oldest: 2002, newest: 2021). We downloaded each as a PDF for our analysis.

We stopped after these 406 documents, as we had already reached saturation, which means that we would not gain new insights from analyzing more documents and advice, i.e., no new themes emerged in Sect. 4 (e.g., advice topics). Repeated advice in multiple documents also indicated saturation. Consequently, we did not survey further participants.

In the survey, we asked our participants about the helpfulness of different information sources (cf. Figure 1). Overall, participants reported a high affinity towards online resources. The top-three sources were official documentation, blog/news articles, and online communities. Similarly, this is also reflected in the different types of documents (Table 3). We found blog posts (some including news) to be the most common type, followed by documentation and wikis and project websites (including source code repositories).

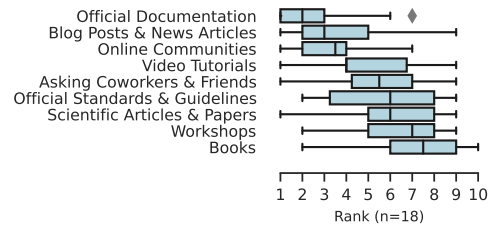


Figure 1: Participants' ranking of information sources for the question: "Which advice or information sources do you find most helpful in general?". From most helpful (1) to least helpful (10), excluding "Other" answers.

Surprisingly, only 15 documents (3.7%) were from Q&A forums, 13 of them from SO (cf. discussion in Sect. 5.4). Minor document types included educational material and courses as well as video and audio resources. 6 documents had no clear type. 4 (1.0%) were not available online anymore when we downloaded the documents.

Summary: Document Corpus. Our final document corpus consisted of 406 distinct documents which we downloaded for analysis. Blog posts, documentation, and wiki pages were the most common document types.

3.4 Limitations

This work has some limitations that are typical for this type of study. Generally, the developer survey is subject to self-reporting, social desirability, under- and over-reporting, and sampling biases. We note that our document corpus cannot contain all online resources and that it is influenced by the participants' experience and especially the personalization of their search results. We addressed this by recruiting a diverse set of participants and extending the document corpus with search results until we reached saturation. Furthermore, we received only English documents as it was the study's language. We decided that English as the only language is an acceptable trade-off as it is the main language used in software development (e.g., documentation or SO). The search task and scenario (Sect. 3.1) itself could bias the document corpus. However, we favored a realistic scenario without many requirements for external validity over a general, undirected search task.

4 ADVICE ANALYSIS

In this section, we provide details on the advice analysis and its results. Our analysis is a multi-stage process that starts with identifying and extracting advice (Sect. 4.1) before analyzing it in detail. In the following subsections, we describe the advice distribution (Sect. 4.2), its topics (Sect. 4.3), and occurrence of recommendable (Sect. 4.4), debatable (Sect. 4.5), and contradicting advice (Sect. 4.6).

Throughout the paper, we report counts, e.g., counts of advice or documents. As this is qualitative work, those counts are used to convey weight and should *not* be interpreted as quantitative statistical results. While we try to give an overview by discussing advice examples, we cannot describe and discuss every single piece of advice due to their high number.

Table 2: Most frequent domains and search query terms and bigrams that participants reported.

(a) Top-10 documents domains.		(b) Top-10 terms from search queries.		(c) Top-10 bigrams from search queries.	
Domain	#	Search Term	#	Search Term Bigram	#
auth0.com	23	authentication	72	web authentication	14
medium.com	16	web	39	authentication methods	10
www.okta.com	8	best	15	for web	8
swoopnow.com	6	security	11	best practices	8
blog.jscrambler.com	6	secure	11	web app	8
cheatsheetseries.owasp.org	5	methods	10	web applications	7
cloud.google.com	5	user	10	user authentication	6
dzone.com	5	app	9	authentication best	5
developers.google.com	4	practices	8	methods for	4
next-auth.js.org	4	applications	7	secure authentication	4

Table 3: Types of the reported documents.

Type	#	%
Blogs (including news)	219	53.9%
Documentation & Wikis	79	19.5%
Project Websites & Code Repositories	41	10.1%
Scientific	35	8.6%
Q&A Forums	15	3.7%
Educational	4	1.0%
Video/Audio	3	0.7%
Undefined Type	6	1.5%
Not Available	4	1.0%

4.1 Identifying & Extracting Advice

Below, we describe our advice identification and extraction approach in detail. It consists of two steps: (1) First, we extracted and systematized each piece of advice from all documents through qualitative coding. (2) Second, we applied inclusion and exclusion criteria to reduce the set of advice to address usable and secure web authentication relevant to our study.

4.1.1 Advice Extraction. We extracted each advice for developers by reading all documents and manually assigning codes that summarize each piece of advice. We used the following definitions to detect advice. The definitions' goal was to make the extraction process transparent, objective, and reproducible.

- **(Concrete) Advice:** We considered explicit prompts to do something in a specific way or specific recommendations concrete advice. Phrases like “we recommend,” “a developer should,” or imperatives like “ensure” often indicate advice, which the researchers therefore used as proxies to detect advice. An example of such a concrete recommendation is: “Allow users to toggle password visibility when typing it.”
- **Descriptive Information:** We deliberately refrained from extracting information as advice that was not explicitly directed at the reader or any other form of specific recommended action. Some document passages just inform and describe, leave room for interpretation, and require reading between the lines. It might be unclear whether the resource's author wanted to give advice, whether a reader perceives it to be advice, and how one might interpret it.

Three researchers each annotated a third of the final dataset in an iterative open coding process. As advice in the documents can be long and wordy, we summarized it by creating codes that briefly paraphrase the advice [47, 55]. Table 4 shows examples of original document excerpts and our assigned advice codes. As shown in

Table 4: Examples of document excerpts with concrete advice and assigned advice codes.

Text in Document & Assigned Advice Code
“As a developer yourself, it's your responsibility to provide users with the best and most secure experience possible.” → Ensure authentication is secure and has good UX
“Make sure to explain each [2FA] step and keep in mind, that not everybody has the same technical knowledge. For example, HubSpot added a description of Google Authenticator when the user would choose that option.” → Explain each 2FA step
“Consider Password-less Authentication [...] Some applications are good candidates for a new approach: eliminating the password altogether. [...] users [...] forget their credentials, password-less authentication is an compelling option.” → Use password alternatives
“If your organisation wants to take its login game to the next level, you should definitely consider password alternatives.” → Use password alternatives
“Send the user an email informing them that their password has been reset.” → Notify users via email when password was reset

the table, different text passages can have the same meaning and were therefore assigned the same advice code (examples 3 and 4). We used these advice codes for our analyses throughout the paper. Each researcher grouped similar or related paraphrases in a code system. The leaf nodes in the code tree contain the advice. The inner nodes are used to build a hierarchy (e.g., group all advice on passwords). After coding sets of 50 documents, at least two coders merged, refined, and regrouped the code system if necessary; all three reviewed the resulting code system. The coders repeated the described process eight times until advice from all 406 documents was extracted. In accordance with McDonald et al. [58], we refrain from reporting inter-rater reliability (IRR) because extracting only concrete advice leaves little room for the researchers' own interpretations and because the following analysis of the advice (following subsections) is our main contribution—not the code system itself.

As we focus on non-scientific online resources, we excluded 31 (7.6%) documents being scientific papers; 4 (1.0%) documents because they contained only a video or audio guide; another 29 (7.1%) documents were not related to our research topic (i.e., not about authentication and usable security); and two documents because they were not available anymore or a duplicate of an existing document. 119 documents contained no advice at all. For the remaining 221 documents, we identified 1,208 distinct pieces of advice.

4.1.2 Inclusion and Exclusion Criteria. In the previous step, we extracted all advice for developers and obtained 1,208 paraphrased pieces of advice. We defined inclusion and exclusion criteria to focus only on advice relevant to this paper's scope. We excluded advice if any of the following criteria were met: (1) if the advice

was not related to security, (2) if the advice was not related to usability aspects and end-users, or (3) if the advice did not cover authentication. The remaining advice was included as it is relevant for usable security of web authentication. Three researchers did this assessment independently and later resolved conflicts until they reached consensus.

Advice Inclusion. 272 distinct pieces of advice out of the initial 1,208 covered usable security for web authentication and were therefore included for further analysis. For example, the advice to “Use a minimum of 8 characters” for passwords matched our inclusion criteria because passwords are a security and authentication feature, and their length affects usability [46, 86, 87, 88]. Overall, the 272 pieces of advice matching the inclusion criteria occurred 655 times across 124 different documents. Hence, only a small fraction of the online advice considers or affects usable security. Instead, most advice is only about security—not considering usability and human factors of security at all.

Advice Exclusion. According to the above-mentioned criteria, we excluded 936 pieces of advice from further analysis because they did not cover usable security of web authentication, thus not being relevant for our research goals for one or multiple of the following reasons. We excluded 257 distinct pieces that did not cover security, e.g., “Provide live form validation” or “Only use single-column forms”. Similarly, we excluded 669 distinct pieces of advice that were not relevant for end-users or usability. For example, “Use object-relational mapping (ORM) to prevent SQL injections” or “Use bearer tokens,” are technical security details that users do not notice. We excluded another 73 pieces of advice that covered usability and security but were not relevant to web authentication, e.g., “Limit upload size for users”. We note that we focussed our analysis on advice in prose text and excluded any code snippets (occurring in 102 documents).

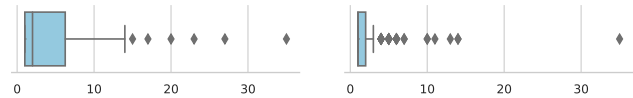
Summary: Identifying & Extracting Advice. We found and extracted 1,208 pieces of advice in 406 documents. Focusing on web authentication advice with relation to usable security, we came up with 272 pieces of advice from 124 documents for further analysis.

4.2 Advice Distribution & Density

As already stated, the advice was widely distributed across 124 documents. A document contained 4.5 different pieces of advice on average (median: 2.0). However, the distribution was overall skewed (cf. Figure 2a): while the document with most advice contained 35 different pieces of advice, the median was only 2.0 different pieces of advice per document. That said, most documents provided only little advice, while a few were much denser. This was similar for the domains on which advice is provided. While individual domains (that contain advice at all) provided on average and median only 6.6 and 3.0 different pieces of advice, respectively, the OWASP Cheat Sheets contained most, having 55 different advice imperatives (see Table 5). Considering that *mobile-security.gitbook.io* by OWASP was also among the top domains, OWASP resources contained by far the most advice. However, authentication provider *swoopnow.com*'s blog and the developer portals *web.dev* and *dzone.com* also contained a considerable amount of advice.

Table 5: Top-5 domains containing most distinct advice.

Domain	#Advice
cheatsheetseries.owasp.org	55
swoopnow.com	27
mobile-security.gitbook.io	27
web.dev	23
dzone.com	18



(a) Number of pieces of advice per document. (b) Occurrence frequency of advice (in number of documents).

Figure 2: Boxplots that describe the advice distribution.

Table 6: Top-10 most frequent pieces of advice measured by occurrence in documents.

Advice	Docs
Offer/Use 2FA/MFA.	35
Enforce a password policy.	14
Allow/Use 3rd-party SSO.	14
Use generic responses for auth. errors to not weaken security.	13
Require strong passwords.	11
Require re-authentication for sensitive actions (step-up auth).	10
Reject most often used passwords and words from dictionaries.	10
Limit the number of login attempts.	7
Protect against brute-force attacks by rate limiting.	7
Use password alternatives.	6

While most documents contained little advice, developers are not limited to considering only a single document. In line with that, our participants found 26.3 different pieces of advice on average among multiple documents (median: 13.0). 81 different pieces of advice was the highest number of different advice imperatives found by a participant. Looking into multiple information sources is also likely to yield new advice because 176 of 272 pieces of advice were only included in one document (cf. Figure 2b). On average, a piece of advice was only contained in 2.1 documents (median: 1.0). However, a few pieces of advice were covered in a higher number of documents, as shown in Table 6. *Offer/Use 2FA/MFA* was the most frequent one, being contained in 35 documents.

All in all, advice is scattered, and most documents contained little advice, with a few exceptions. Hence, developers must consider multiple documents if they want to get a comprehensive set of advice. Looking at a single document would yield only a small subset of advice if not looking at one of the rare dense documents.

Summary: Advice Distribution & Density. While documents contained 4.5 pieces of advice on average, most only contained two or less. However, some documents like *OWASP Cheat Sheets* were denser. 176 of 272 pieces of advice were only covered by a single document.

4.3 Topics of Web Authentication Advice

Early on, we noticed that the advice covered various topics. For further investigation, we took a closer look and identified 12 different topics in our dataset’s advice.

4.3.1 Identifying Topics. Three researchers independently used an inductive coding strategy to identify advice topics within the 272 pieces of advice and assign each advice paraphrase to its respective topic(s). Afterward, the researchers discussed each individual topic in a consensus session to create the final list of topics. It consists of 12 different topics (cf. Table 7). We assigned each topic a definition to achieve a reproducible assignment for each piece of advice and describe the topic. With the final list of topics, each coder independently coded all 272 pieces of advice. After the independent coding, the coders finally compared and resolved existing conflicts until they reached consensus. We allowed assigning multiple topics if a piece of advice fits several topics at the same time. For example, the advice to “Use password strength meters” covers *passwords* and also implies *user interface (UI)/UX* requirements.

4.3.2 Results. Overall, we found 12 different advice topics, as depicted in Table 7 (including each topic’s definition, example advice, and prevalence counts). Within the 12 topics, we found five topics corresponding to major authentication types: *passwords*, *2FA/MFA*, *single sign-on (SSO)/3rd-party login*, *token-based authentication*, and *biometrics*. Besides these five, some minor authentication types emerged that we grouped into an *other* category (e.g., TLS client certificates). Advice on *passwords* forms the most common authentication type, with 101 distinct pieces of advice across 64 documents. The topic covers several aspects, like password policies, reset, entry, managers, or strength meters, and includes usability aspects, e.g., toggling visibility in password fields. Advice on 2FA and MFA is also popular (54 documents), followed by advice on SSO, third-party, and social logins. Token-based and biometric authentication were only rarely discussed.

The remaining six topics were more general and not about a specific authentication type. Usage recommendations, i.e., what authentication mechanism to use, were very common (83 documents). Another topic spanning all authentication types was session and account management (e.g., account recovery). The remaining minor topics were about user and usability aspects. They covered general authentication UI and UX, how to interact with the user through notifications and messages, or how to educate users.

The co-occurrence heatmap of topics within a document in Figure 3 reveals that authentication types are often not jointly covered within the same document. For example, while password advice is in 64 documents, only in 28 documents both passwords and 2FA/MFA are discussed together. This is similar for passwords and SSO/3rd-party login. On average, a document touches 3.2 different topics (median: 3.0). However, when just focusing on the authentication type topic subset, only 1.7 different authentication types are discussed. 61 documents cover only one authentication type.

Summary: Advice Topics. Overall, we found a wide range of 12 different advice topics. Advice on passwords was by far the most frequent, followed by 2FA/MFA. The topics are often discussed in isolation, e.g., not discussing security benefits of 2FA when talking about passwords.

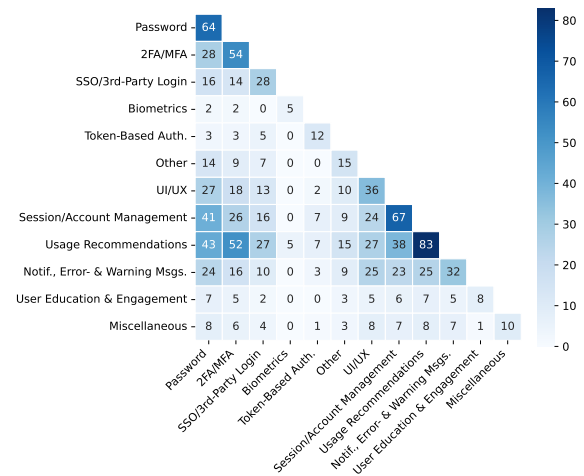


Figure 3: Co-occurrence of the topics from Table 7 within the same document.

4.4 Recommendable Advice

Below, we provide details on advice we deem recommendable for developers. We consider authentication advice recommendable if it contributes to usable and secure authentication on the web.

4.4.1 Identifying Recommendable Advice. We based our assessment on academic literature and official guidelines. In addition to references from related work, we searched for literature by checking the proceedings of top-tier venues, including IEEE S&P, USENIX Security, ACM CCS, NDSS, USENIX SOUPS, ACM CHI, and official resources (e.g., from U.S. National Institute of Standards and Technology (NIST) [64]). This was complemented by searching on Google Scholar and a general online search. We also leveraged *forward* and *backward snowballing* to find further resources [97], i.e., resources citing or cited by an already obtained resource.

Based on these literature references, two authors assessed for each advice whether it could be considered recommendable. We used discussions to obtain consensus decisions. For transparency, we provide supporting references with their publication years to indicate their age. Additionally, we provide written rationales that explain our reasoning within our replication package (Appendix A).

4.4.2 Results. We considered 86 of 272 pieces of advice (31.6%) recommendable. Overall, 92 documents (74.2%) contained recommendable advice. Table 8 provides an overview of the most frequent recommendable advice. Comparing this to the overall most frequent advice (Table 6) shows some minor differences. For example, we did not consider the advice to “Require strong passwords” recommendable because it does not provide concrete guidance. However, we consider the related advice to “Enforce a password policy” recommendable since official institutions generally recommend password policies [64 (2017)], and there is scientific support for the use of password policies (while the concrete policy should be adapted to the context) [46 (2011), 51 (2022)].

Among the advice that is only contained in one or two documents, we found highly recommendable advice for developers. For

Table 7: Categorization of advice in different topics.

Advice Topic	Definition	Advice Example	# ¹	#Adv. ²	#Docs ³
Authentication Type[†]					
Passwords	Includes all advice, which directly mentioned a policy, a feature, or an implementation detail, that is directly or indirectly related to the creation, change, or processing of passwords.	<ul style="list-style-type: none"> Enforce a minimum password length of 8 characters. Reject passwords from breaches. 	264	101	64
2FA/MFA	Advice related to two-factor or in general multi-factor authentication and the different approaches but also backup and recovery and general considerations.	<ul style="list-style-type: none"> Require adding a 2nd factor for features with higher security. Provide OTP as 2FA. 	87	36	54
SSO/3rd-Party Login	Recommendations related to the implementation or use of SSO, 3rd-party authentication services, or social logins, e.g., via Facebook or Google.	<ul style="list-style-type: none"> Use OAuth 2.0 Stick to a few common social/third-party providers. 	50	20	28
Token-Based Authentication	Authentication mechanisms based on tokens, their design, and usage. This also includes token formats such as Bearer or JWT. Cookie- and session-based are authentication also included.	<ul style="list-style-type: none"> Give tokens an expiration. Limit number of token usages. 	23	16	12
Biometrics	Includes all advice which directly is related to any kind of biometric authentication method.	<ul style="list-style-type: none"> Prefer fingerprint scanning above facial recognition. Use a traditional mechanism (e.g., password) beside biometrics. 	11	8	5
Other	Advice related to other authentication types, not belonging to one of the other authentication types.	<ul style="list-style-type: none"> Do not use TLS Client Authentication for public websites. Use passwordless authentication. 	26	10	14
Usage Recommendations	Advice on how to choose or which authentication method to use.	<ul style="list-style-type: none"> Offer/Use 2FA/MFA. Use risk-based auth controls. 	188	71	83
Session/Account Management	Advice on management of sessions and accounts, including login and logout process, account creation/registration, recovery, blocking, user identifiers, etc.	<ul style="list-style-type: none"> Require re-authentication for sensitive actions (step-up auth). Invalidate all sessions after password change. 	225	103	63
UI/UX	Concrete request to change, integrate or implement features and elements in the UI in a specific way, e.g., how to design a password field. Also includes advice that directly mentions some aspect of user experience.	<ul style="list-style-type: none"> Use generic response code for auth/reg/recovery errors Don't prohibit pasting passwords. 	104	50	35
Notifications, Error-, & Warning Messages	Advice on informing the user through notifications and messages in general, their design and usage, including push notifications, email, or SMS/text message, but also warnings, error messages.	<ul style="list-style-type: none"> Notify users via mail when password was reset. Avoid sending sensitive data to push notifications / backups / logs 	59	27	27
User Education & Engagement	Advice on engaging or educating users on different topics of authentication.	<ul style="list-style-type: none"> Educate users to create strong passwords. Encourage users to accept passwords from password managers. 	12	7	8
Miscellaneous	Includes all advice that does not fit into any other category.	<ul style="list-style-type: none"> Use captchas after failed attempts. Involve stakeholders early on for security requirements. 	18	9	10

[†] Container for sub-topics; containing no advice. ¹ Advice occurrences in a topic. ² Distinct pieces of advice in a topic. ³ Documents containing advice from a topic.

Table 8: Top-10 most frequent recommendable advice measured by occurrence in documents.

Advice	Docs
Offer/Use 2FA/MFA.	35
Enforce a password policy.	14
Allow/Use 3rd-party SSO.	14
Require re-authentication for sensitive actions (step-up auth).	10
Reject most often used passwords and words from dictionaries.	10
Limit the number of login attempts.	7
Protect against brute-force attacks by rate limiting.	7
Provide secure password recovery mechanism	6
Enforce minimum password length	6
Use password strength meters.	6

example, “Don’t prohibit pasting passwords” (to not interfere with PWMs [40 (2021)]) or to “Provide a password change feature.” The latter is crucial to change compromised passwords, but to our surprise only mentioned in two documents—maybe because it is too obvious. Advice like “Don’t force users to regularly change passwords” is recommendable, as it replaces the outdated advice to enforce regular password changes [37 (2018), 64 (2017)]. Other advice also gives recommendations on specific pitfalls to avoid. For example, to not use TLS client certificates for public websites is recommendable because TLS client certificates are hard to use [71 (2014)].

Summary: Recommendable Advice. We found 31.6% of the advice to be generally recommendable to developers, as it adheres to state-of-the-art research results or official guidelines. Additionally, we consider 14.7% of advice potentially recommendable that lacked references for evidence.

4.5 Debatable Advice

Besides recommendable advice, we considered other advice to be debatable whether developers should follow it. This includes if it contributes to insecure or unusable authentication on the web, is inconsistent with literature references, is only applicable in certain use cases, or different opinions on the advice exist.

4.5.1 Identifying Debatable Advice. We identified debatable advice as we did for recommendable advice (cf. Sect. 4.4.1). Using the same references, two researchers assessed the advice and discussed debatable advice in a consensus session. In cases without a supporting reference, we only tagged advice as debatable if a clear rationale existed. For traceability, we summarized the rationals in the extended version. All 49 debatable pieces of advice are available in the replication package (Appendix A).

4.5.2 Results. We considered 49 of 272 pieces of advice debatable. Overall, we found debatable advice in 39 documents (31.5%). We grouped similar debatable advice in 14 groups (referred to as D_x) and reported it along with the rationale why it is debatable and examples in the extended version.

Most debatable advice covered password policies (D1) and was discussed in 22 documents. D1 is a good example of outdated advice. Outdated advice might have been considered correct in the past, but in the meantime, new insights emerged that render it incorrect. For example, we found advice to enforce complex composition rules, e.g., passwords should consist of lower- and uppercase letters, numbers, and symbols. Another example is to “Force users to change

passwords regularly.” However, research uncovered its negative effect on authentication usability and security [6 (1999), 37 (2018), 41 (2010), 46 (2011)]. Likewise, official organizations such as NIST changed to less restrictive password policies in 2017 to improve usability and security compared to previous password policies [64 (2017)]. However, we also found up-to-date advice that is less strict and only suggests enforcing a minimal password length, as also recommended by OWASP [67 (2021)].

We found more outdated advice recommends the use of security questions (D3), despite their usability [11 (2015)] and serious security shortcomings [12 (2012), 84 (2009)], and official guidelines advising against them [64 (2017)]. We found other advice that harms (usable) security, such as not allowing to toggle masking in password fields and viewing the password in plain text (D11), automatically terminating all sessions on password change without asking the user (D6), or stating if a user entered an old but not their current password (D7). Some advice is more debatable, and its appropriateness depends on the use case. For example, consider the advice to use text messages (SMS) for authentication purposes (D5). On the one hand, text messages have drawbacks like the possibility of losing access to a phone number [50 (2021), 57 (2021)], SIM swapping attacks [49 (2020)], and being interceptable [65 (2016)]. On the other hand, text messages are convenient as smartphones are ubiquitous. Similarly, PGP-encrypted emails for sending reset tokens (D12) provide good security, but PGP has only poor adoption [92 (2022)]. Some advice is challenging to assess, like fingerprints *always* being preferable to facial recognition (D13), while vendors like Apple changed from fingerprint to *Face ID* [8 (2017)].

Summary: Debatable Advice. In 31.5% of the documents, we found debatable advice that conflicts with research findings and official guidelines. The debatable advice also contains instances of outdated advice, wrong advice, and advice that harms usable security.

4.6 Contradicting Advice

Besides recommendable (Sect. 4.4) and debatable advice (Sect. 4.5), i.e., advice in line or in conflict with literature references, we found several examples of contradicting advice within the document corpus itself. To analyze those, we identified sets of advice that contradict each other (e.g., “Do X” and “Don’t do X”). We refer to those as *contradiction groups*. Overall, we identified 15 contradiction groups with a total of 49 distinct pieces of advice (cf. Table 9). Most contradiction groups contained one pair of pieces of advice that directly contradict each other. We illustrate our key findings below.

Most contradiction groups addressed passwords (C2, C4, C5, C7, C11, C13). For example, C4 is about whether to enforce regular password changes for users or not. That contradiction group’s advice states two opposite approaches: to “*Force users to change passwords regularly*” (6 documents) and “*Don’t enforce regular password updates/changes for users*” (2 documents). Similar disagreements can be found in password length maximum (C2) and minimum (C5) requirements or password visibility (C7). But contradicting advice also exists for other authentication areas, like session timeouts (C3, C14), rate limiting (C8, C12), 2FA (C10), or general discussions on whether to use authentication at all (C6).

Another important aspect is how contradicting advice occurs within a single document and within the document corpus, i.e.,

Table 9: Overview of the 15 contradiction groups, each with a short description and advice examples (in italics).

ID	Description & Advice Examples	#Adv.	#Docs
C1	How authentication errors should be reported to the users. <ul style="list-style-type: none"> <i>Use generic responses for authentication errors to not weaken security</i> <i>Tell if user types in an old password</i> 	6	16
C2	Whether to set a maximum password length or not. <ul style="list-style-type: none"> <i>Set a maximum password length</i> <i>Don’t set a maximum password length</i> 	4	10
C3	Keeping users logged in or automatically log out. <ul style="list-style-type: none"> <i>Provide automatic logout/expire session</i> <i>Keep users signed in</i> 	6	9
C4	Whether to enforce regular password changes or not. <ul style="list-style-type: none"> <i>Force users to change passwords regularly</i> <i>Don’t force users to regularly change/update passwords</i> 	2	8
C5	Different advice on minimum password length. <ul style="list-style-type: none"> <i>Use a minimum of 8 characters</i> <i>Use a minimum of 12–16 characters</i> 	5	7
C6	Whether to use logins or not. <ul style="list-style-type: none"> <i>Always require/enforce authentication</i> <i>Avoid login completely if possible/not necessary</i> 	3	7
C7	Whether to allow toggling password field visibility or not. <ul style="list-style-type: none"> <i>Don’t display passwords on the screen</i> <i>Allow users to toggle password visibility when typing it</i> 	2	6
C8	Whether to use IP addresses or account for rate limits. <ul style="list-style-type: none"> <i>Limit by account, not by IP</i> <i>Limit by IP</i> 	2	6
C9	Using OAuth or OpenID Connect for authentication. <ul style="list-style-type: none"> <i>Implement 3rd-party authentication via OAuth</i> <i>Use OpenID Connect for authentication</i> 	2	6
C10	Which 2FA approach to choose. <ul style="list-style-type: none"> <i>Provide OTPs as 2FA option</i> <i>Use mobile device push notifications as second factor</i> 	3	5
C11	Different speed requirements for password hashing. <ul style="list-style-type: none"> <i>Slow down hash functions</i> <i>Choose work factors so that hashing time <1 sec</i> 	2	5
C12	Different rate limits to prevent brute force attacks. <ul style="list-style-type: none"> <i>Limit to 3 attempts per account in a set timespan</i> <i>Limit to 20 guesses per minute and IP address</i> 	4	4
C13	Entering passwords twice during registration or not. <ul style="list-style-type: none"> <i>Let users confirm passwords by writing twice</i> <i>Don’t ask for password confirmation</i> 	2	3
C14	How long should tokens for authentication live. <ul style="list-style-type: none"> <i>Limit number of token usages</i> <i>Only allow one-time usage</i> 	4	2
C15	Whether to use CAPTCHAs or not. <ul style="list-style-type: none"> <i>Use CAPTCHAs</i> <i>Use alternatives for CAPTCHAs</i> 	2	2

across multiple documents. Regarding *intra-document* contradicting advice, we found this to be almost non-existent. Actually, only 2 documents contain contradicting pieces of advice. However, this low *intra-document* contradiction rate is to be expected because it is very unlikely that an author is inconsistent within a document. Analyzing the *inter-document* advice contradictions, we identified 60 documents (48.4%) that contained at least one piece of advice belonging to a contradiction group. That said, the document corpus contains many contradictions, indicating inconsistency and unreliability of online advice.

Contradiction groups contain opposite advice by definition. Especially for debatable topics, contrary recommendations and opinions exist that are sometimes mixed with recommendable advice (e.g., C1 or C4) and discussed in the document corpus. Therefore, the prevalence of debatable advice (Sect. 4.5) among the contradiction groups is considerably higher at 30.6% compared to the overall 18.0% rate of debatable advice.

Summary: Contradicting Advice. We found 15 groups of contradicting advice. While it is almost non-existent within a single document,

48.4% of the documents contained some piece of contradicting advice. Contradiction groups tend to contain more debatable advice.

5 DISCUSSION

Below, we summarize our key insights and discuss recommendations for developers, advice writers, and academia. This section concludes with a discussion of future work.

5.1 Potential Impact of Advice on Authentication Deployments

Usable and secure web authentication is challenging, and previous work uncovered many issues, such as the widespread use of passwords in general [12, 13], the deployment of complicated password policies [46, 51, 87, 88], and the low adoption of password alternatives such as MFA and FIDO2 [7, 12, 16, 21, 29, 73, 76]. Moreover, developers regularly depend on online resources to solve programming problems [2, 3, 5, 15, 98] and might blindly trust information from blogs, documentation, or platforms such as SO [24, 25, 26].

Our findings illustrate how online advice for web authentication might contribute to unusable and insecure authentication deployments. When developers seek online advice to solve challenges around online authentication, they likely stumble upon advice that is of limited help to make web authentication usable and secure. Most advice covered technical security implementation details that are not relevant for end-users, while only 272 of 1,208 pieces of advice related to end-users' usable security (Sect. 4.1). Contradicting, debatable, and outdated advice might even have a negative effect.

For example, passwords were—despite their known potentially negative impact on security and usability—the most prevalent advice topic and often the only discussed authentication approach (Table 7). The drawbacks and potential alternatives are rarely mentioned. We believe that such advice contributes to passwords' undiminished high popularity. Similarly, many websites do not follow best practices for password policies [51], and advice is also often not in line with best practices (Sect. 4.5, Sect. 4.6). Developers are not necessarily authentication or usable security experts. They might follow the advice unaware of the drawbacks and alternatives.

Considering that developers rely on online resources for solving security programming challenges [2, 24], we think that online advice is a major factor that affects web authentication usable security negatively, e.g., outdated but still followed advice (e.g., changing passwords regularly) or following advice that was always known to counteract usability and security (e.g., using security questions).

5.2 Challenges for Developers

Based on our findings, we identified four main challenges for developers. Overall, the results indicate a mixture of advice that could and should be followed and advice that should be carefully considered or avoided at all.

Advice Diversity & Prioritization. The advice was highly diverse (Sect. 4.3), i.e., we found lots of advice on a single topic (e.g., password policies). This and the high advice volume indicate an overload for developers and a lack of advice prioritization among its authors, as similarly found in related work on end-user advice [77].

Scattered Advice. The advice was also distributed over many documents (Sect. 4.2). While some documents contained more advice, many contained only a single piece of advice. Moreover, most advice was only covered in a single document. This makes it harder for developers to find relevant advice; they might have to consider many documents to find helpful advice for their needs.

Debatable Advice. Debatable advice (Sect. 4.5) challenges developers, as they would have to recognize and not blindly follow detrimental advice. However, it cannot be assumed that developers (i) are usable security experts and, for example, know which advice is outdated and should not be followed anymore; or (ii) can recognize the consequences that poor usability of authentication features can have [36].

Contradicting Advice. Finding contradicting advice (Sect. 4.6) obliges the developer to decide which advice to follow and might be confusing. As with debatable advice, it is unclear whether developers have the skills to make an educated decision and assess related consequences. Even when developers encounter only one piece of contradicting advice, chances are to blindly follow debatable or otherwise disadvantageous advice instead of best practices.

5.3 Recommendations

Given the potential impact of advice on authentication deployments and the associated challenges for developers, we advocate to rethink how advice is given and consumed to achieve authentication for end-users that is both more usable and secure. Based on our findings, we make the following recommendations for improving security advice for developers in the future.

Developers. We found that online security advice on web authentication is diverse and distributed over many resources on the web. Overall, developers should consider their context and not follow advice blindly, as we found debatable advice that could not be unconditionally recommended (Sect. 4.5). Such assessments are mostly impractical for developers since they require lots of expertise, effort, and literature research. Hence, reasonable resources that can be assumed to provide high-quality advice are necessary.

We recommend developers look into the OWASP Cheat Sheets. According to our results, they contain the most advice and, therefore, are the most extensive resource, obviating the need for developers to look at many other resources that provide little or no advice. Another reason is that OWASP Cheat Sheets contain up-to-date advice specifically tailored to software practitioners.

Giving Advice. We found several instances of outdated, incorrect, or otherwise debatable advice. Hence, authors should check their advice to avoid such problems before giving advice. This also implies regularly updating advice to prevent giving outdated advice. Unfortunately, checking advice is not straightforward, requiring expertise and considering resources such as the latest usable security research. Since advice is highly dependent on specific contexts, advice givers should describe in which situations advice should be followed or not to prevent developers from following advice that does not suit their project.

As already discussed, we found parallels between advice and the authentication ecosystem. We argue that advice is related to and could influence the authentication landscape. Therefore, advice givers should carefully choose what authentication topics they

write about. That does not mean that no advice on passwords should be given. However, the authors should focus on approaches that prevent weaknesses in passwords. For example, when writing about passwords, their drawbacks (e.g., phishing, password reuse) should also be mentioned, and countermeasures (e.g., 2FA) or alternatives (e.g., FIDO2) should be outlined.

Official Institutions. While we mentioned OWASP Cheat Sheets, the reader may wonder where official organizations stand as a source of advice. To our surprise, participants rarely reported official documents. For example, no NIST Special Publication was reported. This is especially remarkable as several documents referenced NIST guidelines and considered those authoritative. A closer look at OWASP Cheat Sheets [68] and the related NIST SP [64] quickly reveals major differences. For example, NIST's PDF is much longer at 79 pages than OWASP Cheat Sheets. The language in NIST's documents is also more abstract, e.g., passwords are referred to as "*memorized secret authenticators*" [64]. We hypothesize that these factors lower the accessibility of NIST documents for developers and our participants did not report them for this reason.

Nonetheless, official institutions like NIST are reputable, authoritative organizations whose documents contain reasonable advice. We propose that official institutions provide their advice in a form similar to OWASP cheat sheets. This could be done by condensing the most important advice from their extensive guidelines, standards, and other documents and presenting it in a developer-friendly and actionable way.

Authoritative Advice. We obtained 272 distinct pieces of advice that showed high diversity and were scattered all over the internet. This imposes a challenge, as no developer would search 406 documents for advice as we did. Hence, creating an authoritative source with helpful, actionable, carefully constructed web authentication advice for developers might improve accessibility, advice adoption, and end-user usable security. Moreover, this might prevent encountering debatable or contradicting advice.

It is yet unclear who should create and maintain authoritative advice. As already discussed, official institutions like NIST are considered authoritative but lack accessibility. Since this is mainly a presentation issue, we argue that official institutions could take this role. We found OWASP Cheat Sheets [68] to be the most comprehensive knowledge base for authentication advice. Hence, OWASP should continue the Cheat Sheet Series project, while official institutions should start creating similar resources. The advice extracted in this paper might be a starting point for creating (or extending) authoritative usable security authentication advice.

Knowledge Transfer: Academia to Practice. Unsurprisingly, developers do not consider scientific resources like papers helpful (Figure 1). Developers are unlikely to read dozens of papers that might require expert knowledge or are hidden behind paywalls. Instead, developers prefer online resources and heavily draw on contained advice [2, 3, 5]. However, academic papers yield new usable security knowledge and are the main communication channel for scientific results. Their impact is limited if they do not foster any change and gain real-world traction. For developer authentication advice, we can confirm the knowledge gap between academia and practice that Gutfleisch et al. identified [36]. To overcome this gap, researchers should aim to provide their knowledge in ways that

developers can easily use and access. For example, new research results could be added to OWASP Cheat Sheets. While this can improve the adoption of advice that is grounded on research insights, the adoption is likely limited overall. Recent research indicates that Cheat Sheet advice is only partially followed [42].

5.4 Comparison with Related Work

While previous research investigated online security advice for both end-users [14, 20, 38, 43, 75, 77, 79] and software developers [2, 3, 5, 9, 10, 15, 24, 25, 26, 28, 34, 63, 78, 98], we are the first to explore developer advice on authentication that affects usable security for end-users. As already discussed, we found similar themes compared to related work, like challenges around advice diversity and prioritization [77], and a usable security knowledge gap between academia and industry [36]. Redmiles et al. investigated security advice regarding its quality (e.g., actionability) in a user study [77]. As we performed a researcher evaluation, we refrained from evaluating aspects like developer actionability. Such an assessment with developers, similar to Redmiles et al. [77], is potential future work and possible with our dataset. This might be interesting, as our overall impression is that quality is mixed, e.g., some advice is highly concrete (e.g., containing source code) while other is vague.

Moreover, our work might provide some explanations for observations in related work. While not directly comparable, we see parallels to the findings of Lee et al. [51]. Lee et al. found that only 13% of websites followed all password policy best practices. A contributing factor to this can be developer advice—due to the plethora of different advice. Some advice might be recommendable, and others debatable, while potentially contradicting each other at the same time. Best practices identified by Lee et al., e.g., block lists and strength meters, were also recommendations in our dataset but rarely occurred, reflecting the low adoption of best practices. This low advice adoption might be caused by the challenges (Sect. 5.2) that can become a burden for developers, ultimately resulting in a poor cost–benefit trade-off similar to the one outlined by Herley in 2009 for end-user security advice [39]. Overall, we note that not only developers impact usable security of authentication. Also, system administrators configuring authentication have an impact and draw on advice, e.g., consulting standards and guidelines, as recently found by Sahin et al. [83]. The authors identified similar challenges compared to our study, like administrators facing and following outdated recommendations.

In our dataset, the fraction of documents from Q&A forums like SO (3.7%) is lower than in related work (e.g., 10–40% [2]). As participants were free to choose any resource, we think the lower prevalence is due to different study characteristics. SO posts are likely most useful for programming as they provide code snippets for copy and paste [24]. However, our scenario is a more abstract, conceptual task that does not include programming, while Acar et al. [2] experimented with developers writing code. Other factors might be self-reporting and social desirability biases that prevent participants from reporting SO because the platform is also known to contain low-quality or insecure answers [15, 24]. While we can only speculate about the reasons, future research is needed to investigate which resources software professionals prefer for abstract, conceptual tasks like ours.

5.5 Outlook & Future Work

Changing advice and the authentication landscape can be considered long-term projects requiring suitable technical solutions. The endeavor to replace passwords has been an ongoing effort for decades [6] that turned out to be hard [12] and has not been successful yet. Regarding advice, passwords are the prevailing topic. Therefore, we believe passwords to stay for more years to come. Users also still tend to prefer passwords [100]. However, the industry is trying to offer and shift to passwordless alternatives. In 2022, Apple, Microsoft, and Google announced plans to extend FIDO2 and introduce *passkeys* [22]. Passkeys aim to replace passwords by syncing FIDO credentials across the users' devices. While passkeys seem promising, future research is needed to study their effect. The advent of passkeys also calls for a future replication of this study to check how online advice evolves and how passkeys are covered.

While we qualitatively analyzed developer-reported web authentication advice in this paper, our work does not investigate developers' interaction with advice and advice impact on authentication usable security. Conceptual decisions (e.g., whether 2FA is implemented or not) might not be decided by developers alone [36]. Hence, further investigation of advice adoption and decision processes is needed. One aspect of this is the prioritization of advice. We refrained from indicating recommendation strength for the advice, as a valid assessment requires a broader expert evaluation beyond the involved researchers. Future work could investigate recommendation strength with experienced software professionals. We also propose to conduct a developer experiment focusing on the impact of advice on the final software to understand how advice should be given to developers. Similarly, future work could investigate the impact of code snippets, as we only focussed on textual (prose) advice. We hope that our dataset can lay the foundation for future work in these directions.

6 CONCLUSION

In this work, we qualitatively analyzed 272 pieces of advice containing developer advice on usable and secure authentication. We identify several major challenges for developers. For example, the advice is scattered over many resources and, therefore, hard to find. Despite their shortcoming, passwords are the most discussed authentication approach; the advice lacks the discussion of drawbacks and alternatives. Moreover, unconsciously following advice can induce usable security problems, as we found instances of debatable, outdated, and contradicting advice. Overall, our findings suggest that the advice potentially impacts authentication deployments. Our recommendations can inform better advice giving to improve usable and secure authentication on the web.

ACKNOWLEDGMENTS

We thank the anonymous reviewers for their valuable feedback and for helping us to improve this paper. We thank our study's participants for taking their time and allowing us to gain interesting insights. This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2092 CASA – 390781972.

REFERENCES

- [1] Jacob Abbott and Sameer Patil. 2020. How Mandatory Second Factor Affects the Authentication User Experience. In *Proc. 2020 CHI Conference on Human Factors in Computing Systems (CHI'20)*. ACM.
- [2] Yasemin Acar, Michael Backes, Sascha Fahl, Doowon Kim, Michelle L. Mazurek, and Christian Stransky. 2016. You Get Where You're Looking For: The Impact of Information Sources on Code Security. In *Proc. 37th IEEE Symposium on Security and Privacy (SP'16)*. IEEE.
- [3] Yasemin Acar, Michael Backes, Sascha Fahl, Doowon Kim, Michelle L. Mazurek, and Christian Stransky. 2017. How Internet Resources Might Be Helping You Develop Faster but Less Securely. *IEEE Security & Privacy*, 15, 2, 50–60.
- [4] Yasemin Acar, Sascha Fahl, and Michelle L. Mazurek. 2016. You are Not Your Developer, Either: A Research Agenda for Usable Security and Privacy Research Beyond End Users. In *Proc. 2016 IEEE Secure Development Conference (SecDev'16)*. IEEE.
- [5] Yasemin Acar, Christian Stransky, Dominik Wermke, Charles Weir, Michelle L. Mazurek, and Sascha Fahl. 2017. Developers Need Support Too: A Survey of Security Advice for Software Developers. In *Proc. 2017 IEEE Secure Development Conference (SecDev'17)*. IEEE.
- [6] Anne Adams and Martina Angela Sasse. 1999. Users are not the enemy: Why users compromise security mechanisms and how to take remedial measures. *Communications of the ACM*, 42, 12, 40–46.
- [7] Elham Al Qahtani, Lipsarani Sahoo, Youstra Javed, and Mohamed Shehab. 2022. "Why Would Someone Hack Me out of Thousands of Students": Video Presenter's Impact on Motivating Users to Adopt 2FA. In *Proc. 27th ACM Symposium on Access Control Models and Technologies (SACMAT'22)*. ACM.
- [8] Apple Inc. 2017. Face ID Security. (Nov. 2017). https://www.apple.com/business-docs/FaceID_Security_Guide.pdf.
- [9] David Barrera, Christopher Bellman, and Paul C. van Oorschot. 2022. A Close Look at a Systematic Method for Analyzing Sets of Security Advice. (2022). <https://arxiv.org/abs/2209.04502> arXiv: 2209.04502 [cs.CR].
- [10] David Barrera, Christopher Bellman, and Paul C. van Oorschot. 2023. Security Best Practices: A Critical Analysis Using IoT as a Case Study. *ACM Trans. Priv. Secur.*, 26, 2, Article 13, 30 pages.
- [11] Joseph Bonneau, Elie Bursztein, Ilan Caron, Rob Jackson, and Mike Williamson. 2015. Secrets, Lies, and Account Recovery: Lessons from the Use of Personal Knowledge Questions at Google. In *Proc. 24th Int. Conf. on World Wide Web (WWW'15)*. ACM.
- [12] Joseph Bonneau, Cormac Herley, Paul C. Van Oorschot, and Frank Stajano. 2012. The quest to replace passwords: A framework for comparative evaluation of web authentication schemes. In *Proc. 33rd IEEE Symposium on Security and Privacy (SP'12)*. IEEE.
- [13] Joseph Bonneau, Cormac Herley, Paul C. van Oorschot, and Frank Stajano. 2015. Passwords and the Evolution of Imperfect Authentication. *Commun. ACM*, 58, 7, 78–87.
- [14] Karoline Busse, Julia Schäfer, and Matthew Smith. 2019. Replication: No One Can Hack My Mind Revisiting a Study on Expert and Non-Expert Security Practices and Advice. In *Proc. 15th Symposium on Usable Privacy and Security (SOUPS'19)*. USENIX.
- [15] Mengsu Chen, Felix Fischer, Na Meng, Xiaoyin Wang, and Jens Grossklags. 2019. How Reliable is the Crowdsourced Knowledge of Security Implementation? In *Proc. IEEE/ACM 41st International Conference on Software Engineering (ICSE'19)*. IEEE.
- [16] Jessica Colnago, Summer Devlin, Maggie Oates, Chelse Swoopes, Lujo Bauer, Lorrie Cranor, and Nicolas Christin. 2018. "It's Not Actually That Horrible": Exploring Adoption of Two-Factor Authentication at a University. In *Proc. 2018 CHI Conference on Human Factors in Computing Systems (CHI'18)*. ACM.
- [17] Sanchari Das, Andrew Dingman, and L. Jean Camp. 2018. Why Johnny Doesn't Use Two Factor A Two-Phase Usability Study of the FIDO U2F Security Key. In *Financial Cryptography and Data Security*. Sarah Meiklejohn and Kazuo Sako, (Eds.) Springer, 160–179.
- [18] Brian Dean. 2019. We analyzed 5 million Google Search Results: Here's What We Learned About Organic Click Through Rate. Accessed: 2021-11-26. Backlinko, (2019). <https://backlinko.com/google-ctr-stats>.
- [19] Xavier de Carné de Carnavalet and Mohammad Mannan. 2014. From Very Weak to Very Strong: Analyzing Password-Strength Meters. In *Proc. 21st Annual Network and Distributed System Security Symposium (NDSS'14)*. The Internet Society.
- [20] Michael Fagan and Mohammad Maifi Hasan Khan. 2016. Why Do They Do What They Do?: A Study of What Motivates Users to (Not) Follow Computer Security Advice. In *Proc. 12th Symposium on Usable Privacy and Security (SOUPS'16)*. USENIX.
- [21] Florian M. Farke, Lennart Lorenz, Theodor Schnitzler, Philipp Markert, and Markus Dürmuth. 2020. "You still use the password after all" – Exploring FIDO2 Security Keys in a Small Company. In *Proc. 16th Symposium on Usable Privacy and Security (SOUPS'20)*. USENIX.

- [22] FIDO Alliance. 2022. Apple, Google and Microsoft Commit to Expanded Support for FIDO Standard to Accelerate Availability of Passwordless Sign-Ins. Accessed: 2022-11-27. (2022). <https://fidoalliance.org/apple-google-and-microsoft-commit-to-expanded-support-for-fido-standard-to-accelerate-a-availability-of-passwordless-sign-ins/>.
- [23] FIDO Alliance. 2022. FIDO2. Accessed: 2022-11-09. (2022). <https://fidoalliance.org/fido2/>.
- [24] Felix Fischer, Konstantin Böttinger, Huang Xiao, Christian Stransky, Yasemin Acar, Michael Backes, and Sascha Fahl. 2017. Stack Overflow Considered Harmful? The Impact of Copy&Paste on Android Application Security. In *Proc. 38th IEEE Symposium on Security and Privacy (SP'17)*. IEEE.
- [25] Felix Fischer, Yannick Stachelscheid, and Jens Grossklags. 2021. The Effect of Google Search on Software Security: Unobtrusive Security Interventions via Content Re-Ranking. In *Proc. 28th ACM Conference on Computer and Communication Security (CCS'21)*. ACM.
- [26] Felix Fischer, Huang Xiao, Ching-Yu Kao, Yannick Stachelscheid, Benjamin Johnson, Danial Razar, Paul Fawkesley, Nat Buckley, Konstantin Böttinger, Paul Muntean, and Jens Grossklags. 2019. Stack Overflow Considered Helpful! Deep Learning Security Nudges Towards Stronger Cryptography. In *Proc. 28th Usenix Security Symposium (SEC'19)*. USENIX.
- [27] Anuj Gautam, Shan Lalani, and Scott Ruoti. 2022. Improving Password Generation Through the Design of a Password Composition Policy Description Language. In *Proc. 18th Symposium on Usable Privacy and Security (SOUPS'22)*. USENIX.
- [28] Lisa Geierhaas, Anna-Marie Orloff, Matthew Smith, and Alena Naiakshina. 2022. Let's Hash: Helping Developers with Password Security. In *Proc. 18th Symposium on Usable Privacy and Security (SOUPS'22)*. USENIX.
- [29] Sanam Ghorbani Lyastani, Michael Schilling, Michaela Neumayr, Michael Backes, and Sven Bugiel. 2020. Is FIDO2 the Kingslayer of User Authentication? A Comparative Usability Study of FIDO2 Passwordless Authentication. In *Proc. 41st IEEE Symposium on Security and Privacy (SP'20)*. IEEE.
- [30] Maximilian Golla and Markus Dürmuth. 2018. On the Accuracy of Password Strength Meters. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security (CCS'18)*. ACM.
- [31] Maximilian Golla, Miranda Wei, Juliette Hainline, Lydia Filipe, Markus Dürmuth, Elissa Redmiles, and Blase Ur. 2018. "What Was That Site Doing with My Facebook Password?": Designing Password-Reuse Notifications. In *Proc. 25th ACM Conference on Computer and Communication Security (CCS'18)*. ACM.
- [32] Google Inc. 2022. Programmable Search Engine. (2022). <https://developers.google.com/custom-search/>.
- [33] Peter L. Gorski, Emanuel von Zeuschwitz, Luigi Lo Iacono, and Matthew Smith. 2019. On providing systematized access to consolidated principles, guidelines and patterns for usable security research and development. *Journal of Cybersecurity*, 5, 1.
- [34] Peter Leo Gorski, Luigi Lo Iacono, Dominik Wermke, Christian Stransky, Sebastian Möller, Yasemin Acar, and Sascha Fahl. 2018. Developers Deserve Security Warnings, Too: On the Effect of Integrated Security Advice on Cryptographic API Misuse. In *Proc. 14th Symposium on Usable Privacy and Security (SOUPS'18)*. USENIX.
- [35] Matthew Green and Matthew Smith. 2016. Developers are Not the Enemy!: The Need for Usable Security APIs. *IEEE Security & Privacy*, 14, 5, 40–46.
- [36] Marco Gutfleisch, Jan H. Klemmer, Niklas Busch, Yasemin Acar, M. Angela Sasse, and Sascha Fahl. 2022. How Does Usable Security (Not) End Up in Software Products? Results From a Qualitative Interview Study. In *Proc. 43rd IEEE Symposium on Security and Privacy (SP'22)*. IEEE.
- [37] Hana Habib, Pardis Emami Naeini, Summer Devlin, Maggie Oates, Chelse Swoopes, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. 2018. User Behaviors and Attitudes Under Password Expiration Policies. In *Proc. 14th Symposium on Usable Privacy and Security (SOUPS'18)*. USENIX.
- [38] Ayako A. Hasegawa, Naomi Yamashita, Tatsuya Mori, Daisuke Inoue, and Mitsuaki Akiyama. 2022. Understanding Non-Experts' Security- and Privacy-Related Questions on a Q&A Site. In *Proc. 18th Symposium on Usable Privacy and Security (SOUPS'22)*. USENIX.
- [39] Cormac Herley. 2009. So Long, and No Thanks for the Externalities: The Rational Rejection of Security Advice by Users. In *Proc. 2009 Workshop on New Security Paradigms Workshop (NSPW'09)*. ACM.
- [40] Nicolas Huaman, Sabrina Amft, Marten Oltrögge, Yasemin Acar, and Sascha Fahl. 2021. They Would do Better if They Worked Together: The Case of Interaction Problems Between Password Managers and Websites. In *Proc. 42nd IEEE Symposium on Security and Privacy (SP'21)*. IEEE.
- [41] Philip G. Inglesant and M. Angela Sasse. 2010. The true cost of unusable password policies: password use in the wild. In *Proc. 2010 CHI Conference on Human Factors in Computing Systems (CHI'10)*. ACM.
- [42] Tommaso Innocenti, Seyed Ali Mirheidari, Amin Kharraz, Bruno Crispo, and Engin Kirda. 2021. You've Got (a Reset) Mail: A Security Analysis of Email-Based Password Reset Procedures. In *Proc. 18th International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment (DIMVA'21)*. Springer.
- [43] Iulia Ion, Rob Reeder, and Sunny Consolvo. 2015. "...No one Can Hack My Mind": Comparing Expert and Non-Expert Security Practices. In *Proc. 11th Symposium On Usable Privacy and Security (SOUPS'15)*. USENIX.
- [44] Markus Keil, Philipp Markert, and Markus Dürmuth. 2022. "It's Just a Lot of Prerequisites": A User Perception and Usability Analysis of the German ID Card as a FIDO2 Authenticator. In *Proc. 2022 European Symposium on Usable Security (EuroUSEC'22)*. ACM.
- [45] Erin Kenneally and David Dittrich. 2012. The Menlo Report: Ethical principles guiding information and communication technology research. *SSRN Electronic Journal*.
- [46] Saranga Komanduri, Richard Shay, Patrick Gage Kelley, Michelle L Mazurek, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, and Serge Egelman. 2011. Of Passwords and People: Measuring the Effect of Password-Composition Policies. In *Proc. SIGCHI Conference on Human Factors in Computing Systems (CHI'11)*. ACM.
- [47] Udo Kuckartz. 2019. Qualitative Text Analysis: A Systematic Approach. In *Compendium for Early Career Researchers in Mathematics Education*. Gabriele Kaiser and Norma Presmeg, (Eds.) Springer, 181–197.
- [48] Daniele Lain, Kari Kostiaainen, and Srdjan Capkun. 2022. Phishing in Organizations: Findings from a Large-Scale and Long-Term Study. In *Proc. 43rd IEEE Symposium on Security and Privacy (SP'22)*. IEEE.
- [49] Kevin Lee, Benjamin Kaiser, Jonathan Mayer, and Arvind Narayanan. 2020. An Empirical Study of Wireless Carrier Authentication for SIM Swaps. In *Proc. 16th Symposium on Usable Privacy and Security (SOUPS'20)*. USENIX.
- [50] Kevin Lee and Arvind Narayanan. 2021. Security and Privacy Risks of Number Recycling at Mobile Carriers in the United States. In *Proc. 2021 APWG Symposium on Electronic Crime Research (eCrime)*. IEEE.
- [51] Kevin Lee, Sten Sjöberg, and Arvind Narayanan. 2022. Password policies of most top websites fail to follow best practices. In *Proc. 18th Symposium on Usable Privacy and Security (SOUPS'22)*. USENIX.
- [52] Luigi Lo Iacono, Matthew Smith, Emanuel von Zeuschwitz, Peter Leo Gorski, and Peter Nehren. 2018. Consolidating Principles and Patterns for Human-centred Usable Security Research and Development. In *Proc. 3rd European Workshop on Usable Security (EuroUSEC'18)*. Internet Society.
- [53] Sanam Ghorbani Lyastani, Michael Backes, and Sven Bugiel. 2023. A Systematic Study of the Consistency of Two-Factor Authentication User Journeys on Top-Ranked Websites. In *Network and Distributed System Security (NDSS) Symposium 2023*.
- [54] Philipp Markert, Florian Farke, and Markus Dürmuth. 2019. View The Email to Get Hacked: Attacking SMS-Based Two-Factor Authentication. In *Who Are You?! Adventures in Authentication Workshop (WAY'19)*.
- [55] Philipp Mayring. 2015. Qualitative Content Analysis: Theoretical Background and Procedures. In *Approaches to Qualitative Research in Mathematics Education: Examples of Methodology and Methods*. Angelika Bikner-Ahsbals, Christine Knipping, and Norma Presmeg, (Eds.) Springer, 365–380.
- [56] Michelle Mazurek. 2022. We Are the Experts, and We Are the Problem: The Security Advice Fiasco. In *Proc. 29th ACM SIGSAC Conference on Computer and Communications Security (CCS'22)*. Keynote. ACM.
- [57] Allison McDonald, Carlo Sugatan, Tamy Guberek, and Florian Schaub. 2021. The Annoying, the Disturbing, and the Weird: Challenges with Phone Numbers as Identifiers and Phone Number Recycling. In *CHI Conference on Human Factors in Computing Systems (CHI '21)*. ACM.
- [58] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. 2019. Reliability and Inter-Rater Reliability in Qualitative Research: Norms and Guidelines for CSCW and HCI Practice. *Proc. ACM Hum.-Comput. Interact.*, 3, CSCW, (Nov. 2019), 23 pages.
- [59] Collins W. Munyendo, Yasemin Acar, and Adam J. Aviv. 2023. "In Eighty Percent of the Cases, I Select the Password for Them": Security and Privacy Challenges, Advice, and Opportunities at Cybercafes in Kenya. In *Proc. 44th IEEE Symposium on Security and Privacy (SP'23)*. IEEE.
- [60] Alena Naiakshina, Anastasia Danilova, Christian Tiefenau, Marco Herzog, Sergej Dechand, and Matthew Smith. 2017. Why Do Developers Get Password Storage Wrong?: A Qualitative Usability Study. In *Proc. 24th ACM Conference on Computer and Communication Security (CCS'17)*. ACM.
- [61] Daniela Napoli. 2018. Developing Accessible and Usable Security (ACCUS) Heuristics. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems (CHI EA'18)*. ACM.
- [62] Net Marketshare. 2021. Search Engine Market Share. Accessed: 2021-11-16. (2021). <https://netmarketshare.com/search-engine-market-share.aspx>.
- [63] Duc Cuong Nguyen, Dominik Wermke, Yasemin Acar, Michael Backes, Charles Weir, and Sascha Fahl. 2017. A Stitch in Time: Supporting Android Developers in Writing Secure Code. In *Proc. 24th ACM Conference on Computer and Communication Security (CCS'17)*. ACM.
- [64] NIST. 2017. Digital identity guidelines: authentication and lifecycle management. Accessed: 2022-12-01. (2017). <https://pages.nist.gov/800-63-3/sp800-63b.html>.

- [65] NIST. 2016. Questions... and buzz surrounding draft NIST Special Publication 800-63-3. Accessed: 2022-11-27. (2016). <https://www.nist.gov/blogs/cybersecurity-insights/questionsand-buzz-surrounding-draft-nist-special-publication-800-63-3>.
- [66] Sean Oesch, Scott Ruoti, James Simmons, and Anuj Gautam. 2022. "It Basically Started Using Me." An Observational Study of Password Manager Usage. In *Proc. 2022 CHI Conference on Human Factors in Computing Systems (CHI'22)*. ACM.
- [67] OWASP. 2021. Authentication Cheat Sheet. Accessed: 2022-12-01. (2021). https://cheatsheetseries.owasp.org/cheatsheets/Authentication_Cheat_Sheet.html.
- [68] OWASP. 2021. OWASP Cheat Sheet Series. Accessed: 2022-12-01. (2021). <https://cheatsheetseries.owasp.org/>.
- [69] Kentrell Owens, Olabode Anise, Amanda Krauss, and Blase Ur. 2021. User Perceptions of the Usability and Security of Smartphones as FIDO2 Roaming Authenticators. In *Proc. 17th Symposium on Usable Privacy and Security (SOUPS'21)*. USENIX.
- [70] Bing Pan, Helene Hembrooke, Thorsten Joachims, Lori Lorigo, Geri Gay, and Laura Granka. 2007. In Google We Trust: Users' Decisions on Rank, Position, and Relevance. *Jour. of Comp.-Med. Comm.*, 12, 3, 801–823.
- [71] Arnis Parsovs. 2014. Practical Issues with TLS Client Certificate Authentication. In *Proc. 2014 Network and Distributed System Security Symposium (NDSS'14)*. Internet Society.
- [72] Sarah Pearman, Jeremy Thomas, Pardis Emami Naeini, Hana Habib, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, Serge Egelman, and Alain Forget. 2017. Let's Go in for a Closer Look: Observing Passwords in Their Natural Habitat. In *Proc. 24th ACM Conference on Computer and Communication Security (CCS'17)*. ACM.
- [73] Thanasis Petsas, Giorgos Tsiantonakis, Elias Athanasopoulos, and Sotiris Ioannidis. 2015. Two-Factor Authentication: Is the World Ready? Quantifying 2FA Adoption. In *Proc. 8th European Workshop on System Security (EuroSec'15)*. ACM, 7 pages.
- [74] Olgierd Pieczul, S. Foley, and Mary Ellen Zurko. 2017. Developer-centered security and the symmetry of ignorance. In *Proc. 2017 New Security Paradigms Workshop (NSPW'17)*. ACM.
- [75] Elissa M. Redmiles, Sean Kross, and Michelle L. Mazurek. 2016. How I Learned to Be Secure: A Census-Representative Survey of Security Advice Sources and Behavior. In *Proc. 2016 ACM SIGSAC Conference on Computer and Communications Security (CCS'16)*. ACM.
- [76] Elissa M. Redmiles, Everest Liu, and Michelle L. Mazurek. 2017. You Want Me To Do What? A Design Study of Two-Factor Authentication Messages. In *3rd Who Are You? Adventures in Authentication Workshop (WAY'17)*. USENIX.
- [77] Elissa M. Redmiles, Noel Warford, Amritha Jayanti, Aravind Koneru, Sean Kross, Miraida Morales, Rock Stevens, and Michelle L. Mazurek. 2020. A Comprehensive Quality Evaluation of Security and Privacy Advice on the Web. In *Proc. 29th USENIX Security Symposium (SEC'20)*. USENIX.
- [78] Rob Reeder, E Cram Kowalczyk, and Adam Shostack. 2011. Poster: Helping engineers design NEAT security warnings. In *Proc. 7th Symposium on Usable Privacy and Security (SOUPS'11)*. ACM.
- [79] Robert W. Reeder, Iulia Ion, and Sunny Consolvo. 2017. 152 Simple Steps to Stay Safe Online: Security Advice for Non-Tech-Savvy Users. *IEEE Security and Privacy*, 15, 5, 55–64.
- [80] Ken Reese, Trevor Smith, Jonathan Dutson, Jonathan Armknecht, Jacob Cameron, and Kent Seamons. 2019. A usability study of five two-factor authentication methods. In *Proc. 15th Symposium on Usable Privacy and Security (SOUPS'19)*. USENIX.
- [81] Joshua Reynolds, Nikita Samarin, Joseph Barnes, Taylor Judd, Joshua Mason, Michael Bailey, and Serge Egelman. 2020. Empirical Measurement of Systemic 2FA Usability. In *Proc. 29th USENIX Security Symposium (SEC'20)*. USENIX.
- [82] Joshua Reynolds, Trevor Smith, Ken Reese, Luke Dickinson, Scott Ruoti, and Kent Seamons. 2018. A Tale of Two Studies: The Best and Worst of YubiKey Usability. In *Proc. 39th IEEE Symposium on Security and Privacy (SP'18)*. IEEE.
- [83] Sena Sahin, Suood Al Roomi, Tara Poteat, and Frank Li. 2023. Investigating the Password Policy Practices of Website Administrators. In *Proc. 44th IEEE Symposium on Security and Privacy (SP'23)*. IEEE.
- [84] Stuart Schechter, A.J. Brush, and Serge Egelman. 2009. It's no secret: Measuring the security and reliability of authentication via 'secret' questions. In *Proc. 2009 IEEE Symposium on Security and Privacy*. IEEE.
- [85] Fabian Schwarz, Khue Do, Gunnar Heide, Lucjan Hanzlik, and Christian Rossow. 2022. FeDo: Recoverable FIDO2 Tokens Using Electronic IDs. In *Proc. 29th ACM Conference on Computer and Communications Security (CCS'22)*. ACM.
- [86] Richard Shay, Patrick Gage Kelley, Saranga Komanduri, Michelle L. Mazurek, Blase Ur, Timothy Vidas, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. 2012. Correct Horse Battery Staple: Exploring the Usability of System-Assigned Passphrases. In *Proc. 8th Symposium on Usable Privacy and Security (SOUPS'12)*. ACM.
- [87] Richard Shay, Saranga Komanduri, Adam L. Durity, Phillip Huh, Michelle L. Mazurek, Sean M. Segreti, Blase Ur, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. 2016. Designing Password Policies for Strength and Usability. *ACM Transactions on Information and System Security (TISSEC)*, 18, 4, 1–34.
- [88] Richard Shay, Saranga Komanduri, Adam L. Durity, Phillip (Seyoung) Huh, Michelle L. Mazurek, Sean M. Segreti, Blase Ur, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. 2014. Can Long Passwords Be Secure and Usable? In *Proc. SIGCHI Conference on Human Factors in Computing Systems (CHI'14)*. ACM.
- [89] Stack Overflow. 2022. Stack Overflow Developer Survey 2022. Accessed: 2022-11-10. (2022). <https://insights.stackoverflow.com/survey/2022>.
- [90] Statcounter. 2021. Search Engine Market Share Worldwide. Accessed: 2021-11-16. (2021). <https://gs.statcounter.com/search-engine-market-share>.
- [91] Elizabeth Stobert and Robert Biddle. 2014. The Password Life Cycle: User Behaviour in Managing Passwords. In *Proc. 10th Symposium on Usable Privacy and Security (SOUPS'14)*. USENIX.
- [92] Christian Stransky, Oliver Wiese, Volker Roth, Yasemin Acar, and Sascha Fahl. 2022. 27 Years and 81 Million Opportunities Later: Investigating the Use of Email Encryption for an Entire University. In *Proc. 43rd IEEE Symposium on Security and Privacy (SP'22)*. IEEE.
- [93] Mohammad Tahaei and Kami Vaniea. 2019. A Survey on Developer-Centred Security. In *Proc. 2019 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. IEEE.
- [94] Upwork. 2022. Software Developer Hourly Rates. (2022). <https://www.upwork.com/hire/software-developers/cost/>.
- [95] Blase Ur, Patrick Gage Kelley, Saranga Komanduri, Joel Lee, Michael Maass, Michelle L. Mazurek, Timothy Passaro, Richard Shay, Timothy Vidas, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. 2012. How does your password measure up? The effect of strength meters on password creation. In *Proc. 21st Usenix Security Symposium (SEC'12)*. USENIX.
- [96] Ding Wang, Xuan Shan, Qiying Dong, Yaosheng Shen, and Chunfu Jia. 2023. No Single Silver Bullet: Measuring the Accuracy of Password Strength Meters. In *Proc. 32nd USENIX Security Symposium (SEC'23)*. USENIX.
- [97] Claes Wohlin. 2014. Guidelines for snowballing in systematic literature studies and a replication in software engineering. In *Proc. 18th Int. Conf. on Evaluation and Assessment in Software Engineering (EASE'14)*. ACM.
- [98] Xin-Li Yang, David Lo, Xin Xia, Zhi-Yuan Wan, and Jian-Ling Sun. 2016. What security questions do developers ask? a large-scale study of stack overflow posts. *Journal of Computer Science and Technology*, 31, 5, 910–924.
- [99] Samira Zibaei, Dina Rinoa Malapaya, Benjamin Mercier, Amirali Salehi-Abari, and Julie Thorpe. 2022. Do Password Managers Nudge Secure (Random) Passwords? In *Proc. 18th Symposium on Usable Privacy and Security (SOUPS'22)*. USENIX.
- [100] Verena Zimmermann and Nina Gerber. 2020. The password is dead, long live the password – A laboratory study on user perceptions of authentication schemes. *Int. J. of Hum.-Comp. Studies*, 133, 26–44.

A REPLICATION PACKAGE

For transparency, potential replications of this study, and to support meta-research, we provide artifacts within a replication package.

Material. We provide the following artifacts: (1) The scenario and task we used, (2) the materials used for recruiting, (3) the consent form, (4) the questionnaire from the developer survey, (5) the list of advice that we extracted from the documents, (6) the detailed advice analysis results, (7) the set of documents, queries, and search results, and (8) the extended version of this paper. To ensure participant privacy, we do not provide any raw questionnaire responses.

Link. The replication package is available at <https://doi.org/10.17605/OSF.IO/PFB5H>.