Designing Tools for High-Quality Alt Text Authoring

Kelly Mack
Microsoft Research and University of Washington
Redmond and Seattle, Washington
kmack3@cs.washington.edu

Edward Cutrell
Microsoft Research
Redmond, Washington
cutrell@microsoft.com

Bongshin Lee
Microsoft Research
Redmond, Washington
bongshin@microsoft.com

Meredith Ringel Morris
Microsoft Research
Redmond, Washington
merrie@microsoft.com

ABSTRACT

Alternative (alt) text provides access to descriptions of digital images for people who use screen readers. While prior work studied screen reader users' (SRUs') preferences about alt text and automatic alt text (i.e., alt text generated by artificial intelligence), little work examined the alt text author's experience composing or editing these descriptions. We built two types of prototype interfaces for two tasks: authoring alt text and providing feedback on automatic alt text. Through combined interview-usability testing sessions with alt text authors and interviews with SRUs, we tested the effectiveness of our prototypes in the context of Microsoft PowerPoint. Our results suggest that authoring interfaces that support authors in choosing what to include in their descriptions result in higher quality alt text. The feedback interfaces highlighted considerable differences in the perceptions of authors and SRUs regarding "high-quality" alt text. Finally, authors crafted significantly lower quality alt text when starting from the automatic alt text compared to starting from a blank box. We discuss the implications of these results on applications that support alt text.

CCS CONCEPTS

• **Human-centered computing** → **Accessibility technologies**; *Accessibility systems and tools.*

KEYWORDS

alt text, alt text authoring, feedback on automatic alt text, high-quality alt text, screen reader users

ACM Reference Format:
Kelly Mack, Edward Cutrell, Bongshin Lee, and Meredith Ringel Morris. 2021. Designing Tools for High-Quality Alt Text Authoring. In *The 23rd International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '21), October 18–22, 2021, Virtual Event, USA.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3441852.3471207

ASSETS '21, October 18–22, 2021, Virtual Event, USA

ACM ISBN 978-1-4503-8306-6/21/10…$15.00
https://doi.org/10.1145/3441852.3471207

1 INTRODUCTION

People often use images to convey information in social, work, and educational settings. Therefore, it is important to ensure that the 285 million people worldwide who have visual impairments [16] have equitable access to the content of images. When accessing digital content (e.g., on the internet or in applications), many people with visual impairments use a *screen reader*, a tool that reads aloud formatted content on the screen to allow for consumption in non-visual formats. When the screen reader focuses on an image, the screen reader user (SRU)[1] can issue a command to read the *alternative (alt) text* metadata associated with the image. The alt text serves the "equivalent purpose" as the image and should contain "the *content* and *function* of the images within your web [or document] content" [22, 25].

Creating high-quality alt text is challenging. Two main sources of alt text – humans and AI captioning systems – both have flaws that cause the alt text to fall short of SRUs' preferences. Human-generated alt text can be very high quality and accurate. However, many content authors do not take the time to provide alt text [8, 11, 24], and when they do, the alt text can have quality variance due to authors' understanding of quality alt text. On the other hand, AI captioning systems can be called on demand with an image as input, but often suffer from quality issues; these issues range from comical inaccuracies (e.g., describing a tree as "food"), to incomplete texts (e.g., calling the Taj Mahal a building), to offensive errors (e.g., classifying a human as an animal).

Our work investigates how to improve alt text quality and author engagement with alt text in the context of Microsoft PowerPoint[2], a tool to create slides that often include images. Prior work has mainly focused on 1) alt text content and what should be described [15, 19, 20, 26, 28]; 2) different methods of presenting alt text to SRUs [13, 14, 17, 29]; and 3) a web/social media context [8–10, 12, 27]. Investigation of alt text in document authoring tools like PowerPoint is understudied [6, 19]. Few works study the alt text creation process or feedback interfaces for the alt text despite 1) existing research detailing the challenges and time spent on creating alt text [8]; 2) statistics around the overall lack of alt text associated with images [24]; and 3) research detailing the error rates and other quality issues with automatic alt text [7, 12]. Our work makes novel

[1]Note that not all people who use screen readers also identify as blind or low vision. However, in this paper, when we say SRU, we are referring to people who are blind or low vision and use a screen reader.
[2]Though our work focuses on the PowerPoint context, our findings can extend to other authoring interfaces that include images.

contributions to this under-explored space by improving the alt text authoring experience and creating feedback mechanisms for automatic alt text.

We created two types of interfaces: (1) *authoring interfaces*, designed to support authors in writing high-quality alt text quickly and easily; and (2) *feedback interfaces*, designed to effectively gather author feedback on automatic alt text (throughout this paper, we use the term automatic alt text to refer to alt text that was generated by a machine or artificial intelligence). We harness insights gained from prior work regarding difficulties authors encountered when authoring alt text: authors often do not know what to include or they are constrained by time in authoring [8]. We also build on the idea of scaffolding author alt text with templates, which has been studied in the context of graphs and memes [9, 13]. To our knowledge, interfaces that facilitate the critique or correction of automatic alt text in authoring tools have not been studied. To test our new interface designs, we performed study sessions that combined interview and usability testing with 12 alt text authors, half of whom had prior experience writing alt text.

We sought end-user expertise to validate if the alt text generated with our interface designs matched user expectations and preferences. Therefore, we conducted a second study with six SRU participants. In these interviews we sought to understand their alt text preferences and presentation styles. Specifically, we explored what "high-quality alt text" means to these participants. In these sessions, we also showed SRUs the data generated by our interface designs, and we discussed the alt text quality and effectiveness of our interfaces.

Our work makes three main contributions. First, we present the outcomes of our study of four interface designs for authoring alt text and providing feedback. These findings suggest that authors with different alt text authoring experience preferred different amounts of scaffolding, but all appreciated some additional support. Our findings around the interfaces for providing feedback about automatic alt text revealed a gap in the perceptions of authors and SRUs regarding high-quality alt text. Second, we discuss SRU opinions of high-quality alt text and automatic alt text. These findings highlight the importance of context for effective alt text, which can be challenging for anyone but the author of a document to know. Finally, we examine how the presence or absence of automatic alt text affects the engagement with and quality of alt text. Specifically, we saw evidence of authors viewing automatic alt text as a "gold standard" of quality rather than a last resort description of an image.

2 RELATED WORK

Our research builds on literature relating to automatic alt text and alt text authoring interfaces.

2.1 Guidelines and Design Suggestions for Alt Text

Alt text is a metadata field associated with an image that "serves the equivalent purpose" as the image, according to WCAG 2.0's A level (required) criteria [22]. W3C further lists seven classes of images and the content that should comprise associated alt text [23]. While decorative images should receive no alt text, other classes of images like "informative" should receive a "short description conveying the essential information presented by the image." At an intermediary level, there are "functional" images (e.g., images serving as buttons) that should only convey the functionality of the icon. Other groups, like the Diagram Center, classify images based on content type rather than purpose, like diagram, graph, photo, and art [5]. This group offers further considerations for alt text in general including how to incorporate context, tone, language, and objectivity [4]. Understanding the proper content to include in alt text is non-trivial; indeed, W3C provides a decision tree to help authors select information to include [21]. However, existing alt text interfaces do not expose the complexities of image classes impacting alt text content.

While consumer technologies only provide a single text field for alt text, research prototypes have investigated alternative methods and structures for presenting alt text. Morris et al. characterized the design space for alt text along five dimensions: interactivity, stability, representation, structure, and personalization [14]. They further suggested the designs for six alt text consumption experiences that engage with these five dimensions in different ways, and tested three with SRUs. Results highlighted the potential of richer representations of alt text as well as the diversity in SRU preferences. Several research prototypes have implemented alt text interfaces similar to these suggested designs, including overlaying alt text spatially on the image [17] and providing options to query alt text for further detail [29].

2.2 Automatic Alt Text in the Web and Commercial Products

Creating high-quality, accurate automatic alt text is a complex and challenging problem that draws upon the fields of computer vision and natural language processing. In an effort to be more accessible, several commercial platforms have incorporated automatic alt text. For example, PowerPoint recently announced an update to its automatic captioning system of images in slides [18]. Further, Microsoft Research released CaptionCrawler, a tool that uses image recognition to search for copies of an image on the Internet, which may have alt text that can be repurposed [11]. Looking to social media platforms, Facebook developed an automatic alt text feature, which provides a list of tags of items identified in an image, starting with people, then objects, and then elements of the setting [27]. Recently, researchers created "Twitter A11y," a browser extension that implements six methods of adding alt text to images without captions on Twitter [10]. Each of these features makes strides towards improving alt text coverage on their respective platforms. However, questions related to author interaction with or review of automatic alt text are increasingly important given its growing use, and thus require further investigation.

Though these automatic tools allow for the captioning of considerably more images on these platforms, accuracy is not guaranteed. Moreover, most platforms do not surface a confidence rating or multiple likely versions of alt text for an image. Without this information, SRUs are left questioning the accuracy, and therefore reliability of automatically generated captions, sometimes falsely overestimating the accuracy. With two studies, MacLeod et al. explored how blind or low vision (BLV) individuals experience these

automatic captions for social media images [12]. They found that BLV individuals often attempted to rationalize odd, incorrect captions. Even alt text that is accurate but unclear can be similarly problematic. For example, in a CVPR panel with blind researchers and a disability activist, a panelist commented that Facebook's alt text, comprising a list of keywords, can be too vague to be useful [7, 19]. Another panelist in this session commented that "not all errors are created equal" in automatic captioning, referring to the fact that errors made with respect to sensitive characteristics for minoritized individuals (e.g., mis-classifying a transgender person's gender, failing to recognize someone with dark skin) are particularly harmful. Further research on the subject of photographee identities revealed a preference for inclusion of identity-based information (e.g., gender, race), but only if the alt text author knows how the photographee identifies [1]. When the identity is not known, the paper recommends the use of "appearance based" (e.g., "uses a cane," "person with dark skin") language rather than "identity based" language (e.g., "a disabled, black man"). In this paper, we use appearance based language, as we do not know how the photographees shown in our study identify.

2.3 Alt Text Authoring

While understanding SRU preferences for alt text is critical, it is also important to understand authors' experiences in creating alt text. Gleason et al. asked 20 participants about their alt text habits on Twitter [8]. The most common reasons for not including alt text were the process taking too long or forgetting to add it. These results suggest further work needs to investigate the creation of alt text interfaces that reduce the time and complexity of creating alt text while highlighting what should be included given an image type.

A few research projects have investigated ways to improve the alt text authoring process, mainly through the creation of templates. Morash et al. investigated the benefits of providing alt text templates for different styles of graphs [13], while Gleason et al. provided similar templates for writing alt text for memes [9]. Both of these solutions utilize the common structures present in these types of images (e.g., bar charts often have a title, an x-axis, etc., memes are often image templates). The use of templates in other, less homogeneous scenarios and contexts is not fully explored. Our work innovates in the alt text creation space, with an emphasis on understanding and improving the authoring experience. We expand upon prior ideas of chart or meme templates by creating more general templates, and we develop a new interface that informs alt text authors about important information to include in alt text. Moreover, informed by SRUs' experience with automatic alt text and its prevalence in real-world tools, we developed two prototypes to allow alt text authors to provide feedback on automatic alt text.

3 INTERFACE DESIGNS

Based on prior work and our team's experience, we re-imagined two core tasks for PowerPoint authors regarding alt text: 1) *creating* alt text and 2) *providing feedback* about the quality of automatic alt text. We recognized that the type of an image (e.g., a photograph, a screenshot) can affect what information is integral to the alt text; therefore, we adopted mutability of interfaces as a guiding principle in our designs. After examining images from real-world PowerPoint slides, we developed four interface prototypes to support authors: two for authoring alt text and two for providing feedback about automatic alt text.

3.1 Two Broad Types of Images

Prior work highlights that different features of an image are important for alt text depending on the content in the image and its broader context [13, 19]. Consequently, we realized that interface designs that provide scaffolding and support for authors need to vary based on the type of image. We selected two broad categories of images, supported by prior work [9, 13, 19] that have different alt text content: photographs and non-photographs. While images are diverse in many dimensions outside this binary, we selected these two as a proof of concept dichotomy to test our customizable interfaces. Additionally, this dichotomy is realistic for implementation purposes, as machine learning could feasibly differentiate a photograph from a non-photograph.

Photographs. These images are photos from the real world without significant digital editing (i.e., they still look like photos), see Figure 1a. Photographs include photos of scenes, people, and other objects.

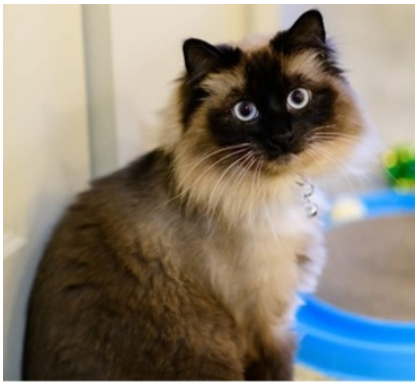
Non-photographic images (non-photographs). These images are digitally created. Often, their purpose is to convey information, see Figure 1b. Non-photographs include graphs, tables, maps, and screenshots of documents and user interfaces. We selected this category because of the unique information requirements that need to be conveyed in the alt text, as opposed to describing the aesthetic, visual features of a photograph. Alt text authors need to describe the key informational pieces of non-photographs like headers, titles, names, words, and data sources.

In our user studies, we included images from both of these categories to ensure our interface designs supported both. While each of these categories can be further subdivided (e.g., non-photographs can be divided into screenshots, graphs, diagrams, etc.), we leave this for future work.

3.2 Alt Text Authoring Interfaces

We designed two interfaces to aid authors in creating alt text through multiple rounds of design iteration. As our primary goal was to increase the quality of the final alt text generated by an author, we surfaced "suggestions" for what to include in the alt text in both interface variants. Our two classes of images required different characteristics in their alt text, and thus we included different suggestions as summarized in Table 1.

We varied the amount of *structure* provided to the author in the two authoring interface variants. In the *free-form* interface (Figure 2a), the suggestions were presented as a bulleted list added above the existing PowerPoint alt text interface. In the *template* interface (Figure 2b), each prompt was listed separately followed by a text box to respond to that prompt. To generate the final alt text from the template interface, we experimented with two ways of combining the responses in the separate text boxes: 1) we concatenated the text boxes together with spaces and 2) we listed the text box responses in the form: "prompt: text box content." For example, one alt text generated in this method was "Subject:



(a) Photographs

(b) Non-photographs

Figure 1: Examples of the two classes of images.

Table 1: The suggestions provided for what to include in the alt text for two classes of images.

Photographs	Non-photographs
• A description of the subject(s) in detail • The main action or interactions between subjects • The setting or background • Anything else that is important for users [SRUs] to take away from this image	• The type of image (e.g., bar chart, screenshot of an application) • The key information in the image (e.g., names, headers) • The main takeaway from the image you want conveyed • Anything else that is important for SRUs to take away from this image

individual sitting at table drinking coffee. Actions: smiling looking at camera. Setting: table and plant."

3.3 Automatic Alt Text Feedback Interfaces

We created two interfaces for authors to provide feedback about the automatic alt text for an image. They varied in two aspects: 1) how the input was provided and 2) location of the interface. The first interface, *check box feedback* (Figure 2c), used a set of four check boxes to ask if alt text was acceptable, unacceptable, offensive, or required "other" feedback. This last field allowed for textual explanations. This interface was appended to the bottom of the current alt text editing pane in PowerPoint. The second interface, *icon feedback* (Figure 2d), featured a thumbs up, thumbs down, and flag icon. Each icon had a tool tip, made visible on mouse over, that described the function of the icon: acceptable, unacceptable, and offensive, respectively. This interface was displayed directly below the image in the slide.

4 METHOD: TWO STUDIES

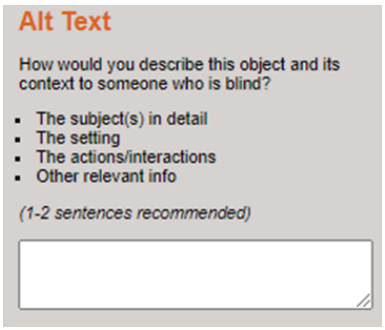
This paper presents the results of two complementary studies: (1) interview and usability testing sessions with 12 sighted authors and (2) interviews with six Screen Reader Users (SRUs), all of whom were blind or low vision. Both of these studies were approved by our organization's Institutional Review Board.

4.1 Author Interviews and Usability Testing

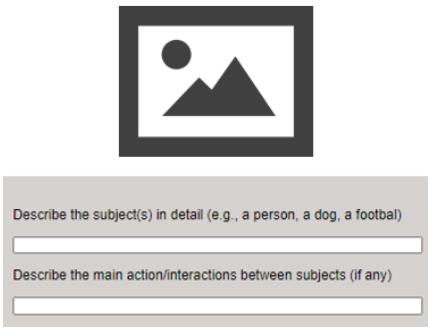
Our first study was for alt text authors and consisted of two phases. The first phase was an interview about their experience with alt text and the second phase was usability testing on our authoring and feedback interface designs.

4.1.1 Participants. We advertised our study to internal mailing lists at our organization, a U.S.-based institution comprising information workers who commonly use PowerPoint, asking those interested to submit a form that asked for demographic and background information. Additionally, those interested in participating were required to submit at least one PowerPoint deck they had created, containing at least five images. These images would be utilized in the second phase of this first study with alt text authors. Each PowerPoint submitted corresponded to one entry for a raffle for a gift card. Those who participated in an optional hour-long interview were also compensated with a $25 gift card.

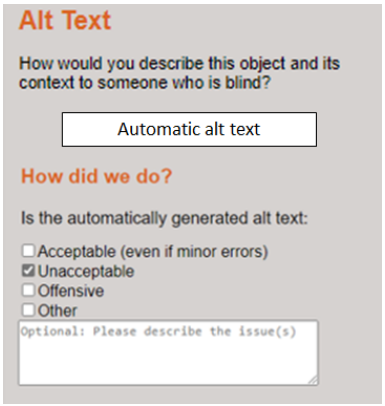
Because our usability testing phase utilized images from our participants' PowerPoints, we strategically selected a subset of the raffle entrants to interview to increase diversity of our sample, first prioritizing based on experience with alt text to ensure we had both novices and experienced alt text authors. To validate if our "suggestions" for the two image classes were suitable for a diverse image set, when looking for diverse images, we prioritized people whose PowerPoints had both photographs and non-photographs. Additionally, we selected images to increase diversity within image classes. For example, if a person had a unique diagram unlike any



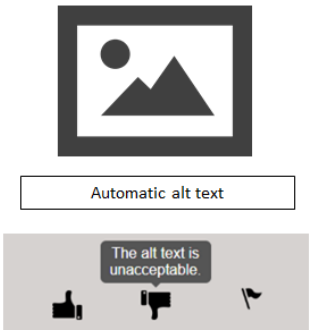
(a) The free-form authoring interface that appears in the current alt text pane.



(b) The template authoring interface.



(c) The checkbox feedback interface that appears in the current alt text pane.



(d) The icon feedback interface that appears in the slide near the image.

Figure 2: Interface variations that we tested with alt text authors: authoring interfaces (top) and feedback interfaces (bottom).

images of other participants, we selected that image to appear in the study.

We selected 12 participants; four identified as men and eight as women. The mean age was 41.5 (*SD* = 9.3), ranging between 25 and 59. With respect to experience with PowerPoint, three participants created or edited a PowerPoint deck daily, six weekly, and three monthly. When asked to rank their experience with alt text, three participants did not know anything about alt text, three knew what alt text was but had never added it in PowerPoint, five sometimes added alt text to images in PowerPoint, and one almost always added alt text to images in PowerPoint.

4.1.2 Study Material. We prepared interactive interface prototypes, using ReactJS, for the second (usability testing) phase of the study. We customized the interfaces for each participant to utilize their own images from their submitted PowerPoints; eight images were shown in total. By showing their own images, we wanted to allow participants to pull on their knowledge of the image's purpose and slide context. We also wanted each participant to experience images of both classes (photograph and non-photograph) with each prototype. Therefore, if there were not enough images of one of the two types in the PowerPoints submitted by the participant, we inserted a public domain image that we selected of the correct type (e.g., if one person was short on non-photographs, we gave them a screenshot of a map). Four participants required one replacement image and two required two. In addition, to compare alt text generated across participants, we included an image of a young-looking adult with light brown skin and curly black hair with glasses smiling at the camera with an espresso in hand for all participants.

4.1.3 Procedure. The study consisted of two phases (Figure 3-top) and took place in one hour on a video conferencing platform (due to the COVID-19 pandemic). The first phase comprised a short interview about the participant's role at Microsoft, their experience with PowerPoint, what they know about alt text and automatic alt text, and experience creating alt text.

Then, in the second (usability testing) phase, participants were asked to visit a website hosting the designs we created and share their screen to allow us to examine their workflow. The usability testing phase included three tasks: 1) authoring alt text with the current PowerPoint interface to serve as a control, 2) authoring

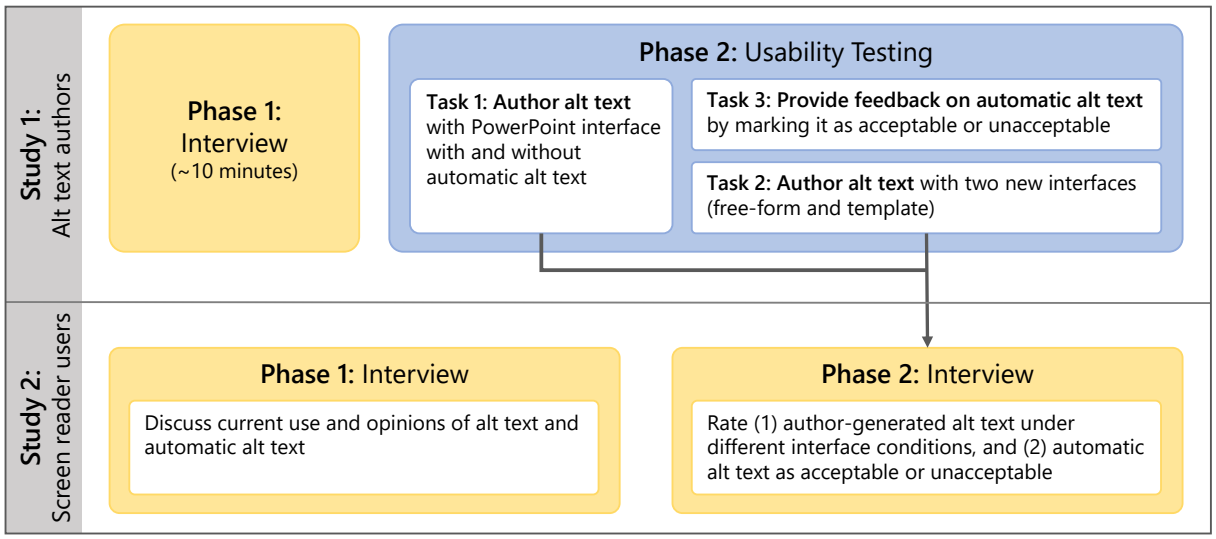


Figure 3: The study procedure that was followed for alt text author-participants (top) and SRU participants (bottom). Alt text authors began with a short background interview. They then performed usability testing with three tasks. The first for all participants was a control condition: authoring alt text in the current PowerPoint interface with and without automatic alt text. Task 2 was to author alt text using our two authoring interface designs. Task 3 was to judge the quality of automatic alt text for images using our two feedback interfaces. The order of interfaces shown in Tasks 2 and 3 were counterbalanced. The order of Tasks 2 and 3 were also counterbalanced across participants. SRUs had two-part interviews. In phase one, they discussed current use and opinions of alt text and automatic alt text. In phase two, they judged the quality of alt text that was generated by author-participants. In this phase, they also judged automatic alt text quality as acceptable or unacceptable.

alt text with two new interface designs, and 3) providing feedback about automatic alt text using another two interface designs.

Participants authored alt text with several interfaces for Tasks 1 and 2. Every session started with Task 1: authors composed alt text for four images using a replica of the PowerPoint interface where the text editing box was first blank and then pre-filled with PowerPoint's automatic alt text (Figure 4). For Task 2, authors generated alt text for three images with the two new interfaces, the free-form and template variants (Figure 2a and 2b). We counterbalanced the order in which participants saw these two interfaces in Task 2. Two images were held constant across all interface variants (a photograph of a young person drinking coffee, which was the same for all participants, and one non-photograph from a participant's own slides); the rest were unique to each interface (eight unique images total). In Task 3, participants used the feedback interfaces to judge the automatic alt text quality for four images in each interface. The order in which participants completed Tasks 2 (authoring) and 3 (feedback) were counterbalanced, as were the order in which the interface variants were displayed in tasks 2 and 3.

In each task, the alt text generated and interactions performed were logged to a server for later analysis. After participants finished interacting with an interface, we discussed its advantages and disadvantages. We concluded by asking which interface they preferred for each task.

4.2 Screen Reader User Interviews

4.2.1 Participants. We recruited employees at Microsoft who use screen readers by reaching out to disability-focused email lists. In total, we interviewed six SRUs: three men and three women. Their mean age was 38.8 (SD = 12.2), ranging between 26 and 54. All participants had experience using PowerPoint and alt text, and characterized their vision status as blind except one, who characterized themselves as low vision. The screen readers they used varied depending on platform and the task at hand and included VoiceOver, TalkBack, NVDA, Jaws, Microsoft Narrator, and a customized system. With respect to other assistive technologies used in computer usage, one participant used hearing aids and another used enlarged font sizes.

4.2.2 Procedure. In the second study, an hour-long interview via a video call, we interviewed our participants about their use of, experience with, and opinions of alt text across different contexts (Figure 3-bottom). We specifically asked about alt text in PowerPoint and opinions of using automatic versus human-authored alt text in the first phase of the interview. In the second phase, we sought to understand our interviewee's opinions of our alt text authoring interfaces and the alt text that was generated using those interfaces. We utilized the alt text generated in Tasks 1 and 2 during our alt text author usability testing study. We asked our SRU interviewees to rank multiple versions of alt text for a single image. The alt texts included the automatic alt text generated by PowerPoint and all of the alt texts for that image that an author interviewee generated with different interfaces. SRU participants were not told how the alt texts were generated. This process was repeated for three to four images per interviewee (depending on time). The images were randomly selected from our author interviews, though

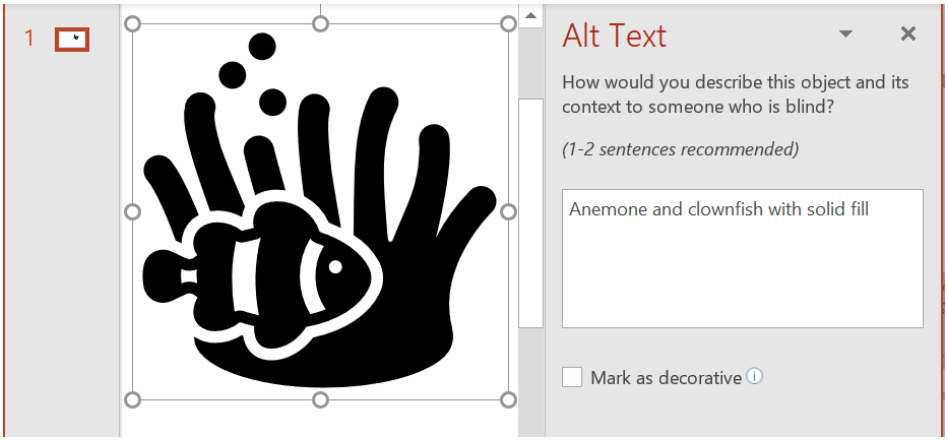


Figure 4: PowerPoint's current alt text editing pane. The text in the box on the right is automatically generated by an AI system. It was pre-populated in the text box before the pane was opened.

we did ensure that at least one person with more and one person with less alt text authoring experience was selected. We also asked participants to note if each alt text created was "acceptable" or not, which allowed us to compare SRU and alt text author views about the quality of automatically created alt text. After ranking the alt texts, we revealed the prompts that were provided to authors for the two images and asked SRUs if they thought any were missing or unnecessary.

4.3 Analysis

Our analysis consisted of two parts: analysis of the interview transcripts and analysis of the alt text generated by participants. To analyze the interview transcripts (both author and SRU), we performed thematic analysis, using an inductive approach but recognizing the reflexive nature of our analysis given the researchers' deep familiarity with alt text [2, 3]. In conducting our thematic analysis, we had three main research questions of interest: 1) what are author views of alt text and alt text authoring, 2) what are SRU preferences for what to include in alt text, and 3) what are authors' and SRUs' opinions of our interface designs and the resulting alt text. The lead author first reviewed all of the transcripts, extracting key information around these topics as well as other salient topics that emerged through the interviews (e.g., authors' views of the purpose of the automatic alt text). From this information, they then developed a set of codes which were grouped into themes. Two authors discussed these codes and broader themes and applied them to one randomly selected, full interview to verify the validity and completeness of the code book. The lead author applied the final set of codes to all transcripts.

We also wanted to assess and quantify the quality of the alt text generated by interviewees. We adopted an existing four-point scale to judge quality of the information presented in alt text from 1 (irrelevant or inaccurate) to 4 ("almost everything is described ... including information that might not be immediately apparent visually") [8]. Two authors independently ranked each alt text generated on this scale, and in the 30% of cases where they did not agree, a third author independently ranked the alt text, and the median value was selected.

5 RESULTS

Our findings contribute to understanding what makes quality alt text and how different interface designs can affect alt text quality. First, SRUs discussed what traits and information characterize high quality alt text; authors shared their views of alt text and their experience generating it. Second, we present the views from alt text author and SRU towards automatic alt text quality. We found that the perspectives of these two groups often do not match. Finally, we describe the impacts of our interface variations: small changes to the interface can affect willingness to edit alt text and alt text quality. We denote the author and SRU participants with the format A# and SRU#, respectively.

5.1 Current Understanding of Alt Text

We first asked about basic alt text preferences and challenges with SRU and author participants. SRUs highlighted that certain characteristics like accuracy, completeness, use of natural language, and conciseness are key to high-quality alt text. Accuracy was widely viewed as the most important factor, but SRUs varied in the prioritization of conciseness and use of natural language. While most authors were familiar with alt text's purpose, they often found authoring alt text challenging and they were unsure of what to include.

5.1.1 Screen reader user preferences. Our SRU participants discussed the key characteristics of high- and low-quality alt text, which varied based on personal preference. Certain characteristics including accuracy and completeness were critical for alt text regardless of image type. **Accuracy** refers to the fact that there is no wrong information in the alt text. **Completeness** was described as including all important aspects of an image for the context. Alt text that was too vague was unacceptable. For example, SRU3 commented on the lack of detail in a common automatic alt text in PowerPoint: *"[The] most annoying one is like 'screenshot of a cell*

phone' ... OK, I know that's kind of what it looks like, but that doesn't help." **Conciseness** was another quality that all but one of our participants preferred in alt text, as SRU3 explains: *"Being concise is also important, 'cause particularly when you're using a linear medium, like speech synthesis, you want to get the most information in the least [space] without any waste."* Four SRUs listed descriptiveness or level of detail as key qualities of good alt text, and two commented on the use of natural, flowing language[3] improving the quality of alt text. However, no participant prioritized naturalness of language over completeness, with two participants specifically stating that completeness is more important.

Descriptiveness vs. Conciseness. Oftentimes, descriptiveness and conciseness are in opposition. For example, when we asked SRU participants to rank versions of alt text for a picture of a person sitting with coffee, the two most popular versions were: "A woman with curly black hair, glasses, and a green sweater sits in a coffee shop or office. She has a cup of espresso in one hand, a saucer in front of her. She is leaning on one arm and looking at the camera, smiling slightly." and "a young lady looking at the camera sitting down drinking a cappuccino." Three participants preferred the former (more detailed) and two preferred the latter (concise). SRU3 and SRU4 agreed that conciseness is important, but they prioritized completeness or descriptiveness of the image over conciseness, as SRU4 explained: *"Probably like, on the [over-doing it] side. Just because I can always choose to skip that or, you know, consume whatever I want to as opposed to not having that information."* Overall, our participants were distributed across the spectrum of preferring very concise to very detailed alt text, suggesting that a one-size-fits-all alt text for an image may not be the best solution.

Important information. Participants listed key types of information that should be included in alt text, which varied based on image type. For photographs, participants were interested in knowing specific names of people and places, and personal characteristics of people like age, gender, and race. Participants listed more criteria for non-photographs, which several participants commented are more critical in the context of PowerPoint than photographs. SRU1 commented that graphs are critical to their productivity in their job: *"This is why you don't see a lot of persons with disability going into very highly technical profession because of a lot of graphical illustration ... are very, very inaccessible still."* Half of participants commented that trends or summaries of the takeaway for charts or graphs should be included, while SRU4 suggested that linking to the data-source to allow for self-investigation would be helpful. Indeed, SRU2 noted that they work on a data analysis-focused team, and they often turned to the raw data used to generate a plot if it has no or poor quality alt text. SRU3 commented that the level of detail required for non-photographs can be overwhelming if forced into a single string, suggesting that a more complex, structured datatype could be useful. Different characteristics are important to include in the alt text depending on image type; particularly for non-photographs, adding more guidance and structure to alt text may be beneficial.

[3]Some captions used by popular platforms include lists of tags in the alt text as opposed to a grammatically correct sentence. An example of this type of alt text from PowerPoint is "A picture containing swimming, table, blue, bird."

Regardless of image type, several participants mentioned the importance of context in alt text. Half of participants mentioned that the purpose an image serves in a given PowerPoint must be included in the alt text, with SRU1 repeatedly questioning: *"Why [does] the author of that article or that piece of collateral choose this image? How is it relevant to what I'm reading?"* In other contexts, like in interface design projects, very visual details (e.g., layout) can be key. Further, three participants mentioned it being difficult to judge the quality of alt text without knowing the broader context of why that image was used. It seems, then, that context is key for deciding what should be included in alt text as well as when judging if the alt text is satisfactory.

Non-desirable information. Some types of information were not desired in alt text by SRUs. For example, SRU4 explained that sometimes more descriptions of visual elements could be unnecessary for them, *"Oh my God, I like, I don't care about the color ... They go to [great] detail."* SRU2 similarly commented that the more specific visual details can be abstracted out of the description. Specifically, for an image where sets of red concentric circles were overlaid on parts of an image to represent noise, SRU2 commented: *"I don't need details like, say, instead of saying red concentric rings ... It's ok to just say that it's indicating noises."* Two participants described that they wouldn't want to consume any alt text for certain types of images altogether. SRU2 thought that repeated elements in slides like a company logo can simply be marked as decorative, even though they are not *decorative* in traditional definitions of alt text. SRU5 suggested that another class of images, similar to "decorative," can similarly be annotated as "alt text unnecessary." This class of images includes those that do not provide additional information to the article or slide the image is a part of. SRU5 explained this category: *"I have to say that to me, if there's meaningful text and [the associated image] doesn't add value to the definition of what is being described or discussed in a slide, then [I] don't need to know it. It's clutter."* These perspectives suggest that a simple dichotomy of "decorative" versus needing detailed alt text may not be sophisticated enough to meet SRU needs.

5.1.2 Author understanding of alt text. Though most of our author-participants were familiar with alt text, a few participants expressed uncertainty about alt text's purpose. Half (n = 6) of our author-participants thought that alt text was used by SRUs or individuals who were blind or low vision, and five participants commented that it was used broadly to improve accessibility. Two participants commented that it was supposed to help people with other differences, including language barriers or learning disabilities. Surprisingly, two participants mentioned non-accessibility uses of alt text, with one using it as a generic description or caption and another using the field to put attribution information (e.g., the photographer). These results suggest that the current simple description that accompanies the alt text field in PowerPoint, "How would you describe this object and its context to someone who is blind?" may be insufficient or overlooked in current alt text workflows.

What information to include. Authors' views for what should be included in alt text fell into two main (non-exclusive) categories: summary information and specific details. Summary information was meant to provide a broad overview of the image for the consumer. Participants specifically mentioned that they included

prominent features (n = 3), the "story" that the image conveys (n = 2), a general overview (n = 1), or what they (the author) would want to know about the image if they could not see it (n = 4). The more specific details that participants discussed included names of prominent individuals or places (n = 2) and personal characteristics of people in photos including gender, race, or style of clothing (n = 4). Finally, participants mentioned two types of information that they include in alt text that do not purely describe visual elements of the image; six participants added additional context (e.g., relevant background information) for the image, and five participants included information about why the photo was included in the PowerPoint or the "main takeaway" that they wanted their consumer to receive from the image. For example, A6 explains: *"Probably the first sentence is what I perceive given all the background knowledge I have about that figure and peripherals."* Only two participants (both of whom were experienced in writing alt text) commented that the alt text that they would provide for an image depends on the broader message of what they want to convey to the consumer.

Difficulty in writing alt text. Several author participants across alt text experience levels conveyed uncertainty about what to include in alt text. A3 questioned: *"What's alt text versus what's a caption? I don't really know."* They further stated that, when writing alt text, they are not sure of *"what kind of context to put in it, [and] where to begin and end."* Particularly with complex images, participants tended to express uncertainty for how to describe them. For a non-photograph infographic with three separate sub-figures (Figure 5-left), A7 recognized that she may not have enough understanding of what SRUs want to hear about such an image: *"I should talk to somebody [who] has accessibility issues to understand what they would like to see as opposed to what I can see ... how do you explain a complex image like that really and give it enough detail that they would understand what you're trying to convey in the different tiles?"* Regardless of expertise, people were unsure of how to write or what to include in alt text.

5.2 Automatic Alt Text

Automatic alt text often comes prepopulated in an application's alt text field (e.g., PowerPoint; Figure 4). When we discussed automatic alt text with participants, we found that SRUs thought it was better than nothing, but that it still needed to improve. Authors also felt the automatic alt text needed improvement and was often wrong. Still, authors preferred to be shown the automatic alt text. However, we saw that alt text quality was often lower when it was modified by editing automatic alt text rather than composed starting with a blank box. Additionally, when discussing what "acceptable" alt text is for SRUs and authors, we saw that their standards were not the same; authors often ranked SRU-defined unacceptable alt text as "acceptable."

5.2.1 SRU perspectives of automatic alt text: improving, but lacking. SRU participants were wary of the quality of automatic alt text, though half (n = 3) of them stated that it is better than nothing. On the other hand, half (n = 3) of the participants discussed that the automatic alt text is often not very accurate, especially in the cases of non-photographs; three participants commented on how screenshots were often described as "a screenshot of a cellphone" in PowerPoint. SRU5 noted concerns regarding an AI system that is getting better but is not yet fully accurate: *"A lot of times it's going to be very accurate. Sometimes it's not and I think that what this creates is a false sense of security in that users may or may not know that that is an actual valid description."* SRU6 further commented that the automatic alt text can only improve so much in quality, since an AI system does not have a good grasp of the context of the image or the reason it was included. Overall, while many SRU participants commented on the improving automatic alt text, it is still unacceptable in many cases (particularly infographics) and is subject to bias.

5.2.2 Authors prefer having automatic alt text as a starting point. In general, sighted author sentiment towards PowerPoint's automatic alt text was negative. Most author participants who commented on its quality fell between "usually unhelpful" and "totally wrong." However, there were a few photographs where author participants commended the system for a high-quality automatic alt text. Interestingly, despite the overall low opinion of the quality, the majority (n = 8) of author participants preferred having the prepopulated automatic alt text suggestion in the PowerPoint interface.

Overall, author participants preferred having the automatic alt text because it provided assistance in the alt text writing process; nine participants appreciated that it was a "starting point" for crafting their own alt text. This assistance made the process faster, easier, and less work. As A3 explained: *"What I like about the AI suggestions is that it makes me feel like somebody is helping me and I like that because I'm more inclined to put the effort into it because I'm like, 'Oh, you don't have to do it alone.' "* Participants stated that automatic suggestions were worth including if they were right some of the time; if it was wrong, it still was not a large cost to them.

5.2.3 The effect of automatic alt text. We compared alt text that was generated with a blank text box and when a text box was prefilled with automatic alt text. By final alt text, we refer to the automatic alt text post-user editing. Table 2 presents an example of the alt text A6 wrote for an image when starting from a blank textbox and a textbox prefilled with the automatic alt text. Interestingly, gender was often excluded when starting from the automatic alt text even though every participant included gender in the very first description they wrote for the shared image of a person sitting at a table with a cup of coffee (without seeing the automatic alt text).

More generally, we found that the presence of automatic alt text as a starting point negatively affected final alt text quality. We ranked the quality of the alt text generated for two images for each participant with and without the presence of the prefilled automatic alt text (24 pairs of alt text total). Overall, the average alt text quality rating when starting from a blank text box (M = 2.96, SD = .81) was higher than that when starting from the automatic alt text (M = 2.38, SD = .50). The difference in the final alt text quality between the two interface conditions was significant (Wilcox signed-rank test: Z = 55, p < .01).

Author participants' reflections provided insight into why automatic alt text may have affected quality: it was viewed as a "gold standard" of good alt text. For example, A3 reflected on their experiences writing alt text with and without the automatic alt text present. When starting from scratch, they wrote 1-2 sentences for each image. They wrote only one sentence when starting from the automatic alt text. They reflected: *"In the automatic alt text one,*

Figure 5: Two of our participants' images. The left is an example of a complex image. It has three sub-figures, each containing a small infographic. The participant indicated that she was unsure how to best write alt text for this image. The alt text created for the right photo, the Taj Mahal, was incomplete. "White building on a green lawn with a walk way leading up to people gathered around" omits the name of the building.

Table 2: A sample of the alt texts one author created in the study. The author wrote alt text for this image a total of four times. The alt texts shown in this table were generated in Task 1 by starting from a blank text box and starting from a box prefilled with the automatic alt text. They provided considerably less detail for alt text when starting from the automatic alt text.

Automatic alt text	Author alt text starting from blank interface	Author alt text from automatic alt text
A person sitting on a table	A young lady with dark curly hair and glasses, sitting down at a coffee table. She is holding an espresso cup with her right arm and leaning her head on her left hand.	A young female person sitting on a table, smiling at the camera.

I think [the provided automatic alt texts] were single sentences ... There was no expectation to make them anything more than that." Similarly, A4 commented that they made their descriptions more vague, since the automatic alt text lacked detail. One participant, A8, commented that they were wary of automatic alt text being interpreted as the "gold standard" of alt text quality. In their opinion: *"It's probably better not to have a prefilled generic description [by default in PowerPoint]. In certain situations, that just erases out important characteristics like demographic, ethnicity, or characteristics of the people described in the images. At scale, I think what that would lead to is for more people, just to accept that as the de facto description, and I think that's not fair and not okay."* In these examples, we found that the automatic alt text acted as a standard of quality for many of our author participants, leading to shorter, less detailed final alt text.

5.2.4 Characteristics of "acceptable" automatic alt text. When asked what characteristics are key for "acceptable" automatic alt text, SRU participants were most adamant about the accuracy and completeness. Five SRUs mentioned inaccurate alt text was unacceptable, with two participants saying this is "misleading." Three SRUs commented on the fact that, for a photo of a person sitting at a table, the automatic alt text said they are sitting *on* the table rather than *at* the table: *"that's a very different picture"* (SRU6). The other key characteristic of "acceptable" alt text was completeness of information (n = 3). Three SRUs were put off when reading the alt text written for the Taj Mahal and some versions did not include the name of the building (Figure 5-right). Overall, it was apparent that accuracy and completeness were critical to acceptable alt text, but not necessarily characteristics like "using natural language."

We found that our author and SRU participants' definitions of "acceptable" alt text differed. While a few authors mentioned looking for accuracy and completeness in alt text, they did mention that their "bar was pretty low," or they expected the AI system to incorporate "just the basics" of the image. After authors rated the automatic alt text for several images, we saw author-rated "acceptable" alt text was not always complete or accurate. For example, A3 said that the automatic alt text "A large crowd of people" was acceptable for a picture, when this alt text left out key information: everyone was dressed in white and red for a festival. The author did include this detail when writing their own alt text from scratch for this photo, but rated the incomplete automatic alt text as acceptable. Regarding inaccuracy, eight of the 11 authors[4] rated the alt text "a person sitting on a table" as acceptable. This same alt text was rated unanimously as unacceptable by our SRU participants because the person is not sitting *on* the table. In summary, there was a mismatch between SRU and author definitions of acceptability for automatic alt text.

5.3 Effects of Interface Variants

Author participants tested interface variants both for providing feedback about the quality of automatic alt text and for authoring

[4]One author ran out of time and did not complete this exercise.

alt text. The feedback interfaces were well received and noted as fast and easy. However, some participants reported they would be less likely to edit the automatic alt text to make it acceptable if it was needed. Authoring interfaces provided the authors with suggestions for what to include in the alt text, which was valued by participants. The interface variants led to higher quality alt text half of the time (n = 12) compared to the original PowerPoint interface; in the other half, interface variants either produced the same (n = 9) or lower quality (n = 3) alt text than the PowerPoint interface.

5.3.1 Author feedback mechanisms for automatic alt text. Two interfaces asked participants to determine if automatic alt text was acceptable or unacceptable; one used thumb icons located directly below the image (Figure 2d), and the other used check boxes in the existing alt text edit pane (Figure 2c). Author participants expressed positive sentiments towards the feedback interface, stating that the interface was fast (n = 3) and easy to use (n = 6). Irrespective of feedback interface style, three authors appreciated the opportunity to provide feedback to the system. The majority of participants (n = 8) preferred interacting with the icon-based interface overall, often because of the naturalness of the icons and the limited reading required. With respect to feedback location, opinion was split. Six participants preferred the pane location for reasons such as freeing up slide space or mimicking PowerPoint interface styles; five preferred the feedback positioned below the image, often commenting on the convenient close proximity to the image. The positive attitudes towards the feedback interface suggest that this interface alteration could provide an engaging method to collect feedback from PowerPoint authors about the quality of automatic alt text, which could be valuable for training future iterations of vision-to-language systems.

Interestingly, we found that surfacing the feedback interface affected self-reported likelihood of editing alt text after providing feedback. Two participants who said it would increase their engagement reasoned that the feedback draws their attention to errors in the alt text. On the other hand, two participants reported they would be less likely to edit the alt text, since they had already provided the system feedback on the alt text via the feedback interface: *"If I'm telling them that this is not right, I would expect them to ... understand that they need to fix it. I wouldn't add any text."* (A5). Further research (preferably a field deployment) with larger samples should investigate if and how asking for feedback impacts automatic alt text editing rates.

5.3.2 Alt text authoring interfaces. Two interface variants supported authors in generating alt text from scratch. Both variants provided suggestions and examples for what to include in alt text. However, the *free-form interface* provided a single text box for entry (Figure 2a), whereas the *template interface* asked participants to respond to each suggestion separately (Figure 2b). Participants commented that the suggestions helped them better structure (n = 8) and know what should be included in the alt text (n = 7): *"It's kind of nice to have that detail ... so you're just not out there, like 'what should I put?' or 'what do they want me to talk about?' So, you know that you're going in on the right areas ..."* (A4). Additionally, the suggestions made five participants look closer at the image and/or increase the detail that they felt they included in the alt text. However, one participant pointed out that the suggestions are only helpful if they are relevant to the image, indicating that the AI algorithm that selects the tips to show would need to be accurate.

Subjective preference. Experience with alt text influenced which interface the participant preferred. Five participants preferred the free-form interface, all but one of whom had experience writing alt text, and four preferred the template interface, all of whom had either never authored alt text in PowerPoint before or had little experience doing so. These preferences indicate that the extra scaffolding of the template interface may be beneficial to people with little experience with alt text.

People generally had positive sentiments towards the free-form interface. Although one participant was adamant that this interface had too much text, the other participants appreciated that it increased the detail they included in the alt text and helped them know what to include. People had stronger feelings towards the template interface. Authors who preferred the free-form interface commented that the template interface was too slow, took too much effort, and was tedious. Additionally, they did not like having to click through multiple boxes, especially when the fields were prepopulated with automatic alt text. Of the participants who preferred the template interface, three stated that it felt easier to respond to each suggestion rather than fill in an empty box.

Quality of alt text. The interface variations affected the quality of the alt text when measured by the researchers using the four-point scale [8] as described in Section 4.3. We compared the alt text generated by a participant for an image with three interfaces: 1) the current PowerPoint interface (Figure 4); 2) the free-form interface (Figure 2a); and 3) the template interface (Figure 2b). We did this for two images per participant (24 images total). The results are summarized in Figure 6. Overall, there were nine instances (five green circles for Template + four for Free-form in Figure 6) where the interfaces we created resulted in higher quality alt text; however, the PowerPoint interface did outperform the interfaces we created three times.

Six SRU participants also ranked the quality of alt text generated under these three authoring interface conditions. Each SRU ranked multiple versions of alt text for three different images (with six participants, this is 18 rankings total). For one image, the PowerPoint interface generated the alt text favored by the most participants (n = 4), and in the other two images, the free-form interface generated the most favored alt text (n = 4 and n = 3, respectively). These data suggest that the tips interface encouraged alt text authors to create text that is more closely aligned to SRU preferences.

6 DISCUSSION AND FUTURE WORK

In this section, we dive deeper into the results and takeaways of our studies. In summary, our results highlight the importance of end-user customization in future alt text solutions. We also found gaps between what SRUs require in alt text (e.g., context) and what authors think to include. We discuss which interface changes are valuable to make now (e.g., adding suggestions for alt text authors) and those that need further studying before integration in mainstream systems (e.g., feedback interfaces). Finally, we briefly reflect on the limitations of our studies and the need of a field deployment study.

Figure 6: The rankings for the alt text created for 24 images with three interfaces each: overall, the alt texts generated with the template interface were rated the best while the alt texts with the free-form one were slightly better than those composed with PowerPoint's default interface. Each circle represents the quality of the alt text for one image under one interface variant; therefore, each row (interface) has 24 circles with the number in the circle indicating the image id. The column groupings indicate the absolute quality of the alt text on a 1 (worst) to 4 (best) scale. The color of the circle indicates the relative quality of the alt text compared to alt text generated by the other interfaces for the same image. If an alt text was the best of all three interface variants for an image, the circle is green. Many times, there was a tie in quality, which are represented by purple and orange color.

6.1 One-Size-Fits-All Solution is Suboptimal

Our findings support the need for personalized alt text, and suggest that a one-size-fits-all solution is suboptimal. SRUs differed in key dimensions of alt text preferences, including what information to include in alt text (e.g., colors, names), the level of detail of alt text (e.g., some SRUs considered level of detail more important than conciseness), and even what images should have alt text (e.g., two SRUs noted classes of images for which they would not desire to hear alt text). Customization of alt text is a challenging problem; given that human time and attention is scarce, it is likely unreasonable to ask an author to write multiple versions of alt text (e.g., a detailed and a summarized version).

One potential method to support customization is to encourage annotating portions of alt text so that information can be included or excluded easily. For example, attributes about the scenery or the colors present in a photo could be annotated as such. SRUs could select the classes of information they care about and exclude the others by default. This solution echos the ideas from SRUs of having more metadata about visual descriptions that supports drilling down to find specific types of information (e.g., the trend of a graph). We explored one method of annotation with our template authoring interface. Each text box in the template inherently received annotation via the associated prompt (e.g., "what was the main purpose of this image?"). However, this solution was not perfect. The words written in the different text boxes of the template were often redundant or fragmented and hard to understand when the alt text was read by SRUs.

As an alternative to author-generated labels, tools could be built to allow crowd workers to assign annotations to specific parts of the alt text. A dataset of enough of these labels could be used to train an AI system to annotate the captions. Future research should investigate what categories of annotations would be valuable for SRUs (e.g., graph trends, colors, layout, subject), building upon existing work [1, 19].

Finally, AI may be able to help generate summarized forms of alt text using natural language processing techniques (e.g., extraction-based text summarization) based on user preferences. However, it may be challenging to eliminate specific types of descriptions (e.g., colors, layout) and maintain the grammatical correctness of the sentence.

6.2 Context is Key

SRU participants highlighted the importance of including context (i.e., why an image was included in a slide deck) in the alt text, which was not a trait considered by many of our author participants. This insight expands prior work that notes that the context of an image (e.g., e-commerce site, dating app) determines how the image should be described [19]. SRUs' desire to understand the reason an image was used implies that people who did not author the slide deck lack all of the information required to write high-quality (or judge the quality of) alt text. Even so, crowd workers or AI systems may still hold a place in the alt text workflow. For example, authors can be encouraged to include information that only they know (e.g., why an image was used) over other types of information. Then, information that is purely visual description (e.g., describing what is in the image) can be created by crowd workers or AI systems. Indeed, the crowd worker or AI generated captions may be higher quality if they know the image's purpose.

One way to make aspects of context explicit is to create classes of "image purposes." "Decorative" is an existing example of denoting image purpose in alt text. Other options for the purpose an image serves could include "presenting new information" or "visually reaffirming textual content." Asking authors to annotate the image purpose via checkbox (in addition to writing alt text) could provide multiple benefits. First, this metadata would support screen reader customization; a reader could specify that they only want to hear alt text for images that contribute new information to a slide. Second, the image purpose could influence suggestions for what to include in the alt text. For example, images that are included to reaffirm textual content may need brief, high level descriptions, but images that contribute new content may need detailed alt text. Exploring the different classes of purposes that images serve and how the alt text should change based on image purpose should be investigated in future work.

6.3 Interface Changes Affect Alt Text Engagement and Quality

We found that interface variants used in our study affected authors' willingness to engage with the alt text (i.e., through feedback and/or edits) and also affected the final alt text quality. Regarding

feedback interfaces for automatic alt text, the lightweight interface encouraged more interaction for many participants. However, some participants mentioned that they would not be willing to both provide feedback *and* edit the automatic alt text. For this reason, field deployments need to compare alt text editing behavior before and after a feedback interface is introduced before any platform adopts such an interface at scale.

Similarly, the authoring interface variations affected alt text engagement and quality. Both of our authoring interface designs provided tips for what to include in alt text, improving ease of text creation and author confidence. Our interface designs on average increased the alt text quality as judged by both a qualitative alt text scale and SRU rankings, though our sample of SRUs was small (n = 6) and time allowed for ranking of only three images. While future studies should investigate if this trend persists with a larger sample size, our results show promise for integrating such a solution into content creation platforms to encourage higher quality alt text production that better matches SRU expectations.

Our solution also suggests the use of a relatively simple AI that recognizes the difference between photographs and non-photographs, surfacing tailored prompts accordingly. Even without the application of AI, simple prompts encouraging authors to include the subject of the image and the main takeaway and/or purpose of the image might lead to higher quality alt text. Experienced authors may do well with a simple list of suggestions. However, authors that are less experienced may benefit from a more structured approach that probes for specific information, as supported by prior alt text template research [13].

Finally, our findings suggest that automatic alt text affected the quality of the final alt text. Therefore, consideration should be taken as to how and when to surface it. In our study, presence of automatic alt text never increased, and often decreased, the final alt text quality. Participant comments suggest that they viewed automatic alt text as a quality standard that needed to be met. While this view had one positive implication in our study (it steered participants away from assuming gender identities of photographees [1]), it also led to generally less detailed alt text. Educating users about the purpose and limitations of automatic alt text could help correct this view of automatic alt text as a gold standard. One way to educate users would be to encourage them to check for completeness and accuracy of the automatic alt text or prompt them to make specific edits, like adding information that SRUs desire but is challenging for an AI to recognize (e.g., context). Without such instructions, platforms should consider if it is appropriate to surface automatic alt text to authors; interface designers must balance the desire to not prime authors with low quality automatic alt text with the fact that without automatic alt text, there may be a higher incidence of images with no alt text at all.

6.4 Authors Need a Better Understanding of High Quality Alt Text

We found a clear mismatch between SRU and author understandings of "quality" alt text. Though SRUs prioritized traits such as accuracy and completeness, over half of the author participants ranked an inaccurate alt text as acceptable. These results indicate that, until differences between authors' and SRUs' mental models of quality alt text are resolved, authors might not provide accurate ratings about the quality of automatic alt text. On the other hand, allowing content authors to provide insights into the quality of automatic alt text is critical for building systems that are resilient to errors that are hard to diagnose with AI (e.g., avoiding offensive mistakes). A feedback interface that lists the necessary criteria of alt text quality for SRUs may help authors provide more accurate feedback about quality.

6.5 Study Limitations

Our work provides initial insights into the design of more effective and engaging methods of authoring alt text: our sample of 12 author interviewees from a technology company was enough to validate interface concepts. However, our study measured effectiveness and opinions through interview and usability testing sessions. While our participants reported that they would engage with alt text more given our interfaces, this cannot be validated without a broader field study, which would also allow for measures of actual instead of predicted use. In addition, experiments with a larger and more diverse set of participants would be beneficial 1) to confirm generalizability beyond information workers, 2) to understand how author expertise can affect interface engagement and preferences, and 3) if adjustments need to be made to these interfaces in non-PowerPoint contexts. We note that, though our results were scoped to the context of PowerPoint, we suspect that our findings generalize to other contexts with minor adjustments. For example, the same alt text interface is utilized by most Microsoft Office products, and a similar pane as shown in this work could be integrated into web interfaces as well. Finally, future work should investigate how to support SRU feedback about automatic alt text, as end user feedback is critical for improving AI systems.

7 CONCLUSION

In this work, we developed interface variants to facilitate authoring alt text and providing feedback for automatic alt text in Microsoft PowerPoint. We performed both combined interview and usability testing sessions with 12 sighted alt text authors, and interviews with six SRUs. Through our analysis, we found that authoring interfaces that support the authors in choosing what to include in the alt text were both well received by authors and resulted in higher quality alt text compared to the current interface (as judged by the researchers and SRUs). Interview results about automatic alt text and key aspects of "high-quality" alt text revealed a gap in alt text author and SRU opinions about alt text quality; these results suggest that alt text authors, particularly those who do not know the context for an image, may not provide accurate feedback about automatic alt text quality. Finally, we found a significant difference in quality of alt text generated from scratch versus by editing automatic alt text, suggesting that current author perceptions of and engagement with automatic alt text result in lower quality alt text. We hope that this work influences the construction of alt text authoring interfaces and the use of feedback mechanisms and automatic alt text in the future.

ACKNOWLEDGMENTS

We would like to thank all of our participants for their time and insights and Michael Barnett for his engineering expertise and support.

REFERENCES

[1] Cynthia L. Bennett, Cole Gleason, Morgan Klaus Scheuerman, Jeffrey P Bigham, Anhong Guo, and Alexandra To. 2021. "It's Complicated": Negotiating Accessibility and (Mis) Representation in Image Descriptions of Race, Gender, and Disability. (2021).

[2] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[3] Virginia Braun and Victoria Clarke. 2019. Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health* 11, 4 (2019), 589–597.

[4] Diagram Center. 2019. *General Guidelines.* http://diagramcenter.org/general-guidelines-final-draft.html

[5] Diagram Center. 2019. *Specific Guidelines: Art, Photos& Cartoons.* http://diagramcenter.org/specific-guidelines-final-draft.html#20

[6] Kay Alicyn Ferrell, Silvia M Correa-Torres, Jennifer Johnson Howell, Robert Pearson, Wendy Morrow Carver, Amy Spencer Groll, Tanni L Anthony, Deborah Matthews, Bryan Gould, Trisha O'Connell, et al. 2017. Audible image description as an accommodation in statewide assessments for students with visual and print disabilities. *Journal of Visual Impairment & Blindness* 111, 4 (2017), 325–339.

[7] Chancey Fleet, Cynthia L. Bennett, Venkatesh Potluri, and Meredith Ringel (panel moderator) Morris. 2020. *CVPR 2020 VizWiz Grand Challenge Workshop – Panel Discussion with Blind Technology Experts.* https://www.microsoft.com/en-us/research/video/cvpr-2020-vizwiz-grand-challenge-workshop-panel-discussion-with-blind-technology-experts/

[8] Cole Gleason, Patrick Carrington, Cameron Cassidy, Meredith Ringel Morris, Kris M. Kitani, and Jeffrey P. Bigham. 2019. "It's Almost like They're Trying to Hide It": How User-Provided Image Descriptions Have Failed to Make Twitter Accessible. In *The World Wide Web Conference* (San Francisco, CA, USA) *(WWW '19)*. Association for Computing Machinery, New York, NY, USA, 549–559. https://doi.org/10.1145/3308558.3313605

[9] Cole Gleason, Amy Pavel, Xingyu Liu, Patrick Carrington, Lydia B. Chilton, and Jeffrey P. Bigham. 2019. Making Memes Accessible. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility* (Pittsburgh, PA, USA) *(ASSETS '19)*. Association for Computing Machinery, New York, NY, USA, 367–376. https://doi.org/10.1145/3308561.3353792

[10] Cole Gleason, Amy Pavel, Emma McCamey, Christina Low, Patrick Carrington, Kris M. Kitani, and Jeffrey P. Bigham. 2020. Twitter A11y: A Browser Extension to Make Twitter Images Accessible. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376728

[11] Darren Guinness, Edward Cutrell, and Meredith Ringel Morris. 2018. Caption Crawler: Enabling Reusable Alternative Text Descriptions Using Reverse Image Search. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3173574.3174092

[12] Haley MacLeod, Cynthia L. Bennett, Meredith Ringel Morris, and Edward Cutrell. 2017. Understanding Blind People's Experiences with Computer-Generated Captions of Social Media Images. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 5988–5999. https://doi.org/10.1145/3025453.3025814

[13] Valerie S. Morash, Yue-Ting Siu, Joshua A. Miele, Lucia Hasty, and Steven Landau. 2015. Guiding Novice Web Workers in Making Image Descriptions Using Templates. *ACM Trans. Access. Comput.* 7, 4, Article 12 (Nov. 2015), 21 pages. https://doi.org/10.1145/2764916

[14] Meredith Ringel Morris, Jazette Johnson, Cynthia L. Bennett, and Edward Cutrell. 2018. Rich Representations of Visual Content for Screen Reader Users. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3173574.3173633

[15] Meredith Ringel Morris, Annuska Zolyomi, Catherine Yao, Sina Bahram, Jeffrey P Bigham, and Shaun K Kane. 2016. " With most of it being pictures now, I rarely use it" Understanding Twitter's Evolving Accessibility to Blind Users. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 5506–5516.

[16] World Health Organization. 2010. *Global data on visual impairment.* https://www.who.int/blindness/publications/globaldata/en/

[17] Kyle Reinholt, Darren Guinness, and Shaun K. Kane. 2019. EyeDescribe: Combining Eye Gaze and Speech to Automatically Create Accessible Touch Screen Artwork. In *Proceedings of the 2019 ACM International Conference on Interactive Surfaces and Spaces* (Daejeon, Republic of Korea) *(ISS '19)*. Association for Computing Machinery, New York, NY, USA, 101–112. https://doi.org/10.1145/3343055.3359722

[18] John Roach. 2020. *What's that? Microsoft's latest breakthrough, now in Azure AI, describes images as well as people do.* https://blogs.microsoft.com/ai/azure-image-captioning/?utm_source=stories-pointer

[19] Abigale Stangl, Meredith Ringel Morris, and Danna Gurari. 2020. "Person, Shoes, Tree. Is the Person Naked?" What People with Vision Impairments Want in Image Descriptions. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376404

[20] Violeta Voykinska, Shiri Azenkot, Shaomei Wu, and Gilly Leshed. 2016. How blind people interact with visual content on social networking services. In *Proceedings of the 19th acm conference on computer-supported cooperative work & social computing*. 1584–1595.

[21] W3C. 2019. *An alt Decision Tree.* https://www.w3.org/WAI/tutorials/images/decision-tree/

[22] W3C. 2019. *Guideline 1.1 – Text Alternatives.* https://www.w3.org/WAI/WCAG21/quickref/?versions=2.0#text-alternatives/

[23] W3C. 2019. *Images Concepts.* https://www.w3.org/WAI/tutorials/images/

[24] WebAIM. 2020. *The WebAIM Million.* https://webaim.org/projects/million/

[25] WebAIM. 2021. *Alternative Text.* https://webaim.org/techniques/alttext/

[26] Qi Wu, Peng Wang, Chunhua Shen, Anthony Dick, and Anton Van Den Hengel. 2016. Ask me anything: Free-form visual question answering based on knowledge from external sources. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4622–4630.

[27] Shaomei Wu, Jeffrey Wieland, Omid Farivar, and Julie Schiller. 2017. Automatic Alt-Text: Computer-Generated Image Descriptions for Blind Users on a Social Network Service. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, Oregon, USA) *(CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 1180–1192. https://doi.org/10.1145/2998181.2998364

[28] Yuhang Zhao, Shaomei Wu, Lindsay Reynolds, and Shiri Azenkot. 2017. The effect of computer-generated descriptions on photo-sharing experiences of people with visual impairments. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 1–22.

[29] Hong Zou and Jutta Treviranus. 2015. ChartMaster: A Tool for Interacting with Stock Market Charts Using a Screen Reader. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility* (Lisbon, Portugal) *(ASSETS '15)*. Association for Computing Machinery, New York, NY, USA, 107–116. https://doi.org/10.1145/2700648.2809862