Is Generative AI Mature for Alternative Image Descriptions of STEM Content?

Marina Buzzi[1][a], Giulio Galesi[2][b], Barbara Leporini[2,3][c] and Annalisa Nicotera[3]

[1]*IIT-CNR, Institute of Informatics and Telematics, National Research Council, Pisa, Italy*
[2]*ISTI-CNR, Institute of Information Science and Technologies, National Research Council, Pisa, Italy*
[3]*University of Pisa, Largo B. Montecorvo, Pisa, Italy*
marina.buzzi@iit.cnr.it, giulio.galesi@cnr.it, a.nicotera2@studenti.unipi.it, barbara.leporini@unipi.it

Abstract: Alternative descriptions of digital images have always been an accessibility issue for screen reader users. Over time, numerous guidelines have been proposed in the literature, but the problem still exists. Recently, artificial intelligence (AI) has been introduced in digital applications to support visually impaired people in getting information about the world around them. In this way, such applications become a digital assistant for people with visual impairments. Increasingly, generative AI is being exploited to create accessible content for visually impaired people. In the education field, image description can play a crucial role in understanding even scientific content. For this reason, alternative descriptions should be accurate and educational-oriented. In this work, we investigate whether existing AI-based tools on the market are mature for describing images related to scientific content. Five AI-based tools were used to test the generated descriptions of four STEM images chosen for this preliminary study. Results indicate that answers are prompt and context dependent, and this technology can certainly support blind people in everyday tasks; but for STEM educational content more effort is required for delivering accessible and effective descriptions, supporting students in satisfying and accurate image exploration.

1. INTRODUCTION

Graphic content, such as photos, diagrams, graphs, etc., have always been a problem to be addressed in the accessibility field. Such elements, in fact, can be a challenge for people with visual impairments, and especially for screen reader users. Alternative description is often the easiest and fastest way to provide information about an image to a reader who cannot access it visually.

The W3C Web Content Accessibility Guidelines include recommendations for non-textual content (WCAG2.2, (https://www.w3.org/TR/WCAG22/). More specific recommendations have been proposed to match alternative descriptions to images for screen reader users [DC 2012], [Lundgard 2021]. Despite this, in a recent study conducted by interviewing twenty visually impaired people, image Captioning,

Personal Object Recognition, Safe Path Detection and Fill Paper Forms all received a highest score as tasks of their interests [Gamage 2023]. One participant emphasized the importance of accurate Image Captioning, to render them able to comprehend the visual content.

To make digital images accessible to screen reader users, image descriptions (alternative text) need to be added for describing the information contained within the image. Unfortunately, this might be time-consuming because (1) the system (website, document, eBook, etc.) might not be designed to allow the user to easily add the alternative text to each graphical item; (2) a person must prepare the description of each image; (3) the developer/operator needs match alternative text to each image. Developers might not remember to add alternative text, not have time to add it, or may not know what to

[a] https://orcid.org/0000-0003-1725-9433
[b] https://orcid.org/0000-0001-6185-0795
[c] https://orcid.org/0000-0003-2469-9648

include when writing the descriptions [Gleason 2019]. So, these activities require time, expertise and tools [Mack 2021]. This is why images do not frequently have meaningful accessible descriptions.

Complex images contain substantial information – more than can be conveyed in a short phrase or sentence. Examples are graphs, charts, diagrams, etc. Web accessibility guidelines from W3C provide best-practices for writing descriptions of "complex images", either in a short description alt text attribute, or as a long textual description displayed alongside the visual image [WAI 2022]. The description for complex images is crucial in the science, technology, engineering, and mathematics (STEM) content.

Publishers have developed guidelines for describing graphical elements appearing in STEM materials [Lundgard 2021]. However, visual data or content in scientific articles are still not accessible to people with visual disabilities [Sharif 2021], [Splendiani 2014].

Students need to learn what complex images represent in term of important educational concepts (i.e. the semantic meaning). For example, a student of computer science needs a description of a graph, an automaton or the graphical representation of a relationship between two sets. A physics student needs a description of a principle; an electrical engineering student needs to be able to understand a circuit diagram, etc.. More complex concepts may be better understood by a blind person through tactile reproduction, but these require considerable effort [Supalo 2014]. Consequently, the available reproductions are limited. On the other hand, it is crucial that any digital book, app or web resource should be equipped with images - even complex ones - with descriptions useful for educational purposes.

Recently, AI has been introduced to support blind and visually impaired people in their everyday tasks [Walle et al., 2022]. More and more, apps or web services are appearing on the market to generate alternative image descriptions. Generative artificial intelligence can be a valuable support to tackle this type of accessibility task. Facebook [Wu 2017], and Microsoft [Yu 2022], for instance, are among the first players to have introduced image description generation functions in their services to overcome such an accessibility issue. People with visual impairments are increasingly making use of camera and AI-based tools to obtain a description of objects and scenes around them, or of images and photographs from social networks. Examples are 'Seeing AI' and 'Be My Eyes' [Kim 2023].

In this work, we investigate how popular AI-based tools can describe complex images in the field of STEM. The paper is organised in six sections: after an overview of the related work, the methodology, study and results are presented. A brief discussion and conclusions close the paper.

2. RELATED WORK

Images have the power to deliver important information, especially in the case of scientific and educational content. For this reason, alternative and equivalent content is crucial for people who cannot see. The alternative text should be accurate and descriptive while concise as possible, to not overload the user with useless information (Leotta et al 2023). Preparing alternative and narrative description is not an easy task and requires competence and accuracy, especially in the STEM field ([Lundgard 2021], [Mack 2021]). Williams et al. (2023) investigate how the accessibility of images is implemented by designers and developers in productive contexts since the point of view of accessibility practitioners might differ from those of researchers.

Thanks to recent progress, Artificial Intelligence has the potential to greatly enhance accessibility. In recent years, Image Captioning, the process of generating a textual description of an image has become an emerging research topic. Exploiting Natural Language Processing and Computer Vision, deep learning systems can generate captions [Sharma et al. 2020, Stefanini et al. 2022, Leotta et al, 2023]. For instance, OpenAI ChatGPT showed the potential to effectively support clinical decision making in the medical field, by exploiting a combination of language models for tuning the automatic generation of image captioning [Selivanov et al., 2023].

AI can significantly improve the daily life of blind people by enabling the understanding of images and visual contexts or even face recognition (Mott et al., 2023). This leads to more satisfaction in life, it is easy to use, quick to learn and effective [Kubulleket al. 2023]. However, to fully reach this target additional research steps are still needed, as suggested by recent studies (Leotta et al. (2022), Williams et al. (2022)).

Leotta et al. (2022) investigate Services such as Azure Computer Vision Engine, Amazon Rekognition, Cloudsight, and Auto Alt-Text for Google Chrome which process images and return textual descriptions, for understanding if they can be exploited for generating alt-text in the web content. Results showed that none of the analysed systems are mature enough to replace the human-based preparation of alternative texts although some tools

can generate good descriptions for specific categories of images.

Williams et al. (2022) investigated the state of alt text in HCI publications by analysing 300 figures (including data representations, and diagrams). Results revealed that the quality of alt text is highly variable, and nearly half of figure descriptions have few helpful information. More guidelines need to expand to address different content types of complex images, composed of multiple elements.

To the best of the author's knowledge there is no study comparing answers in automatic captioning generation of images in the STEM field, evaluating the perceived quality. In this study, we analyse and compare how popular AI-based systems describe STEM images belonging to different science domains, in response to 3 different levels of prompt.

3. THE STUDY

One of the aims of the project is to promote ICT solutions to enhance accessibility in order to overcome the barriers faced by people with visual impairments in accessing scientific studies and careers. One of the most important accessibility issues encountered by screen reading users relates to complex content conveyed through images, diagrams, graphs, and so on. Digital solutions aimed at supporting screen reading users in accessing STEM content are in the scope of our study.

In this work, we investigate how available AI-based tools, such as apps or online services, are suitable for describing complex images with STEM content. To this end, we selected a representative data set of STEM images, and exploited some AI Digital Assistants (ADAs), popular in the blind community, to generate alternative and textual descriptions of them.

3.1 Images Data Set

As data set, four STEM subjects were selected: biochemistry, physics, mathematics and information technology. For each subject, one image has been selected according to expected different levels of complexity (evaluated by authors), in terms of: image text and object recognition, deductibility of the context, and overall content analysis for detecting image semantic meaning.

1) Equation (Mathematics, no difficulty)
2) Krebs cycle (Biochemistry, basic difficulty).
3) Magnetic field (Physics, medium difficulty).
4) Diagram of a Finite State Machine (computer science, moderate difficulty).

The first picture, selected in the mathematics field, is a simple addition, shown in Figure 1. We can expect that an AI digital assistant recognizes it easily even with no information about the context being provided (i.e. no information about the math field). For example, a possible accurate description includes two information, i.e. the context and then the content. The image contains a math content. This content is the simple equation 1+1=2 in a b/w format.

$$1+1=2$$

Figure 1: A Simple math equation

The second image is a more complex picture in the chemistry field, the Krebs Cycle (Fig. 2). This picture incorporates labels that are very specific to the subject, therefore we can assume that the AI will recognize it but with some difficulty.
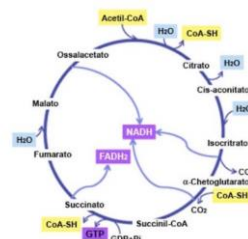

Figure 2: Krebs cycle (Biochemistry, basic difficulty). www.chimicaonline.it

We assume that the AI could give us more information about the concept and hopefully a detailed description of it. This picture is exploited for simplifying and assign an order to a series of chemical reactions.

The third picture is related to a Magnetic Field. It can be easily found in high school physics textbooks, and it is optimal to describe the shape and function of a magnetic field. It has some form of context due to the text in Italian (see figure 3).
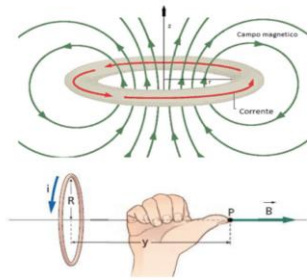
Figure 3: Magnetic Field (Physics, Image 3: magnetic field (physics, medium difficulty). www.chimica-online.it/

We expect the AI is able to recognize the language and translate it. This picture has been selected since using a hand to visually explain how a magnetic field works is a common method. It could be interesting to see how a visually impaired student could access this information through text and screen readers.

The fourth picture is the diagram of a Finite State Machine (see figure 4). It has no text that can lead to context and has little to no explanation: its simplicity is misleading and needs an accurate description for a blind student to understand its shape.
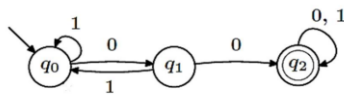

Figure 4: State Diagram (Computer Science, Moderate difficulty with no context).

3.2 AI Digital Assistants

AI Digital Assistants can assist blind users in exploring complex images. In our study we tested two apps designed for supporting visually impaired and blind people, and two popular ADAs (of two big players) which offer a good degree of accessibility for easy interaction via screen reader:

Seeing AI (https://www.seeingai.com) is a free app that describes the surrounding world by exploiting the power of generative AI. Seeing AI can read text, describing photos or images, and identify objects. Seeing AI provides a summary of what an image depicts. Tapping on the "more info" icon, the app generate a far more in-depth description. Moving the finger over the screen the app announces the locations of various objects.

Be My Eyes (https://www.bemyeyes.com) is a visual assistance app for people with low vision letting the user request video support any time via smartphone. Taking a picture Be My Eyes describes it. The goal is to make the world more accessible for blind or visually impaired people.

Microsoft Copilot (formerly Bing Chat Enterprise) is a public web service available on copilot.microsoft.com. It provides AI-powered chat for the web built on the latest large language models, GPT-4 and DALL-E 3. It's grounded in the Bing search index.

Google Gemini (https://gemini.google.com/app/) is the current iteration of Bard, a large language model from Google AI, trained on a massive dataset of text and code. It incorporates several advancements in terms of data, capabilities, performance, and availability.

Google Gemini Advanced (the paid version of Google Gemini) (https://one.google.com/explore-plan/gemini-advanced) is designed for highly complex tasks; it can understand, explain, and generate high-quality code in many programming languages.

3.3 Method

Two authors of this work performed tests with the data set images using the selected apps and AI digital assistants (see 3.2). We proceeded in two different ways to get the image descriptions.

(1) Each picture was displayed and shared with the 'Be My Eyes' app on an Android device (after clicking on the "take a picture" button). On the other hand, as this mode does not work with the 'Seeing AI' app, each image was first displayed on a computer screen Lenovo Yoga C640-13IML (13.3" 1920x1080) and then a photo was taken by the camera of an android Redmi 7 smartphone.

(2) When using 'Microsoft Copilot' and 'Google Gemini' a dialogue approach guided the user in the interaction with the AI digital assistant to obtain the description of the images.

Three types of prompts were used when interacting with each tool to ask for different types of description:

- Prompt 1: 'What is in this picture?'. This prompt is intended to identify what is in the image ('identification'), what context it belongs to and what it is about (in short). No context information is provided by the user.

- Prompt 2: 'Can you describe this picture?'. This second prompt is aimed at obtaining more details on the image, i.e. a 'image description'.

- Prompt 3: I'm a blind person. Can you describe this picture?'. The purpose is to obtain a more precise description which includes the visual representation, for a person who cannot see it. This can be a 'graphic description'.

For the two applications 'Seeing AI' and 'Be my eyes', the three prompts were not applied because the user cannot have a dialogue with the system itself. Therefore, we limited to analyse whether they were able to identify the subject and give a description.

For the ADAs, the three prompts were performed to obtain increasing details: identification (prompt1), description (prompt2) and a more accurate visual description suitable for blind users (prompt 3).

The evaluation is based on the quality of the ADAs output to the three prompts.

For the image/context recognition/identification (output of prompt 1) Yes/No is assigned.

For the image description (prompt 2) and visual description (prompt 3) outputs a score from 0 to 5 is assigned, according the following Scale: '0' No description; '1' Wrongly described; '2' Partially correct but insufficient/useless information; '3' Sufficient description but some content incomplete or inaccurate; '4' Good description; '5' Complete and precise description. This scale is subjective, based on the perceived quality assessed by this paper's authors. The consensus on the final rating was assessed by all authors (via videoconference discussions).

4. RESULTS

To make results more comparable, they have been arranged in tables, one per each AI digital assistant. In each table, the columns contain the evaluation of the three specifications of the three prompts. The last column in each table refers to Additional data, i.e. the information that can be added by the AI assistant, but which can result in too much verbosity for visually impaired people, interacting via screen reader.

For the two applications, the analysis was limited to only the two descriptions - identification and description.

Seeing AI. We observed that Seeing AI is not precise enough to describe a specific didactic illustration. Table 1 summarizes the results.

Table 1 - Test with Seeing AI

Figure	Identification	Image description
Equation	No	2
Krebs cycle	No	0
Magnetic field	No	1
State Diagram	No	1

Be My Eyes. This tool was really useful to get a first idea of the picture. It does not give us additional data, but it describes the picture with superficial accuracy

while making some misinterpretations. Table 2 summarizes the results.

Table 2 - Test with Be My Eyes

Figure	Identification	Image description
Equation	Yes	4
Krebs cycle	Yes	3
Magnetic field	Yes	5
State Diagram	Yes	3

Bing Copilot. This tool recognizes all pictures and gives a full detailed description giving useful additional information. However, it makes some misinterpretations, especially while describing the State Diagram. Table 3 summarizes the results.

Table 3 - Test with Bing Copilot

Figure	Identification	Image description	Visual description	Additional data
Equation	Yes	5	5	5
Krebs cycle	Yes	5	5	3 options for more info
Magnetic field	Yes	5	4	5
State Diagram	Yes	4	4	5

Google Gemini. It recognizes the first three pictures but fails to recognize the State Diagram. For the Krebs cycle and the magnetic field, it delivers a full detailed description. Concerning the equation, it delivers the right information, but with an inappropriate graphic description of the background. The image and didactic descriptions are the same. Table 4 summarizes the results.

Table 4 - Test with Gemini

Figure	Identification	Image description	Visual description	Additional data
Equation	Yes	4	4	Too much
Krebs cycle	Yes	5	5	Yes
Magnetic field	Yes	5	5	Yes
State Diagram	No	2	1	Yes

Google Gemini Advanced. Analogously to the free version, Gemini Advanced recognizes the first three pictures but fails to recognize the State

Diagram. For the Krebs cycle and the magnetic field, it delivers a full detailed description. Concerning the equation, it delivers the right information, but with an inappropriate graphic description of the background. The image and didactic descriptions are the same. Table 5 summarizes the results.

Table 5 - Test with Gemini Advanced

Figure	Identifi cation	Image description	Visual description	Additional data
Equation	Yes	4	4	Too much
Krebs cycle	Yes	5	4	Yes
Magnetic field	Yes	5	5	Yes
State Diagram	No	2	1	Yes

5. DISCUSSION

When considering images for educational purposes, it is mostly important to have descriptions that do more than just give a summary of what the image contains. If in the textbook or handout the image represents a directed graph, for instance, it is not enough that the AI digital assistant says that it is a (directed) graph. In addition, a visually impaired student needs to learn how the graph is made up, that is there are nodes (i.e. points) shown in random order, maybe with a label, there are arcs represented by arrows connecting two nodes, etc. This is a very simple example. In case of a graph contains a loop or a not connected node, the graph description should not tell the user "the graph contains a loop and has an isolated node", because by doing so, it does not represent the graphic description to the user: i.e. what a loop means and how it is represented. This type of information can be effective when the student has already learned these concepts (even in a graphical sense). So, in descriptions like this one, the educational purpose might be missing. On the other hand, when the user has already understood what a graph looks like, then the digital assistant should provide a brief description such as "the graph consists of 5 nodes, and the following arcs...". Sometimes, however, it may only be useful to know that the image represents a directed graph, and nothing more. It depends on the context and the purpose for which the image is being analysed. This is why we proposed three types of approaches to understand whether using differentiated and more precise prompts would result in more adequate image descriptions for the intended purpose.

The study conducted with the 4 selected images showed that with the first prompt "What is this picture?" the digital assistant should provide the context. With the second prompt "Could you describe this picture?", the description provided should contain in short what it is about. Finally, with the third prompt "I'm a blind user. Could you describe to me this picture?", the description should be much more precise for a blind person, including information on the graphic representation so that a person who cannot see the representation can understand what it looks like. In addition to this kind of detail, the context (e.g. educational) and purpose (e.g. learning, exercise or examination) could be better defined. This, however, would need to be better detailed.

After testing two applications for mobile devices and three AI assistants we observed the following pattern: the mobile applications are useful for a quick scan and recognition of the picture, but they rarely provide any didactic information and sometimes even fail to recognise the context of the picture. On the other hand, AI assistants such as Google Gemini (in the simple and advanced versions), Bing Copilot, which delivered the best results, is very useful in providing a full description of the picture as well as explaining the context and information related to it. We also noticed that by repeating the same prompt the answer is refining over time.

To sum up, some aspects emerged from our analysis:
- The behaviour of the tool in interpreting the target images may change over time, so the answer risks not being unique.
- The answer depends on the training set. It seems that the Gemini tool is trained with images with background, maybe more suitable for photos and pictures. So, in the first answers the tool describing it would analyse the context and it is not properly focused on the specific content.
- Results may be affected by the tool usage in terms of how images are provided to the tool. 'Be My Eyes' requires taking a photo by the smartphone's camera; as a result, the image resolution can vary according to the different models or how the image is captured; this may provide a negative effect. Moreover, capturing a picture through a photo might be very complex for blind users.
- Last, the more accurate the answer the more time is needed for the AI assistant to describe the picture, as it happens with Bing Copilot.

The results obtained are dependent on the prompts used. The results may change or be completely different depending on the questions posed to the tool. This is certainly a limitation of this work, which however intends to be a preliminary study on this type of research.

6 CONCLUSIONS

The use of STEM materials should be guaranteed for everyone to have equal access to scientific studies and careers. Unfortunately, this is not the case for everyone, as many accessibility problems in the use of STEM content continue to exist. For example, screen reading users have difficulty using graphical content. Alternative descriptions for images, photos, graphs and diagrams are a possible solution to address this issue in an accessible manner. Unfortunately, they have to be prepared manually, which is time-consuming and requires the necessary skills.

Recently, artificial intelligence has revolutionised our lives by influencing many fields, including accessibility. Several AI-based tools and applications are already on the market as Digital Assistants to support accessibility. This study is focused on investigating whether generative AI used by digital assistants is suitable for generating descriptions of STEM graphic content. The literature shows that generative AI is increasingly being used to produce descriptions of commonly used images, especially photos and art images. Scientific content is little considered, with the risk of excluding many people from access to STEM studies and careers. Generative AI could be well exploited to support the task of producing complex image descriptions also for STEM content. In this study we analysed alternative descriptions automatically generated for STEM graphical content (with different levels of difficulty) by five AI Digital Assistants and applications. Based on the tests conducted on even a few images, we can say that 'Seeing AI' overall is unsuitable because it cannot identify STEM content. 'Be my eyes', identifies objects correctly, and produces good descriptions of simple content, and less accurate descriptions of more complex content. Gemini has limitations in more complex images such as the state diagram and generates descriptions that are too verbose. Bing Copilot also seems to perform well with more complex images, both in identification and descriptions, including the visual one.

As we discussed, the tested AI assistants can be useful to visually impaired students, although we saw that some very promising AI assistants are not always reliable, especially for complicated images like STEM subjects. Moreover, while using the tools we could find that some parts of the descriptions were not appropriate. Those little mistakes can be challenging for a student who is trying to learn or while taking an exam. It needs work in terms of accuracy and accessibility but in the end, current AI assistants need some improvements for effectively assisting blind students. Last, the AI assistants should be able to adapt the various descriptions to the level of the student, i.e. whether he/she is new to the subject, or has already learned several concepts.

The study is certainly too limited to be able to say whether the tools are mature or not for interpreting STEM content. However, it emerges that some tools are beginning to provide appropriate descriptions, albeit with many limitations and inaccuracies. A more in-depth study may provide more guidance. We can conclude that when image descriptions related to STEM content are to be generated to the user, they should be provided according to the student's learning level with respect to a certain subject. Furthermore, the student may be in different contexts: learning, review/practice, examination. The system should also consider these three different contexts to produce appropriate descriptions.

REFERENCES

Diagram Center (2012). Making Images Accessible, online Available: http://diagramcenter.org/making-images-accessible.html/.

Gamage, B., Do, T. T., Price, N. S. C., Lowery, A., & Marriott, K. (2023). What do Blind and Low-Vision People Really Want from Assistive Smart Devices? Comparison of the Literature with a Focus Study. In Proc. of the 25th Int. ACM SIGACCESS Conference (pp. 1-21).

Gleason, C., Carrington, P., Cassidy, C., Morris, M. R., Kitani, K. M., & Bigham, J. P. (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 (pp. 549-559).

Hilal, A. M., Alrowais, F., Al-Wesabi, F. N., & Marzouk, R. (2023). Red Deer Optimization with Artificial Intelligence Enabled Image Captioning System for Visually Impaired People. Computer Systems Science & Engineering, 46(2).

Kim, H. N. (2023). Digital privacy of smartphone camera-based assistive technology for users with visual disabilities. International Journal of Human Factors and Ergonomics, 10(1), 66-84.

Kubullek, A. K., & Dogangün, A. (2023). Creating Accessibility 2.0 with Artificial Intelligence. In Proceedings of Mensch und Computer 2023 (pp. 437-441).

Leotta, M., Mori, F., & Ribaudo, M. (2023). Evaluating the effectiveness of automatic image captioning for web accessibility. Universal access in the information society, 22(4), 1293-1313.

Lundgard, A., & Satyanarayan, A. (2021). Accessible visualization via natural language descriptions: A four-level model of semantic content. IEEE transactions on visualization and computer graphics, 28(1), 1073-1083.

Mack, K., Cutrell, E., Lee, B., & Morris, M. R. (2021). Designing tools for high-quality alt text authoring. In Proc. of the 23rd Int. ACM SIGACCESS (pp. 1-14).

Mott, M. E., Tang, J., & Cutrell, E. (2023). Accessibility of Profile Pictures: Alt Text and Beyond to Express Identity Online. In Proceedings of the CHI Conference on Human Factors in Computing Systems (pp. 1-13).

Selivanov, A., Rogov, O. Y., Chesakov, D., Shelmanov, A., Fedulova, I., & Dylov, D. V. (2023). Medical image captioning via generative pretrained transformers. Scientific Reports, 13(1), 4171.

Sharif, A., Chintalapati, S. S., Wobbrock, J. O., & Reinecke, K. (2021). Understanding screen-reader users' experiences with online data visualizations. In Proc. of the 23rd Int. ACM SIGACCESS Conference (pp. 1-16).

Sharma, H., Agrahari, M., Singh, S. K., Firoj, M., & Mishra, R. K. (2020). Image captioning: a comprehensive survey. In 2020 International Conference PARC (pp. 325-328). IEEE.

Splendiani, B., & Ribera, M. (2014). Accessible images in computer science journals. Procedia computer science, 27, 9-18.

Stangl, A., Morris, M. R., & Gurari, D. (2020). "Person, Shoes, Tree. Is the Person Naked?" What People with Vision Impairments Want in Image Descriptions. In Proc. of the 2020 CHI conference on human factors in computing systems (pp. 1-13).

Stangl, A., Verma, N., Fleischmann, K. R., Morris, M. R., & Gurari, D. (2021). Going beyond one-size-fits-all image descriptions to satisfy the information wants of people who are blind or have low vision. In Proc. of the 23rd International ACM SIGACCESS (pp. 1-15).

Stefanini, M., Cornia, M., Baraldi, L., Cascianelli, S., Fiameni, G., & Cucchiara, R. (2022). From show to tell: A survey on deep learning-based image captioning. IEEE transactions on pattern analysis and machine intelligence, 45(1), 539-559.

Supalo, C. A., & Kennedy, S. H. (2014). Using commercially available techniques to make organic chemistry representations tactile and more accessible to students with blindness or low vision. Journal of Chemical Education, 91(10), 1745-1747.

WAI Web Accessibility Tutorials: Complex Images, Updated 17 January 2022, online Available at https://www.w3.org/WAI/tutorials/images/complex/.

Walle, H., De Runz, C., Serres, B., & Venturini, G. (2022). A survey on recent advances in AI and vision-based methods for helping and guiding visually impaired people. Applied Sciences, 12(5), 2308.

Williams, C., de Greef, L., Harris III, E., Findlater, L., Pavel, A., & Bennett, C. (2022). Toward supporting quality alt text in computing publications. In Proc. of the 19th Int. Web for All Conference (pp. 1-12).

Wu, S., Wieland, J., Farivar, O., & Schiller, J. (2017). Automatic alt-text: Computer-generated image descriptions for blind users on a social network service. In proc. of the 2017 ACM conference on computer supported cooperative work and social computing (pp. 1180-1192).