AI and Media Accessibility

An Overview

Authors:

Letícia Seixas Pereira (LASIGE, Faculdade de Ciências da Universidade de Lisboa)

Carlos Duarte (LASIGE, Faculdade de Ciências da Universidade de Lisboa)

Disclaimer

This document contains description of work conducted in the scope of the COST Action LEAD-ME and the H2020 WAI-CooP project.

The authors of this document have taken any available measure in order for its content to be accurate, consistent and lawful. Neither the project consortium as a whole nor the individual partners that implicitly or explicitly participated in the creation and publication of this document hold any responsibility for actions that might occur as a result of using its content.

Table of Contents

1 Executive Summary

This report presents the outcomes of a systematic review of the literature to investigate the use of artificial intelligence in digital accessibility studies. We queried 5 scientific research databases for papers related to digital accessibility and AI published in the last 5 years and added papers that had been cited multiple times by the papers retrieved by the original query. This resulted in 280 papers. After several rounds of screening, we considered a total of 71 papers for the review.

Our findings suggest that most studies on this context focus on machine learning, natural language processing, computer vision, and deep learning, with particular emphasis on media accessibility. Blind and low vision users were the focus of most studies. In what concerns ways that AI can leverage digital accessibility, researchers identified key areas for further investigation, including the use of real data to automate processes currently reliant on human judgment and exploring different AI methods, as well as optimizing and reducing the costs of training machine learning models so to improve their performance and scalability. AI systems were envisioned to be able to assist users in a multitude of ways: providing multi-model representations of content, improving semantic services used for implementation and assessment, exploring personalization alternatives, categorizing web components, or providing machine translation of content. On the other hand, the survey also identified how AI can hinder digital accessibility. Two major groups of issues emerged from the analysis. The lack of accuracy and reliability of current AI-based models is raised my numerous studies. The second group reflects potential ethical issues with AI-based systems, including social biases, privacy, and legal responsibility.

2 Introduction

This research aims to analyse the impacts of current AI research in digital accessibility. For that, we aim to investigate (1) how current AI research addresses digital accessibility, (2) how current AI research can leverage digital accessibility, and (3) how current AI research can hinder digital accessibility.

AI is a complex subject, thus can be divided into several subdomains. To better investigate current solutions, challenges, and opportunities regarding digital accessibility, this analysis focuses on AI and its branches. Considering the different categorizations currently used to represent this complex theme, for this research, we followed the ACM Computing Classification System[1] (ACM CSS), as described below:

- Artificial Intelligence
  - Natural language processing
  - Knowledge representation and reasoning
  - Planning and scheduling
  - Search methodologies
  - Control methods
  - Philosophical/theoretical foundations of artificial intelligence
  - Distributed artificial intelligence
  - Computer vision
- Machine Learning
- Deep Learning

In this report, we first present a description of the methodology used, followed by the results obtained, and a brief discussion of these findings. Finally, we draw some conclusions on how these findings can help us to identify current challenges and opportunities raised by the increasing use of AI regarding digital accessibility.

[1] https://dl.acm.org/ccs

3 Methods and Analysis

This review followed the guidelines of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses 2020 (PRISMA)[2].

3.1 Eligibility criteria

This research considered:

- studies published in the English language;
- peer-reviewed documents, including studies published in journals, conferences, symposia, or workshops;
- studies published during the preceding 5 years (2017 to 2022);
- digital accessibility studies employing or discussing artificial intelligence techniques.

Thus, the research did not consider:

- studies not published in the English language;
- documents not peer-reviewed, such as books, Ph.D. and Master thesis, technical reports, white papers, posters, newsletters, editorials, and prefaces;
- studies published before 2017;
- digital accessibility studies not employing or discussing artificial intelligence techniques.

3.2 Information sources

The database selection in this research focused on providing a comprehensive sample of general, technological, and medical sciences sources. Databases considered in this review comprised ACM Digital Library, IEEE Xplore, Web of Science, ScienceDirect, and Scopus. Filter features provided by these platforms were used to support the paper search, such as features for filter results by title, keywords, and abstract fields, as well as filter results by peer-reviewed sources.

3.3 Search strategy

3.3.1 Electronic databases

Considering the scope established for this research, the complete list of strings inserted into each database was based on the concepts available in the ACM CSS and is provided in Annex 1. The databases previously enumerated were searched from 2017 and no end date was indicated. Information on each database search, such as filters and date of research, is provided in Annex 2.

[2] http://www.prisma-statement.org/PRISMAStatement/

3.3.2 Manual searches

The reference and citation lists of the studies included in this research was also searched to identify additional potentially relevant studies, according to the scope of this review. For these studies two criteria of inclusion were modified, as further detailed.

3.4 Selection process

3.4.1 Preliminary screening

One researcher first removed the duplicates and screened titles and abstracts of each paper according to the inclusion and exclusion criteria previously described. In cases of uncertainty, the same assessment was conducted by a second researcher, and discussed among the two to reach consensus. Priority was given to including studies for further analysis rather than excluding them.

From 376 records retrieved, 180 were single studies that were further screened. From these, 111 did not meet the inclusion criteria - with 105 of them out of the research scope. From that, 70 potential studies remained for full text screening. In brief, at the end of this step, studies that were duplicated, not relevant for the scope of the research, or did not meet the eligibility criteria were excluded. Information on the number of studies is summarized in Table 1.

Table 1: Number of studies identified on electronic databases in the preliminary screening phase.

Records retrieved	376
Single studies retrieved	180
Studies outside of the inclusion criteria	111
Studies included in the review	70

Another step included on the first screening phase was to gather the list of references of each paper in this first list, as well as other papers citing each paper included. This task was conducted with the support of the ResearchRabbit[3] literature mapping tool, Google Scholar[4], and the digital libraries previously mentioned. The same process conducted before was repeated for this new dataset of papers. However, for this step, two eligibility criteria were modified:

- We chose to not exclude papers based on publication period, as we did not want to exclude any paper with major influence on these topics,
- Considering the multidisciplinarity of this research, only papers that (1) were cited by at least two papers from the first list or (2) cite at least two papers from the first list were included.

[3] https://www.researchrabbit.ai

[4] https://scholar.google.com

At the end of this phase, 100 potential studies remain for screening on full text. Information on the number of studies is summarized in Table 2.

Table 2: Number of studies identified on manual searches on the preliminary screening phase.

	References	Citations
Records retrieved	919	122
Studies considered for review	72	28

3.4.2 Final screening

After the initial screening, 170 potential studies remain for full text screening. This final screening aims to confirm that the studies selected follow the inclusion and exclusion criteria as well as its relevance for this research, i.e., the research scope of digital accessibility and Artificial Intelligence. All the information collected during this process concerning the final screening of the electronic database dataset is summarized in Table 3, and Table 4 presents the information concerning the final screening of the manual searches dataset. This information is also reported according to the PRISMA Flow Diagram in Annex 3.

Table 3: Number of studies identified on electronic databases on the final screening phase.

Initial dataset	70
Studies outside of the inclusion criteria	1
Studies outside of the research scope	27
Full text not retrieved (full text not found)	1
Studies included	41

Table 4: Number of studies identified manual searches on the final screening phase.

	References	Citations
Initial dataset	72	28
Studies outside of the inclusion criteria	-	2
Studies outside of the research scope	53	9
Studies already included on the first dataset	3	2
Full text not retrieved (full text not found)	-	1
Studies included	16	14

At the end of the final screening, 71 studies were included in this review, and proceeded to data extraction. The full list of papers is provided in Annex 4.

3.5 Data extraction

From each study included in this review, a set of pre-established information was collected by at least one researcher. This information was gathered in a spreadsheet to be further analysed, covering the following items:

- Study details: reference, country/countries, year
- Study context: AI subdomain, digital accessibility subdomain
- Study design: objective, type of research (i.e., qualitative, quantitative, mixed, literature review, survey)
- Participants demographics (if any): population size and demographics, inclusion and exclusion criteria, settings, assistive technologies (if any)
- Scientific contribution: outcomes analysed
- Conclusions: authors and researcher conclusions
- Summary of possible impacts on digital accessibility

3.6 Data analysis

Studies were grouped by AI subdomain, followed by digital accessibility subdomain. With that information, we expect to respond to our pre-established research questions, namely: how current AI research addresses digital accessibility, how current AI research can leverage digital accessibility, and how current AI research can hinder digital accessibility.

4 Findings

4.1 Study characteristics

Most of the studies analysed conducted both quantitative and qualitative data analysis, followed by those conducting only a quantitative analysis. This proportion is to be expected given the AI component in the studies reviewed. In a smaller number, 10 studies conducting only a qualitative data analysis were also retrieved. Finally, 15 studies did not provide any data analysis as they just provided a proposal of an approach or a literature review. These numbers are presented in Table 5.

Table 5: Data analysis conducted on the studies analysed.

Mixed (Qualitative and Quantitative)	26
Quantitative	20
Qualitative	10
Other	15

4.2 Demographic information

4.2.1 Country

From the 71 studies reviewed, most of them were from authors based on United States with 31 papers, followed by China, with 11 papers. Studies conducted from European Union member states represented 21% of the analysed dataset with 18 studies. The distribution of this total by member state is presented in Table 6.

Table 6: Number of studies included by member state.

Italy	5
France	4
Spain	4
Belgium	1
Germany	1
Greece	1
Ireland	1
Portugal	1

4.2.2 User groups

As previously mentioned, AI-oriented studies are expected to rely on a stronger quantitative analysis. However, the digital accessibility aspect also requires users to be involved wherever possible. In this context, we seek to better understand which audience these works are targeting. Of the 71 studies included in this survey, 39 included subjects at some stage of their analysis. We identified 27 studies working closely with users with disabilities in different contexts [6–9,12,13,16–21,25,26,28,30,31,33,34,40–42,44–46,53,54]. Additionally, a group of 6 studies investigated their proposal with digital accessibility experts, such as AAC experts, and people with expertise in web design and development [13,14,29,39–41]. Two groups of studies investigated their approach with accessibility practitioners and content authors. The first one counted with sign language experts, teachers, professional translators, and experts in linguistic [3,13,16,43]. The second investigated content authoring in different context and worked closely with people without any disability in different roles such as presentation authors, caption authors, video editors, and audio description professionals [33,35,36]. Finally, two works also recruited external workers, such as through Amazon Mechanical Turk[5] on their analysis [10,32]. Finally, 9 studies included other users, without further specification or criteria [2,12,15,25,28,32,40,42,45]. This is common when conducting open surveys or gathering a group control data. Table 7 details this information.

Table 7: Summary of users included on the studies analysed.

Users with disabilities	27
Digital accessibility experts	6
Accessibility practitioners	4
Content authors	3
External workers	2
Other users	9

Concerning the type of disabilities included on these studies, as also expected, most of them included blind and low vision (BLV) users. From the remaining, deaf and hard-of-hearing (DHH) users, users with motor impairments, users with speech disabilities, Augmentative and alternative communication (AAC) technology users, and users with intellectual or development disabilities (IDD) were involved. It is also important to highlight that some studies included different users' profiles on their analysis. This information is presented in Table 8.

Table 8: Summary of user data on the studies analysed.

BLV users	24

[5] https://www.mturk.com

DHH users	4
Users with motor impairments	2
Users with speech disabilities	2
AAC users	1
IDD users	1

Table 9 presents the mean number of participants per user group included in quantitative, qualitative, and mixed studies. It is important to highlight that some studies did not provide detailed information or distinction of the number between the analysed groups. For instance, Song et al. [40] evaluated the performance of their proposed according to the data provided by 49 people with motor disability, hearing disability, visual impairment or speech disability, not specifying how this group is distributed. In these cases, the total number of users was used to provide the information below.

Table 9: Mean number of participants per quantitative, qualitative, and mixed studies.

	Quantitative	Qualitative	Mixed
BLV users	3	8	512
DHH users	3	2	-
Users with motor impairments	3	5	-
Users with speech disabilities	3	5	-
AAC users	2	-	-
IDD users	-	8	-

4.3 Themes

From each paper analysed, extracted data from two domains of the study context will guide this analysis: AI subdomain, and digital accessibility subdomain. The AI-related subdomains were categorized according to the previously mentioned ACM concepts. To obtain more detail on the topics covered, we opted to use all terms in the same hierarchy - except for AI itself. Therefore, the papers were categorized according to the following list:

- Natural Language Processing
- Knowledge representation and reasoning
- Planning and scheduling
- Search methodologies
- Control methods
- Philosophical/theoretical foundations of artificial intelligence
- Distributed artificial intelligence

- Computer vision
- Machine learning
- Deep learning

Initially, all the information regarding the AI subdomain was gathered and it was possible to analyse that most of the publications investigate machine learning techniques, followed by NLP techniques. These numbers are summarized in Table 10.

Table 10: Total of papers identified by the different AI subdomains analysed.

Machine learning	40
Natural Language Processing	32
Computer vision	17
Deep learning	14
Knowledge representation and reasoning	2
Distributed artificial intelligence	1
Philosophical/theoretical foundations of artificial intelligence	1

The next step of this analysis focused on the most referred subdomains, i.e., Machine learning, Natural Language Processing, Computer vision, and Deep Learning. For each one of them, the digital accessibility application domains were identified, giving us the following ranking: Machine learning for web accessibility evaluation, NLP for media accessibility, and Computer vision for media accessibility. These numbers are presented in Table 11. All the other combinations had less than 10 papers addressing its topic and the complete list is provided in Annex 5.

Table 11: Number of papers in the Top 3 AI domains and digital accessibility application domains.

Machine learning for web accessibility evaluation	16
NLP for media accessibility	11
Computer vision for media accessibility	10

With the studies identified for each theme, in the following sections we will present the current scenario, plus the challenges and opportunities identified by researchers in their works.

4.3.1 Machine learning for web accessibility evaluation

Web accessibility evaluation aims to find barriers for people with disabilities in accessing and interacting with web content. Automatic evaluations are very useful to identify several accessibility issues, in a fast and scalable way, but there are still limitations, such as not being able to cover all the accessibility checkpoints. In this scenario, web accessibility still relies on manual assessments, that, on the other hand, incur expensive cost in evaluating large websites. This is just a brief overview of current challenges. Web accessibility evaluation has a lot of potential to expand, and AI-based solutions have a lot to contribute to fill some of these gaps. As previously mentioned, manual assessment are very costly, so most current evaluation methodologies rely on a representative sample of the website to presume its accessibility. However, there are a few drawbacks to this, such as possible biases on the web pages chosen, or leaving out important web pages.

Previous research indicate that several factors can influence the results obtained from experts' assessments. Li et al. [27] explore the challenge of assigning experts to conduct accessibility assessments by using evaluators' historical evaluation records and experts' review to train a minimum cost model via machine learning methods. This is then used to obtain an optimal task assignment map. Li et al. [27] discuss future work on this topic as including using other measurements for assignment as well as employing different machine learning methods. They also pointed out further investments in supporting the decision of the proper number of evaluators that a task requires.

Wu et al. [47], Bajammal et al. [10] and Zhang et al. [52] propose a predictive approach to provide accessibility results. Wu et al. [47] use a sample of web pages previously evaluated by experts to train their semi-supervised machine learning models. This model is then used to obtain the accessibility evaluation results for all pages in a web site. Bajammal et al. [10] proposed an approach that automates web accessibility testing from a semantic perspective. It analyses web pages using a combination of visual analysis, supervised machine learning, and natural language processing, and infers the semantic groupings present in the page and their semantic roles. It then asserts whether the page's markup matches the inferred semantics. Zhang et al. [52] detail their approach to gather this initial dataset with a first active learning method to select the most informative web pages on the website. This dataset is then evaluated by experts and used as input for the prediction model. Zhang et al. [52] point out that further efforts could be in evaluating the results obtained by using different machine learning methods to the prediction model so to obtain more precise results. Bajammal et al. [10] highlights one limitation of their work being that it only covers a subset of accessibility requirements, and, for that, it would require a novel technique to address other accessibility requirements.

Yu et al. [49] also proposes an active learning method, but focused on identifying the number of samples needed for a specific website, as well as which pages to use as samples. Harper et al. [22] and Zhang et al. [51] invest in further investigating the representativeness of web page sampling, in particular, for methods that aim to automatize this process. Both works explore clustering techniques to gather web pages with similar structures to be used in a manual assessment. Harper et al. [22] develop a solution that crawls a website comparing and clustering web pages with similar structures, based on DOM-block similarity, while Zhang et al. [51] exploit similarities in URL patterns. Zhang et al. [50] propose a metric-specific sampling method that uses a greedy algorithm to approximately solve the optimization problem for Web Accessibility Quantitative Metric (WAQM) in an efficient way.

Harper et al. [22] provide a few thoughts on future directions for their work and for the sampling context. One of them is including a combinatorial approach to improve web accessibility evaluation effectiveness. Future work would be to deploy this approach and test it against standard measures to evaluate the validity of the outcomes. Harper et al. [22] also mention how web pages that cannot directly be crawled, such as those behind a password or pay wall will not be accessible by the tool. Interactive functionalities could be integrated to mitigate this problem. Finally, Harper et al. [22] and Zhang et al. [51] stress out the challenge of handling websites with complicated structures or that do not provide a strong common structure. Another challenge highlighted by both works was handling single page applications. For Harper et al. [22] the difficulty lays on hidden and visible parts of the DOM, depending on external factors. Zhang et al. [51] suggest sampling from the incremental content instead of the whole website to improve the efficiency of the accessibility assessment.

Metrics are also a theme discussed by researchers on the scope of web accessibility evaluation. For instance, Song et al. [41] propose a metric that can better match the accessibility evaluation results with the user experience of people with disabilities by aligning the evaluation metric with the partial user experience order (PUEXO). To achieve this, a machine learning model is developed to derive the optimal checkpoint weights from the PUEXO. In a follow up work, Song et al. [40] evolve the previous metric practical applicability by developing a reliability aware model which considers the heterogeneous reliability barriers and design an Expectation Maximization (EM) based algorithm to build this model. Their results showed that considering reliability in user experience outperforms state of the art approaches.

Another challenging topic in the web accessibility evaluation context, that was also mentioned in previous discussions, is handling dynamic content. Antonelli et al. [5] discuss current research on this field as well as limitations of current automatic tools that assess the accessibility of dynamic web pages. Through a meta-review of some multidisciplinary techniques and applications, they show that there is a lack of structural standardization for the creation of interactive elements. The use of AI, ML, and statistical inference techniques can contribute to the exploration and analysis of the structural and behavioral patterns of a web application. Concerning machine learning techniques, when applied with syntactic and semantic analysis of dynamic web pages source code, it can contribute to the identification and classification of web components, for instance. A proper recognition of web components can be useful to check compliance with accessibility standards. One approach to this was conducted in a follow up study. In Antonelli et al. [4], the authors propose an approach focused on the identification of drop-down menu widgets. This approach also works toward evolving automatic evaluation strategies and adaptation of web applications. The task of identifying the widget was modeled as a Classification Problem to determine whether every change in the DOM structure of a web application is a widget element or not. In a similar note, Rizo et al. [37] propose a machine learning pipeline for automatic classification of web components. In addition to handling different web components, this approach also targets identifying the items that compose them. In this way, it is possible to obtain not only a dataset for widgets but also for their subcomponents. One limitation of this work is that results cannot be generalized since datasets were only composed by dropdown menu widgets. On both works, researchers reinforced the need to direct further efforts to other types of web widgets. On Rizo et al. [37], if the same results are obtained for different elements. This investigation topic can also be expanded towards ARIA conformance.

While topics such as sampling, metrics, and dynamic content were mentioned, Duarte et al. [15] explores semantic content analysis to support web accessibility evaluation, in this case, media accessibility. They propose an algorithm to automatically rate the similarity between a media content and its textual description in a web accessibility evaluation context. This algorithm relates the content descriptors and description through a direct similarity metric, combined with an indirect metric. With these results, Duarte et al. [15] also argue that, in an automated accessibility evaluation context, it is important to clearly define evaluation procedures and disambiguate all concepts required for evaluation. Further work considers focus on improving the recall and specificity of the algorithm classifications.

Yu and Bu [48] provides an overview of current AI-based systems towards improving access for visually impaired people in China. While this is a broader scope, conclusions can be easily applied to this topic. This overview mentions some of the current challenges in digital accessibility as the lack of accessibility awareness amongst developers, inadequate understanding of the real needs of users, and the inability to simulate real user behaviors. Yu and Bu [48] reported that more efforts should be put into making technology barrier-free.

Abou-Zahra et al. [1] restate the value of leveraging manual assessment results as input for machine learning, as seen in some current studies. Also, deploying AI-based systems for detecting accessibility barriers can be used to support accessible content authoring as well as for code augmentation. While discussing how AI-based systems can enhance digital accessibility, Abou-Zahra et al. [1] gives an important highlight for conformance evaluation services. Firstly because, as seen in this analysis, the data is already available. Secondly because there is already a business case for it.

4.3.2 NLP for media accessibility

Natural language processing is a subdomain of AI that benefits from linguistics and computational power to create systems able to “understand” natural language data. This powerful tool was explored on the studies investigated to support machine-generation alternative media descriptions. Researchers explored different types of media, such as images, videos, and audio, and combined different NLP techniques such as semantic analysis, natural language generation, and speech recognition.

For instance, Singh [38] obtained a large dataset of images and their alternate texts and tags and then trained a model based on the image and associated tags. This model aimed to improve machine-generated captions to provide the most fit for a specific image. In addition, these captions are first generated in English, but could also be provided in user-specific language with a machine translation API. Singh et al. [38] points out the costs of training a Machine learning model on high performant GPUs, especially since one of their challenges is to improve accuracy of descriptions generated by doing more training and fine tuning their model.

Wang et al. [44] and Sreedhar and Tan, et al. [42] investigated how contextual information available on a web page can help improve product image descriptions for e-commerce platforms. Wang et al. [44] implemented a mechanism that supports question-answering interactions on a reconstructed product page. To achieve this, they formulated syntactic rules to extract review snippets, which were used to generate image descriptions and responses to users’ queries related to product appearances. NLP techniques are used to handle responses to both keywords queries as well as natural language questions and extracting concepts from the reviews provided for the product. Sreedhar et al. [42] employed a series of models, trained for each topic extracted from the review comments, to parse semantically relevant keywords from the unstructured review comment and then associate keywords to these predefined topics. They also combined scene description and the output of the parsed review comment and generate human-readable alt text. Sreedhar et al. [42] also identified the need to further investigate how blind users experience customized preferences to personalize their preferred balance between product-specific and scene-specific descriptors. Wang et al. [44] reported there still exists a gap between user expectations and the answer quality in other details beyond the attributes included. On a similar note, they envisioned investment on human helpers through crowd workers, and formulating guidelines to support product image description.

Huh et al. [25] also explored contextual information to augment image description, in this case, webtoon, a type of digital comics read online. The mechanism developed allows users to have active control over the reading pace and the level of detail by requesting additional details or relevant comments per webtoon panel on demand. Their computational pipeline uses NLP techniques to score and abridge descriptions, and to extract descriptive comments. Huh et al. [25] suggests further investigation on complementary descriptions with audio augmentations, as well as in objectivity and output consistency for the descriptions created. Finally, a concern is how to handle incorrect output, as no user reported that they noticed irrelevant comments in their user study.

Audio augmentation for improving media accessibility was also explored by Pavel et al. [33] and Zhang et al. [53], in this case, video and emoji accessibility. Pavel et al. [33] worked on a mechanism that allows audio description describers to efficiently produce audio descriptions that maintain the length and audio quality of the existing video. This study is strongly based on speech segmentation techniques to identify description locations by classifying audio regions. Following that, they generate candidate descriptions, based on a parse tree and ranking the simplified candidates. Finally, they score audio description compositions so they can be optimized and rendered. Their results indicate that their solution allows users to automatically retarget extended descriptions to inline audio descriptions improving the ease and efficiency of creating such descriptions without overlapping content. However, they pointed out that synthesizing speech leads to low-quality results when few training examples exist, providing results that are too general or inaccurate to be useful. Authors point out that a possible approach to handle this issue is to explore a model fed by descriptions previously provided by describers themselves. In addition, they also suggest further investigation on audience preferences covering new types of media and content. The combination of audio augmentation and new types of media is present in the work of Zhang et al. [53] that explores the design of a speech-based emoji entry system. For that, speech-to-text techniques are used to explicitly search for emojis as well as to allow flexible search queries with natural language understanding. This work also comprises searches on different languages, English and Chinese. Authors defined further investigation on evaluation network latency on real-world settings, integrating this tool into an actual keyboard, and extend capabilities to other forms of visual media, such as stickers and memes.

Although the mentioned works address media accessibility in different contexts, another common denominator between them is the constant difficulty in providing machine-generated descriptions with enough quality to support users in interpreting media content. He et al. [23] provides an in-depth look at the current state of AI-based technologies to enhance media accessibility. They reiterate the challenge of creating supervised training models for image captioning, often resulting in ambiguous descriptions, or lacking fine-grained details originally present in the image. This leads to a second challenge: validate the accuracy of the results found. Authors state that, in practice, human studies are often employed, but this approach is not scalable and effective to a real-life setting. Otherwise, automatic metrics are proposed.

Hessel et al. [24], Duarte et al. [15], and Draffan et al. [14] explore approaches to tackle the accuracy of machine-generated descriptions. Hessel et al. [24] investigated the use of a cross-modal model pretrained on image/caption pairs corpora consisting of human Likert-scale judgments from the web to be used as an offline evaluation metric for literal caption quality. Duarte et al. [15] addresses the use of semantic analysis tools to enable automatic comparison of contents by combining several semantic similarity measures computed from the descriptors extracted from the media and the textual description. Although this work focus on enhancing web accessibility evaluation, the algorithm developed can be employed for other contexts, as its focus is to identify if the description adequately describes the content for users that cannot perceive it. Combining the topics of evaluating web media content descriptions, Draffan et al. [14] discuss how AI can be helpful to improve the reliability of automatic accessibility checks. They propose an AI-based approach that supports additional checks when a possible mismatch between the content in the alt text and in the image happens. Results obtained by Hessel et al. [24] allowed them to outline one remaining challenge concerning the volatile quality obtained through different domains. Duarte et al. [15] discuss the further advancements on this area as improving the recall and specificity of the algorithm classifications. For that they draw some possible approaches, such as exploring the different types of relations between the content and its description, improving the semantic services used in the implementation as well as the assessment of more domains. Another topic pointed out by the authors is the importance to clearly define evaluation procedures and disambiguate all concepts required for evaluation. Concerning more comprehensive discussions, Hessel et al. [24] stress out the urge of exploring potential social biases of candidate generations.

Abou-Zahra et al. [1] discuss how digital accessibility can be empowered by AI-based systems, such as providing automatic translations or automatic captions for online videos. Even though the discussion about the quality is a current and relevant issue, they also convey that the growth of pre-indexed data repositories allows rapidly growing also of the accuracy of image captioning services. An important aspect that one should consider is that many of mainstream AI-based services are not specifically built for accessibility. Authors suggest that the accuracy and reliability of these services for accessibility purposes could be improved with training data that is specific to people with disabilities, such as content that has been evaluated with input from human experts – as partially explored by Hessel et al. [24].

4.3.3 Computer vision for media accessibility

Papers in this category addressed support for alternative text generation through different techniques of image and text recognition. Over the years, research on this area has focused on filling the gap of the lack of image descriptions found on the web. While it is possible to observe significant advances in this area, media content has been gaining a lot of ground, representing a considerable part of the content present on the web nowadays as well as gaining different shapes and uses.

Bigham et al. [11], Singh et al. [38] and Guinness et al. [21] address the lack of alternative text for images found on the web with the support of computer vision techniques. They all propose a similar structure that uses a main server to maintain and reuse alternative text, embedding it on web pages on-the-fly. One main challenge in this context is the poor quality of machine-generated descriptions. In Bigham et al. [11] their approach is complemented by a mechanism for users to request that an image be sent to a labelling service for labelling by humans. Singh et al. [38] uses a web-site scrapping technique to train a model based on the image and associate tags so to improve their quality. Finally, Guinness et al. [21] relies on providing, initially, alternative text retrieved from other web pages, not relying solely on computer vision. They argue that finding images having alt attributes and propagating this alt text to copies of the image that lack the description, produces human-quality captioning.

Researchers also shared some of the challenges they face as limitations or future work. For instance, Bigham et al. [11] and Singh et al. [38] highlighted the need to investigate effective ways to handle incorrect or ambiguous descriptions. Guinness et al. [21] identified underperformance when finding unique images that are not hosted in many locations, such as those in personal photos. For that they suggest using computer vision or other AI techniques to identify images that are not an exact match to the target image, but which are nonetheless highly similar, to expand the set of images for which captions can be retrieved. Another technical challenge identified by Guinness et al. [21] consists in the latency in their approach to provide a caption queue. A potential solution suggested by them is to add a single caption to each image on the page before fleshing out extra captions for the queues. Bigham et al. [11] and Singh et al. [38] opted for providing one alternative text at time, recurring to different methods to offer users other captions retrieved. Bigham et al. [11] planned to include a tool for web authors that will provide suggestions for alternative text and coordinate the labelling of images across an entire site. Finally, Bigham et al. [11] identify some non-technical challenges such as legal troubles faced by image search engines, copyright issues, and responsibility for the captions produced.

As highlighted, one of the challenges faced in media accessibility is to handle personal photos. This is particularly challenging not only because these images are not likely to be found on other web pages, but also because they carry a strong personal meaning. This scenario has been further amplified by the increased use of social networks, in which such images represent a large part of the content. Wu et al. [46], and Gleason et al. [19] investigated image descriptions in the context of social networks, or a particular type of media highly used on these types of systems, Memes [20]. Wu et al. [46] built a system that uses computer vision technology to identify basic objects and themes in photos on Facebook, and constructs alt text using the identified concepts. This research focused on improving machine-generated descriptions to better fit users' desires for more information about the images, with a higher-quality and more socially aware computer algorithm. Gleason et al. [20] explored image descriptions on Twitter, that, on the other hand, do not provide machine-generated descriptions for their content, leaving this task up to its users. The researchers created Twitter A11y, a browser extension to add alternative text on Twitter using six different methods, such as looking for previous descriptions for the same image on the web, OCR, or crowdsourcing. They identify a great trade off present in this context, as users tend to prefer the text recognition and automatic captioning methods because they were quick and often descriptive, while crowdsourcing produces the highest-quality alt text, but asking crowd workers to label images is likely prohibitively expensive at scale and might be perceived as too slow to wait for when browsing social media sites. In a follow-up research, Gleason et al. [19] investigated the accessibility of internet memes, exploring computer vision techniques, such as OCR, to extract text from these media and provide an audio macro meme. To preserve the emotional tone or humour embedded in the meme of this content, they also propose a set of structured questions for writing alt text of memes. Concerning their conclusion on avenues for this field, Wu et al. [46] highlights the need of constructing captions that not only list objects and themes but also reveal the relationship among them. In addition, as identity, emotion, and appearance are personal, social, and fluid, it is extremely difficult to train computer algorithms to interpret these concepts in context. Gleason et al. [19] identified the need to explore multi-modal representations of, not only memes, but online content in general, that, especially with the increased use of social networks, has become more diverse. Gleason et al. [20] also pointed out the need of employing approaches to score the quality of image descriptions from multiple methods to ensure the best alt text is always returned. Authors also reported some concerns about privacy [46], and inaccuracies - that are not easily noticed by people with vision impairments [20]. Finally, both Wu et al. [46] and Gleason et al. [20] discuss about ethical issues, such as agency and accommodations. The first one being related to the decision to design an AI system that acts on behalf of the photo owner to describe to blind people what the image is about. The second one towards distinguishing accessibility and accommodation, and the importance to consider additional accessibility features and user education that could improve accessibility, not just rely on accommodations such as scene description methods.

Recent research efforts identified the relevance of the context on media accessibility alternative descriptions. On that note, Wang et al. [44] and Sreedhar et al. [42] investigated the use of contextual information to improve image descriptions in e-commerce platforms. Wang et al. [44] employed a rule-based approach using customer reviews to improve image descriptions for online products. Computer vision techniques were used to provide details such as colour and shape, however, this information was only used in case there were not enough reviews to extract further information. Further work concerning computer vision considers including OCR techniques to extract textual information in the images provided by sellers. Sreedhar et al. [42] used computer vision techniques to capture the context of the scene automatically, including key objects in the scene to feed an image captioning model. Sreedhar et al. [42] suggest improving their approach by verifying the usefulness of a customization of accessibility features such as the balance between product-specific and scene-specific descriptors.

While the works described focused on employing technical solutions to handle the accessibility of media content, Draffan et al. [14] and Abou-Zahra et al. [1] discuss potential avenues for this field. Draffan et al. [14] summarizes opportunities to update a web accessibility evaluator with current needs that are very well fit to this discussion. For instance, the use of AI models to support the evaluation of the accuracy of the alternative text for images. This can be of assistance when handling false negatives, such as when the alternative text description is considered as accepted by a tool. By adding additional checks, it is possible to notify the evaluator to what appears to be a mismatch between the content in the alt text and the actual image. Abou-Zahra et al. [1] also pointed out that the accuracy of automatic image recognition should not only be measured against which objects are depicted in images, but also against how well resulting text alternatives serve the equivalent purpose of images. This statement can be applied not only on the evaluation context, but on quality measurements in general. Finally, Abou-Zahra et al. [1] summarizes some limitations found by studies analysed through this section stating that a significant drawback of artificial intelligence for web accessibility at this time is a lack of accuracy and reliability.

5 Discussion

In what follows, we will discuss how these findings can be used to answer our previously established research questions.

5.1 How current AI research addresses digital accessibility?

From the analysis conducted it was possible to observe that several studies analyzed included users with disabilities, mostly blind and low vision users. Even though in a smaller number, it was possible to identify a positive practice of also including users without disabilities. These users represent a key role on this context, such as web designers and developers, and content authors, as they are the ones in charge of creating (or not) an accessible web.

During this analysis, it was possible to observe that several studies addressed their research topic employing different AI techniques. Most efforts were concentrated in machine learning, natural language processing, computer vision, and deep learning. Concerning digital accessibility, studies were mainly focused on web accessibility evaluation, and media accessibility. Other topics worth mentioning were voice browsing, and accessible communication.

AI was used on the studies on **web accessibility evaluation** to:

- support web page sampling to enhance the representativeness of the pages to be evaluated,
- predict accessibility results for the whole website based on previous results of evaluations conducted by experts,
- improve web accessibility metrics to better match user experience, and
- improve web components identification and classification to better improve accessibility of dynamic web pages.

The topic of media accessibility was also highly explored with NLP and computer vision techniques. These technologies combined are essential to improve current automated approaches to improve **media accessibility**. Researchers investigated:

- media alternatives in different contexts, such as social media, e-commerce platforms, as well as different types of medias, such as images, videos, and internet memes,
- speech-based media entry system, supporting searching on different languages,
- approaches to improve and assess the quality of machine-generated descriptions – as transposing image concepts obtained by image recognition to textual descriptions serving the equivalent purpose of images is not trivial,
- using contextual information to augment machine-generated descriptions, and
- training models with image descriptions generated by humans.

5.2 How current AI research can leverage digital accessibility?

Some potential avenues to AI leveraging accessibility were present in studies addressing all the domains analyzed.

Researchers on different domains identified that current **AI and digital accessibility research** can be further improved by:

- investigating how results obtained can be refined through exploring different AI methods, as results may vary according to it, and better results can be obtained,
- using real data to automatize and support processes that, currently, rely strictly on human judgment, and
- optimizing and reducing the costs of training machine learning models so to improve their performance and scalability.

It is important to emphasize that neither this analysis nor any of the papers included on this study advocate replacing human judgment with computer analysis, but rather using these techniques to optimize these processes.

Improving current models with real data also needs to consider the context of use. That is, data used for accessibility purposes must be gathered and trained according to its purpose. For instance, as previously mentioned, image description quality is highly influenced by the context in which this image is in. For that, the criteria to measure the quality of the alt text provided also needs to differ. In addition, one must consider the difference between describing the components present on an image and properly providing the equivalent purpose of images. Several approaches on this context were mentioned by researchers, such as using AI-based systems to:

- provide multi-modal representations of online content,
- explore the different types of relations between the content and its description,
- improve the semantic services used in the implementation as well as the assessment of more domains,
- investigate how user can benefit from customizable preferences for the level of the details provided, and
- support the evaluation of the accuracy of the alternative text for images, considering effective ways to handle incorrect or ambiguous descriptions.

On a different topic, researchers also envision the use of AI-based systems to:

- improve the identification and categorization of web components, to support dynamic content authoring as well as its evaluation.

Finally, one AI technique employed by one study that deserves more attention is:

- provide users with content on different languages through machine translation techniques.

5.3 How current AI research can hinder digital accessibility?

AI is a powerful tool with great potential to empower different stakeholders involved in the digital accessibility context: content authors, web designers and developers, accessibility practitioners, but, most important, users. While it is possible to perceive technical challenges as opportunities, some important aspects must be thoroughly discussed and investigated for future efforts to be made in the right direction.

As observed, one major drawback at this time is **the lack of accuracy and reliability**. One of the possible reasons for that is that AI-based services are not specifically built for accessibility. This can be evidenced by the number of studies reporting the challenge of conveying particular details present in an image, such as emotions and personal traits, as it is extremely difficult to train computer algorithms to interpret these concepts in context. While improving current models is a doable task and it can be perceived as an opportunity for the research community, as mentioned, it is important to consider the nuance that inaccuracy has for this community. While in other domains this can be easily perceived and further reported by users, researchers pointed out that inaccuracies are not well perceived by people with vision impairments. This leads us to tackle the urge of discussing the ethical issues with AI-based systems. Researchers reported the need of exploring potential **social biases**, **privacy**, and **social and legal responsibility** for accessed and generated data. Most of machine-generated approaches rely on image datasets and search engines provided by private companies. The legal implications concerning copyright issues are still unknown when employing this setting outside of a research scope, i.e., in a real-life setting. This could be a potential limitation to expand the capabilities of proposed approaches. Another challenge that comes along with this is summarized by researchers as identifying the boundary of algorithms. As previously mentioned, AI-based systems have the potential to support humans by automatizing several processes, and, for that, real data, i.e., provided by a human is needed. For instance, for improving the accuracy of machine-generated descriptions. However, each step must be carefully thought as several ethics concerns are involved. Users report that these descriptions are not currently providing enough details for perceiving the content of an image, and more personal details could be useful on that matter. However, for that, it is necessary to train algorithms and models with more personal data, and **privacy** issues are not yet widely discussed so that this development can take place in an ethical manner. Technologies are already capable of identifying people through face recognition approaches, but privacy implications still need to be assessed and furthermore accounted for. On that topic, it is first important to better understand the perspectives of photo owners and the implication on creative ownership. Secondly, a particularly important and sensitive subject is the potential **social bias** of classification algorithms. Researchers pointed out a recent case of a trained model that can make disproportionate incorrect classifications of people, e.g., “male images were misclassified into classes related to crime”. And for that, it is possible to conclude that we still need to further discuss how to address the **social and legal responsibility** when evolving current approaches.

Finally, it is essential to bear in mind that that future efforts on AI and digital accessibility must be conducted towards empowering users to access and produce accessible content rather than replacing human responsibility and agency.

6 Conclusions

Through a systematic literature review, we identified several key areas where AI can both enhance and hinder digital accessibility. While AI has the potential to automate processes currently reliant on human judgment and improve the performance and scalability of machine learning models, addressing challenges such as accuracy, reliability, social biases, privacy, and legal responsibility is crucial to ensure ethical and effective use of AI in digital accessibility. Our findings underscore the importance of ongoing research and development in this area and the need for collaborative efforts to bridge the gap between emerging technologies and accessibility solutions.

7 References

[1] Shadi Abou-Zahra, Judy Brewer, and Michael Cooper. 2018. Artificial Intelligence (AI) for Web Accessibility: Is Conformance Evaluation a Way Forward? In *Proceedings of the 15th International Web for All Conference*, ACM, Lyon France, 1–4. DOI:https://doi.org/10.1145/3192714.3192834

[2] M Elgin Akpinar and Yeliz Yeşilada. 2017. Discovering Visual Elements of Web Pages and Their Roles: Users' Perception. *Interact. Comput.* 29, 6 (November 2017), 845–867. DOI:https://doi.org/10.1093/iwc/iwx015

[3] Rodrigo Alarcon, Lourdes Moreno, and Paloma Martinez. 2021. Lexical Simplification System to Improve Web Accessibility. *IEEE Access* 9, (2021), 58755–58767. DOI:https://doi.org/10.1109/ACCESS.2021.3072697

[4] Humberto Lidio Antonelli, Rodrigo Augusto Igawa, Renata Pontin De Mattos Fortes, Eduardo Henrique Rizo, and Willian Massami Watanabe. 2018. Drop-Down Menu Widget Identification Using HTML Structure Changes Classification. *ACM Trans. Access. Comput.* 11, 2 (June 2018), 1–23. DOI:https://doi.org/10.1145/3178854

[5] Humberto Lidio Antonelli, Leonardo Sensiate, Willian Massami Watanabe, and Renata Pontin de Mattos Fortes. 2019. Challenges of automatically evaluating rich internet applications accessibility. In *Proceedings of the 37th ACM International Conference on the Design of Communication*, ACM, Portland Oregon, 1–6. DOI:https://doi.org/10.1145/3328020.3353950

[6] Aboubakr Aqle, Dena Al-Thani, and Ali Jaoua. 2022. Can search result summaries enhance the web search efficiency and experiences of the visually impaired users? *Univers. Access Inf. Soc.* 21, 1 (March 2022), 171–192. DOI:https://doi.org/10.1007/s10209-020-00777-w

[7] Vikas Ashok, Yevgen Borodin, Yury Puzis, and I. V. Ramakrishnan. 2015. Capti-speak: a speech-enabled web screen reader. In *Proceedings of the 12th International Web for All Conference*, ACM, Florence Italy, 1–10. DOI:https://doi.org/10.1145/2745555.2746660

[8] Vikas Ashok, Yury Puzis, Yevgen Borodin, and I.V. Ramakrishnan. 2017. Web Screen Reading Automation Assistance Using Semantic Abstraction. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces*, ACM, Limassol Cyprus, 407–418. DOI:https://doi.org/10.1145/3025171.3025229

[9] Ali Selman Aydin, Shirin Feiz, Vikas Ashok, and Iv Ramakrishnan. 2020. SaIL: saliency-driven injection of ARIA landmarks. In *Proceedings of the 25th International Conference on Intelligent User Interfaces*, ACM, Cagliari Italy, 111–115. DOI:https://doi.org/10.1145/3377325.3377540

[10] Mohammad Bajammal and Ali Mesbah. 2021. Semantic Web Accessibility Testing via Hierarchical Visual Analysis. In *2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE)*, IEEE, Madrid, ES, 1610–1621. DOI:https://doi.org/10.1109/ICSE43902.2021.00143

[11] Jeffrey P. Bigham, Ryan S. Kaminsky, Richard E. Ladner, Oscar M. Danielsson, and Gordon L. Hempton. 2006. WebInSight:: making web images accessible. In *Proceedings of the 8th international ACM SIGACCESS conference on Computers and accessibility -*

Assets '06, ACM Press, Portland, Oregon, USA, 181. DOI:https://doi.org/10.1145/1168987.1169018

[12] Jeffrey P. Bigham, Tessa Lau, and Jeffrey Nichols. 2009. Trailblazer: enabling blind users to blaze trails through the web. In *Proceedings of the 14th international conference on Intelligent user interfaces*, ACM, Sanibel Island Florida USA, 177–186. DOI:https://doi.org/10.1145/1502650.1502677

[13] Chaohai Ding, E. A. Draffan, and Mike Wald. 2020. AI and Global AAC Symbol Communication. In *Computers Helping People with Special Needs*, Klaus Miesenberger, Roberto Manduchi, Mario Covarrubias Rodriguez and Petr Peňáz (eds.). Springer International Publishing, Cham, 59–66. DOI:https://doi.org/10.1007/978-3-030-58796-3_8

[14] E. A. Draffan, Chaohai Ding, Mike Wald, Harry Everett, Jason Barrett, Abhirami Sasikant, Calin Geangu, and Russell Newman. 2020. Can a Web Accessibility Checker Be Enhanced by the Use of AI? In *Computers Helping People with Special Needs*, Klaus Miesenberger, Roberto Manduchi, Mario Covarrubias Rodriguez and Petr Peňáz (eds.). Springer International Publishing, Cham, 67–73. DOI:https://doi.org/10.1007/978-3-030-58796-3_9

[15] Carlos Duarte, Inês Matos, and Luís Carriço. 2018. Semantic Content Analysis Supporting Web Accessibility Evaluation. In *Proceedings of the 15th International Web for All Conference*, ACM, Lyon France, 1–4. DOI:https://doi.org/10.1145/3192714.3196828

[16] Eleni Efthimiou, Stavroula-Evita Fotinea, Panos Kakoulidis, Theodore Goulas, Athansia-Lida Dimou, and Anna Vacalopoulou. 2017. Sign Search and Sign Synthesis Made Easy to End User: The Paradigm of Building a SL Oriented Interface for Accessing and Managing Educational Content. In *Universal Access in Human–Computer Interaction. Designing Novel Interactions*, Margherita Antona and Constantine Stephanidis (eds.). Springer International Publishing, Cham, 14–26. DOI:https://doi.org/10.1007/978-3-319-58703-5_2

[17] Javedul Ferdous, Hae-Na Lee, Sampath Jayarathna, and Vikas Ashok. 2022. InSupport: Proxy Interface for Enabling Efficient Non-Visual Interaction with Web Data Records. In *27th International Conference on Intelligent User Interfaces*, ACM, Helsinki Finland, 49–62. DOI:https://doi.org/10.1145/3490099.3511126

[18] Javedul Ferdous, Sami Uddin, and Vikas Ashok. 2021. Semantic table-of-contents for efficient web screen reading. In *Proceedings of the 36th Annual ACM Symposium on Applied Computing*, ACM, Virtual Event Republic of Korea, 1941–1949. DOI:https://doi.org/10.1145/3412841.3442066

[19] 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*, ACM, Pittsburgh PA USA, 367–376. DOI:https://doi.org/10.1145/3308561.3353792

[20] 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*, ACM, Honolulu HI USA, 1–12. DOI:https://doi.org/10.1145/3313831.3376728

[21] 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*, ACM, Montreal QC Canada, 1–11. DOI:https://doi.org/10.1145/3173574.3174092

[22] Simon Harper, Anwar Ahmad Moon, Markel Vigo, Giorgio Brajnik, and Yeliz Yesilada. 2015. DOM block clustering for enhanced sampling and evaluation. In *Proceedings of the 12th International Web for All Conference*, ACM, Florence Italy, 1–10. DOI:https://doi.org/10.1145/2745555.2746649

[23] Xiaodong He and Li Deng. 2017. Deep Learning for Image-to-Text Generation: A Technical Overview. *IEEE Signal Process. Mag.* 34, 6 (November 2017), 109–116. DOI:https://doi.org/10.1109/MSP.2017.2741510

[24] Jack Hessel, Ari Holtzman, Maxwell Forbes, Ronan Le Bras, and Yejin Choi. 2021. CLIPScore: A Reference-free Evaluation Metric for Image Captioning. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, Online and Punta Cana, Dominican Republic, 7514–7528. DOI:https://doi.org/10.18653/v1/2021.emnlp-main.595

[25] Mina Huh, YunJung Lee, Dasom Choi, Haesoo Kim, Uran Oh, and Juho Kim. 2022. Cocomix: Utilizing Comments to Improve Non-Visual Webtoon Accessibility. In *CHI Conference on Human Factors in Computing Systems*, ACM, New Orleans LA USA, 1–18. DOI:https://doi.org/10.1145/3491102.3502081

[26] Claire Kearney-Volpe and Amy Hurst. 2021. Accessible Web Development: Opportunities to Improve the Education and Practice of web Development with a Screen Reader. *ACM Trans. Access. Comput.* 14, 2 (June 2021), 1–32. DOI:https://doi.org/10.1145/3458024

[27] Liangcheng Li, Can Wang, Shuyi Song, Zhi Yu, Fenqin Zhou, and Jiajun Bu. 2017. A task assignment strategy for crowdsourcing-based web accessibility evaluation system. In *Proceedings of the 14th International Web for All Conference*, ACM, Perth Western Australia Australia, 1–4. DOI:https://doi.org/10.1145/3058555.3058573

[28] Jalal U. Mahmud, Yevgen Borodin, and I. V. Ramakrishnan. 2007. Csurf: a context-driven non-visual web-browser. In *Proceedings of the 16th international conference on World Wide Web - WWW '07*, ACM Press, Banff, Alberta, Canada, 31. DOI:https://doi.org/10.1145/1242572.1242578

[29] Rafaela Mazalu and Alejandra Cechich. 2020. Web Accessibility Assessment Through Multi-Agent Support for Visually Impaired Users. *Int. J. Coop. Inf. Syst.* 29, 03 (September 2020), 2050005. DOI:https://doi.org/10.1142/S0218843020500057

[30] Emma J. McDonnell, Ping Liu, Steven M. Goodman, Raja Kushalnagar, Jon E. Froehlich, and Leah Findlater. 2021. Social, Environmental, and Technical: Factors at Play in the Current Use and Future Design of Small-Group Captioning. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2 (October 2021), 1–25. DOI:https://doi.org/10.1145/3479578

[31] Valentyn Melnyk, Vikas Ashok, Valentyn Melnyk, Yury Puzis, Yevgen Borodin, Andrii Soviak, and I. V. Ramakrishnan. 2015. Look Ma, no ARIA: generic accessible interfaces for web widgets. In *Proceedings of the 12th International Web for All Conference*, ACM, Florence Italy, 1–4. DOI:https://doi.org/10.1145/2745555.2746666

[32] Hendrik Meutzner, Santosh Gupta, Viet-Hung Nguyen, Thorsten Holz, and Dorothea Kolossa. 2017. Toward Improved Audio CAPTCHAs Based on Auditory Perception and

Language Understanding. *ACM Trans. Priv. Secur.* 19, 4 (February 2017), 1–31. DOI:https://doi.org/10.1145/2856820

[33] Amy Pavel, Gabriel Reyes, and Jeffrey P. Bigham. 2020. Rescribe: Authoring and Automatically Editing Audio Descriptions. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, ACM, Virtual Event USA, 747–759. DOI:https://doi.org/10.1145/3379337.3415864

[34] Yi-Hao Peng, Jeffrey P Bigham, and Amy Pavel. 2021. Slidecho: Flexible Non-Visual Exploration of Presentation Videos. In *The 23rd International ACM SIGACCESS Conference on Computers and Accessibility*, ACM, Virtual Event USA, 1–12. DOI:https://doi.org/10.1145/3441852.3471234

[35] Yi-Hao Peng, JiWoong Jang, Jeffrey P Bigham, and Amy Pavel. 2021. Say It All: Feedback for Improving Non-Visual Presentation Accessibility. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, ACM, Yokohama Japan, 1–12. DOI:https://doi.org/10.1145/3411764.3445572

[36] Xin Qian, Eunyee Koh, Fan Du, Sungchul Kim, Joel Chan, Ryan A. Rossi, Sana Malik, and Tak Yeon Lee. 2021. Generating Accurate Caption Units for Figure Captioning. In *Proceedings of the Web Conference 2021*, ACM, Ljubljana Slovenia, 2792–2804. DOI:https://doi.org/10.1145/3442381.3449923

[37] Eduardo Henrique Rizo, Renata Pontin de Mattos Fortes, Humberto Lidio Antonelli, and Willian Massami Watanabe. 2019. Automatic Identification of Widgets and their Subcomponents Based on a Classification Pipeline for DOM Mutation Records. In *Proceedings of the 16th International Web for All Conference*, ACM, San Francisco CA USA, 1–10. DOI:https://doi.org/10.1145/3315002.3317555

[38] Shivam Singh, Anurag Bhandari, and Nishith Pathak. 2018. Accessify: An ML Powered Application to Provide Accessible Images on Web Sites. In *Proceedings of the 15th International Web for All Conference*, ACM, Lyon France, 1–4. DOI:https://doi.org/10.1145/3192714.3192830

[39] Osama Sohaib, Mohsen Naderpour, and Walayat Hussain. 2018. SaaS E-Commerce Platforms Web Accessibility Evaluation. In *2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, IEEE, Rio de Janeiro, 1–7. DOI:https://doi.org/10.1109/FUZZ-IEEE.2018.8491621

[40] Shuyi Song, Jiajun Bu, Chengchao Shen, Andreas Artmeier, Zhi Yu, and Qin Zhou. 2018. Reliability Aware Web Accessibility Experience Metric. In *Proceedings of the 15th International Web for All Conference*, ACM, Lyon France, 1–4. DOI:https://doi.org/10.1145/3192714.3192836

[41] Shuyi Song, Can Wang, Liangcheng Li, Zhi Yu, Xiao Lin, and Jiajun Bu. 2017. WAEM: A Web Accessibility Evaluation Metric Based on Partial User Experience Order. In *Proceedings of the 14th International Web for All Conference*, ACM, Perth Western Australia Australia, 1–4. DOI:https://doi.org/10.1145/3058555.3058576

[42] Rachana Sreedhar, Nicole Tan, Jingyue Zhang, Kim Jin, Spencer Gregson, Eli Moreta-Feliz, Niveditha Samudrala, and Shrenik Sadalgi. 2022. AIDE: automatic and accessible image descriptions for review imagery in online retail. In *Proceedings of the 19th International Web for All Conference*, ACM, Lyon France, 1–8. DOI:https://doi.org/10.1145/3493612.3520453

[43] Carlos Teixeira, Joss Moorkens, Daniel Turner, Joris Vreeke, and Andy Way. 2019. Creating a Multimodal Translation Tool and Testing Machine Translation Integration Using Touch and Voice. *Informatics* 6, 1 (March 2019), 13. DOI:https://doi.org/10.3390/informatics6010013

[44] Ruolin Wang, Zixuan Chen, Mingrui Ray Zhang, Zhaoheng Li, Zhixiu Liu, Zihan Dang, Chun Yu, and Xiang “Anthony” Chen. 2021. Revamp: Enhancing Accessible Information Seeking Experience of Online Shopping for Blind or Low Vision Users. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, ACM, Yokohama Japan, 1–14. DOI:https://doi.org/10.1145/3411764.3445547

[45] Yujia Wang, Wei Liang, Haikun Huang, Yongqi Zhang, Dingzeyu Li, and Lap-Fai Yu. 2021. Toward Automatic Audio Description Generation for Accessible Videos. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, ACM, Yokohama Japan, 1–12. DOI:https://doi.org/10.1145/3411764.3445347

[46] 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*, ACM, Portland Oregon USA, 1180–1192. DOI:https://doi.org/10.1145/2998181.2998364

[47] Yue Wu, Zhi Yu, Liangcheng Li, Wei Wang, Jiajun Bu, Yueqing Zhang, and Lianjun Dai. 2018. Using Semi-supervised Group Sparse Regression to Improve Web Accessibility Evaluation. In *Computers Helping People with Special Needs*, Klaus Miesenberger and Georgios Kouroupetroglou (eds.). Springer International Publishing, Cham, 52–56. DOI:https://doi.org/10.1007/978-3-319-94277-3_10

[48] Chun Yu and Jiajun Bu. 2021. The practice of applying AI to benefit visually impaired people in China. *Commun. ACM* 64, 11 (November 2021), 70–75. DOI:https://doi.org/10.1145/3481623

[49] Zhi Yu, Jiajun Bu, Chao Shen, Wei Wang, Lianjun Dai, Qin Zhou, and Chuanwu Zhao. 2020. A Multi-site Collaborative Sampling for Web Accessibility Evaluation. In *Computers Helping People with Special Needs*, Klaus Miesenberger, Roberto Manduchi, Mario Covarrubias Rodriguez and Petr Peňáz (eds.). Springer International Publishing, Cham, 329–335. DOI:https://doi.org/10.1007/978-3-030-58796-3_39

[50] Mengni Zhang, Can Wang, Jiajun Bu, Zhi Yu, Yi Lu, Ruijie Zhang, and Chun Chen. 2015. An optimal sampling method for web accessibility quantitative metric. In *Proceedings of the 12th International Web for All Conference*, ACM, Florence Italy, 1–4. DOI:https://doi.org/10.1145/2745555.2746663

[51] Meng-ni Zhang, Can Wang, Jia-jun Bu, Zhi Yu, Yu Zhou, and Chun Chen. 2015. A sampling method based on URL clustering for fast web accessibility evaluation. *Front. Inf. Technol. Electron. Eng.* 16, 6 (June 2015), 449–456. DOI:https://doi.org/10.1631/FITEE.1400377

[52] Mengni Zhang, Can Wang, Zhi Yu, Chao Shen, and Jiajun Bu. 2017. Active learning for Web accessibility evaluation. In *Proceedings of the 14th International Web for All Conference*, ACM, Perth Western Australia Australia, 1–9. DOI:https://doi.org/10.1145/3058555.3058559

[53] Mingrui Ray Zhang, Ruolin Wang, Xuhai Xu, Qisheng Li, Ather Sharif, and Jacob O. Wobbrock. 2021. Voicemoji: Emoji Entry Using Voice for Visually Impaired People. In

Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, ACM, Yokohama Japan, 1–18. DOI:https://doi.org/10.1145/3411764.3445338

[54] Shaojian Zhu, Daisuke Sato, Hironobu Takagi, and Chieko Asakawa. 2010. Sasayaki: an augmented voice-based web browsing experience. In *Proceedings of the 12th international ACM SIGACCESS conference on Computers and accessibility - ASSETS '10*, ACM Press, Orlando, Florida, USA, 279. DOI:https://doi.org/10.1145/1878803.1878870

8 Annex 1

Table 12: Search terms used

Topic	Search terms
Artificial Intelligence	
Natural language processing	(“artificial intelligence”) AND (“natural language processing” OR “NLP”)
Knowledge representation and reasoning	(“artificial intelligence”) AND (“knowledge representation and reasoning” OR “KRR”)
Planning and scheduling	(“artificial intelligence”) AND (“planning and scheduling” OR “AI planning” OR “automated planning” OR “APS”)
Search methodologies	(“artificial intelligence”) AND (“search technique*” OR “search method*”)
Control methods	(“artificial intelligence”) AND (“control method*”)
Philosophical/theoretical foundations of artificial intelligence	(“artificial intelligence”) AND (“machine learning approach*”)
Distributed artificial intelligence	(“artificial intelligence”) AND (“distributed artificial intelligence” OR “distributed AI” OR “DAI”)
Computer vision	(“artificial intelligence”) AND (“computer vision” OR “machine vision”)
Learning paradigms	(“artificial intelligence”) AND (“learning paradigm*” OR “paradigm* of learning”)
Learning settings	(“artificial intelligence”) AND (“learning setting*”)
Machine learning approaches	(“artificial intelligence”) AND (“machine learning approach*”)

Machine learning algorithms	("artificial intelligence") AND ("machine learning algorithm*")
Cross-validation	("artificial intelligence") AND ("cross validation" OR "cross-validation")
Machine learning	"machine learning"
Deep learning	"deep learning"
AND	
Digital accessibility	"web accessibility" OR "web accessible" OR "accessible web" OR "digital accessibility" OR "a11y" OR "accessibility on the web" OR "accessibility of websites" OR "accessible website"

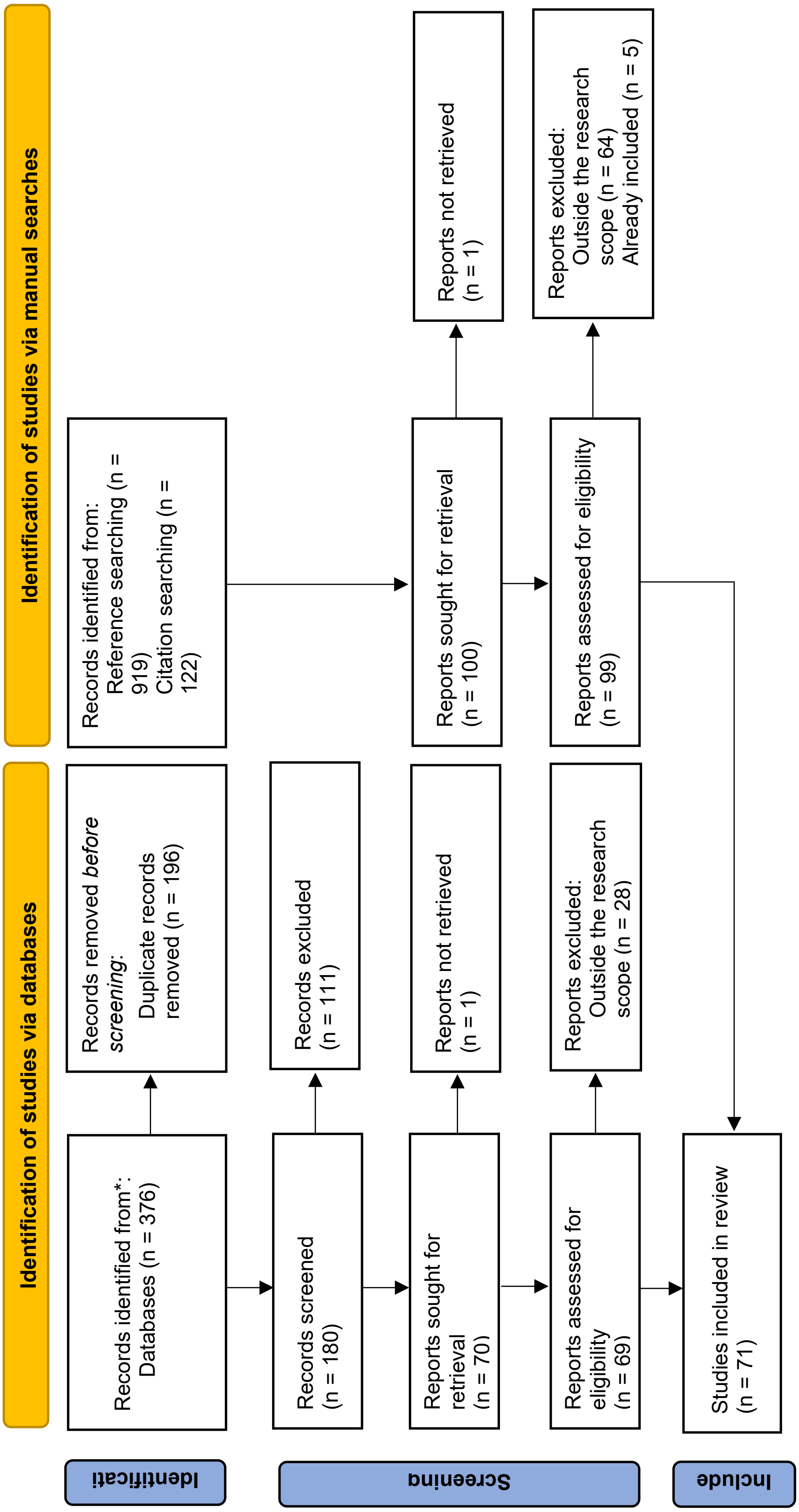
9 Annex 2

Table 13: Database search information

Database	Search date	Filters on
ACM Digital Library	25 February 2022	Article Type: Research Article Publication Date: 01/01/2017 to * ACM Content: DL
IEEE Xplore	25 February 2022	Conferences, Journals, Early Access Articles Year: 2017-2022
Web of Science	25 February 2022	Document Type: Articles or Proceedings Papers or Early Access Publication Years: 2022 or 2021 or 2020 or 2019 or 2018 or 2017
ScienceDirect	25 February 2022	Article Type: Research articles Years: 2022, 2021, 2020, 2019, 2018, 2017
Scopus	25 February 2022	Doctype: Article, Conference Paper Year: 2022, 2021, 2020, 2019, 2018, 2017

10 Annex 3

PRISMA diagram



11 Annex 4

List of papers included:

(1) Abou-Zahra, S. et al. 2018. Artificial Intelligence (AI) for Web Accessibility: Is Conformance Evaluation a Way Forward? Proceedings of the 15th International Web for All Conference (Lyon France, Apr. 2018), 1–4.

(2) Akpinar, M.E. and Yeşilada, Y. 2017. Discovering Visual Elements of Web Pages and Their Roles: Users' Perception. Interacting with Computers. 29, 6 (Nov. 2017), 845–867. DOI:https://doi.org/10.1093/iwc/iwx015.

(3) Alarcon, R. et al. 2021. Lexical Simplification System to Improve Web Accessibility. IEEE Access. 9, (2021), 58755–58767. DOI:https://doi.org/10.1109/ACCESS.2021.3072697.

(4) AlSadoun, W. et al. 2020. Towards Intelligent Technology in Art Therapy Contexts. HCI International 2020 - Late Breaking Papers: Multimodality and Intelligence. C. Stephanidis et al., eds. Springer International Publishing. 397–405.

(5) Antonelli, H.L. et al. 2018. Drop-Down Menu Widget Identification Using HTML Structure Changes Classification. ACM Transactions on Accessible Computing. 11, 2 (Jun. 2018), 1–23. DOI:https://doi.org/10.1145/3178854.

(6) Antonelli, H.L. et al. 2019. Challenges of automatically evaluating rich internet applications accessibility. Proceedings of the 37th ACM International Conference on the Design of Communication (Portland Oregon, Oct. 2019), 1–6.

(7) Aqle, A. et al. 2022. Can search result summaries enhance the web search efficiency and experiences of the visually impaired users? Universal Access in the Information Society. 21, 1 (Mar. 2022), 171–192. DOI:https://doi.org/10.1007/s10209-020-00777-w.

(8) Ashok, V. et al. 2015. Capti-speak: a speech-enabled web screen reader. Proceedings of the 12th International Web for All Conference (Florence Italy, May 2015), 1–10.

(9) Ashok, V. et al. 2017. Web Screen Reading Automation Assistance Using Semantic Abstraction. Proceedings of the 22nd International Conference on Intelligent User Interfaces (Limassol Cyprus, Mar. 2017), 407–418.

(10) Aydin, A.S. et al. 2020. SaIL: saliency-driven injection of ARIA landmarks. Proceedings of the 25th International Conference on Intelligent User Interfaces (Cagliari Italy, Mar. 2020), 111–115.

(11) Azmi, A.M. and Aljafari, E.A. 2018. Universal web accessibility and the challenge to integrate informal Arabic users: a case study. Universal Access in the Information Society. 17, 1 (Mar. 2018), 131–145. DOI:https://doi.org/10.1007/s10209-017-0522-3.

(12) Baez, M. et al. 2020. Conversational Web Interaction: Proposal of a Dialog-Based Natural Language Interaction Paradigm for the Web. Chatbot Research and Design. A. Følstad et al., eds. Springer International Publishing. 94–110.

(13) Baez, M. et al. 2021. Chatbot Integration in Few Patterns. IEEE Internet Computing. 25, 3 (May 2021), 52–59. DOI:https://doi.org/10.1109/MIC.2020.3024605.

(14) Bajammal, M. and Mesbah, A. 2021. Semantic Web Accessibility Testing via Hierarchical Visual Analysis. 2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE) (Madrid, ES, May 2021), 1610–1621.

(15) Bigham, J.P. et al. 2006. WebInSight:: making web images accessible. Proceedings of the 8th international ACM SIGACCESS conference on Computers and accessibility - Assets '06 (Portland, Oregon, USA, 2006), 181.

(16) Bigham, J.P. et al. 2009. Trailblazer: enabling blind users to blaze trails through the web. Proceedings of the 14th international conference on Intelligent user interfaces (Sanibel Island Florida USA, Feb. 2009), 177–186.
(17) Borodin, Y. et al. 2007. The HearSay non-visual web browser. Proceedings of the 2007 international cross-disciplinary conference on Web accessibility (W4A) - W4A '07 (Banff, Canada, 2007), 128–129.
(18) Chittò, P. et al. 2020. Automatic Generation of Chatbots for Conversational Web Browsing. Conceptual Modeling. G. Dobbie et al., eds. Springer International Publishing. 239–249.
(19) Cormier, M. et al. 2016. Purely vision-based segmentation of web pages for assistive technology. Computer Vision and Image Understanding. 148, (Jul. 2016), 46–66. DOI:https://doi.org/10.1016/j.cviu.2016.02.007.
(20) De Los Rios Perez, C. et al. 2017. Holistic Approach for Sustainable Adaptable User Interfaces for People with Autism Spectrum Disorder. Proceedings of the 26th International Conference on World Wide Web Companion - WWW '17 Companion (Perth, Australia, 2017), 1553–1556.
(21) Ding, C. et al. 2020. AI and Global AAC Symbol Communication. Computers Helping People with Special Needs. K. Miesenberger et al., eds. Springer International Publishing. 59–66.
(22) Draffan, E.A. et al. 2020. Can a Web Accessibility Checker Be Enhanced by the Use of AI? Computers Helping People with Special Needs. K. Miesenberger et al., eds. Springer International Publishing. 67–73.
(23) Duarte, C. et al. 2018. Semantic Content Analysis Supporting Web Accessibility Evaluation. Proceedings of the 15th International Web for All Conference (Lyon France, Apr. 2018), 1–4.
(24) Efthimiou, E. et al. 2017. Sign Search and Sign Synthesis Made Easy to End User: The Paradigm of Building a SL Oriented Interface for Accessing and Managing Educational Content. Universal Access in Human–Computer Interaction. Designing Novel Interactions. M. Antona and C. Stephanidis, eds. Springer International Publishing. 14–26.
(25) Ferdous, J. et al. 2021. Semantic table-of-contents for efficient web screen reading. Proceedings of the 36th Annual ACM Symposium on Applied Computing (Virtual Event Republic of Korea, Mar. 2021), 1941–1949.
(26) Ferdous, J. et al. 2022. InSupport: Proxy Interface for Enabling Efficient Non-Visual Interaction with Web Data Records. 27th International Conference on Intelligent User Interfaces (Helsinki Finland, Mar. 2022), 49–62.
(27) Gleason, C. et al. 2019. Making Memes Accessible. The 21st International ACM SIGACCESS Conference on Computers and Accessibility (Pittsburgh PA USA, Oct. 2019), 367–376.
(28) Gleason, C. et al. 2020. Twitter A11y: A Browser Extension to Make Twitter Images Accessible. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (Honolulu HI USA, Apr. 2020), 1–12.
(29) González-Mora, C. et al. 2021. Open Data Accessibility Based on Voice Commands. Web Engineering. M. Brambilla et al., eds. Springer International Publishing. 456–463.
(30) Guinness, D. et al. 2018. Caption Crawler: Enabling Reusable Alternative Text Descriptions using Reverse Image Search. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (Montreal QC Canada, Apr. 2018), 1–11.

(31) Harper, S. et al. 2015. DOM block clustering for enhanced sampling and evaluation. Proceedings of the 12th International Web for All Conference (Florence Italy, May 2015), 1–10.
(32) He, X. and Deng, L. 2017. Deep Learning for Image-to-Text Generation: A Technical Overview. IEEE Signal Processing Magazine. 34, 6 (Nov. 2017), 109–116. DOI:https://doi.org/10.1109/MSP.2017.2741510.
(33) Hessel, J. et al. 2021. CLIPScore: A Reference-free Evaluation Metric for Image Captioning. Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (Online and Punta Cana, Dominican Republic, 2021), 7514–7528.
(34) Huh, M. et al. 2022. Cocomix: Utilizing Comments to Improve Non-Visual Webtoon Accessibility. CHI Conference on Human Factors in Computing Systems (New Orleans LA USA, Apr. 2022), 1–18.
(35) Kearney-Volpe, C. and Hurst, A. 2021. Accessible Web Development: Opportunities to Improve the Education and Practice of web Development with a Screen Reader. ACM Transactions on Accessible Computing. 14, 2 (Jun. 2021), 1–32. DOI:https://doi.org/10.1145/3458024.
(36) Li, L. et al. 2017. A task assignment strategy for crowdsourcing-based web accessibility evaluation system. Proceedings of the 14th International Web for All Conference (Perth Western Australia Australia, Apr. 2017), 1–4.
(37) Lundgard, A. 2020. Measuring justice in machine learning. Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (Barcelona Spain, Jan. 2020), 680–680.
(38) Madaan, H. and Gupta, S. 2021. AI Improving the Lives of Physically Disabled. Proceedings of the 12th International Conference on Soft Computing and Pattern Recognition (SoCPaR 2020). A. Abraham et al., eds. Springer International Publishing. 103–112.
(39) Mahmud, J.U. et al. 2007. Csurf: a context-driven non-visual web-browser. Proceedings of the 16th international conference on World Wide Web - WWW '07 (Banff, Alberta, Canada, 2007), 31.
(40) Mazalu, R. and Cechich, A. 2020. Web Accessibility Assessment Through Multi-Agent Support for Visually Impaired Users. International Journal of Cooperative Information Systems. 29, 03 (Sep. 2020), 2050005. DOI:https://doi.org/10.1142/S0218843020500057.
(41) McDonnell, E.J. et al. 2021. Social, Environmental, and Technical: Factors at Play in the Current Use and Future Design of Small-Group Captioning. Proceedings of the ACM on Human-Computer Interaction. 5, CSCW2 (Oct. 2021), 1–25. DOI:https://doi.org/10.1145/3479578.
(42) Melnyk, V. et al. 2014. Widget Classification with Applications to Web Accessibility. Web Engineering. S. Casteleyn et al., eds. Springer International Publishing. 341–358.
(43) Melnyk, V. et al. 2015. Look Ma, no ARIA: generic accessible interfaces for web widgets. Proceedings of the 12th International Web for All Conference (Florence Italy, May 2015), 1–4.
(44) Meutzner, H. et al. 2017. Toward Improved Audio CAPTCHAs Based on Auditory Perception and Language Understanding. ACM Transactions on Privacy and Security. 19, 4 (Feb. 2017), 1–31. DOI:https://doi.org/10.1145/2856820.

(45) Pavel, A. et al. 2020. Rescribe: Authoring and Automatically Editing Audio Descriptions. Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology (Virtual Event USA, Oct. 2020), 747–759.
(46) Peng, Y.-H. et al. 2021. Say It All: Feedback for Improving Non-Visual Presentation Accessibility. Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (Yokohama Japan, May 2021), 1–12.
(47) Peng, Y.-H. et al. 2021. Slidecho: Flexible Non-Visual Exploration of Presentation Videos. The 23rd International ACM SIGACCESS Conference on Computers and Accessibility (Virtual Event USA, Oct. 2021), 1–12.
(48) Pina, A. et al. 2021. Bringing Cognitive Augmentation to Web Browsing Accessibility. Service-Oriented Computing – ICSOC 2020 Workshops. H. Hacid et al., eds. Springer International Publishing. 395–407.
(49) Prasad, G.V.S. 2017. Renarrating Web Content to Increase Web Accessibility. Proceedings of the 10th International Conference on Theory and Practice of Electronic Governance (New Delhi AA India, Mar. 2017), 598–601.
(50) Qian, X. et al. 2021. Generating Accurate Caption Units for Figure Captioning. Proceedings of the Web Conference 2021 (Ljubljana Slovenia, Apr. 2021), 2792–2804.
(51) Rizo, E.H. et al. 2019. Automatic Identification of Widgets and their Subcomponents Based on a Classification Pipeline for DOM Mutation Records. Proceedings of the 16th International Web for All Conference (San Francisco CA USA, May 2019), 1–10.
(52) Rizvi, N. et al. 2019. Margin call: An accessible web-based text viewer with generated paragraph summaries in the margin. INLG 2019 - 12th International Conference on Natural Language Generation, Proceedings of the Conference.
(53) Saggion, H. et al. 2017. Able to Read My Mail: An Accessible e-Mail Client with Assistive Technology. Proceedings of the 14th International Web for All Conference (Perth Western Australia Australia, Apr. 2017), 1–4.
(54) Shaqoor Nengroo, A. and Kuppusamy, K.S. 2018. Machine learning based heterogeneous web advertisements detection using a diverse feature set. Future Generation Computer Systems. 89, (Dec. 2018), 68–77. DOI:https://doi.org/10.1016/j.future.2018.06.028.
(55) Singh, S. et al. 2018. Accessify: An ML Powered Application to Provide Accessible Images on Web Sites. Proceedings of the 15th International Web for All Conference (Lyon France, Apr. 2018), 1–4.
(56) Sohaib, O. et al. 2018. SaaS E-Commerce Platforms Web Accessibility Evaluation. 2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE) (Rio de Janeiro, Jul. 2018), 1–7.
(57) Song, S. et al. 2017. WAEM: A Web Accessibility Evaluation Metric Based on Partial User Experience Order. Proceedings of the 14th International Web for All Conference (Perth Western Australia Australia, Apr. 2017), 1–4.
(58) Song, S. et al. 2018. Reliability Aware Web Accessibility Experience Metric. Proceedings of the 15th International Web for All Conference (Lyon France, Apr. 2018), 1–4.
(59) Sreedhar, R. et al. 2022. AIDE: automatic and accessible image descriptions for review imagery in online retail. Proceedings of the 19th International Web for All Conference (Lyon France, Apr. 2022), 1–8.
(60) Teixeira, C. et al. 2019. Creating a Multimodal Translation Tool and Testing Machine Translation Integration Using Touch and Voice. Informatics. 6, 1 (Mar. 2019), 13. DOI:https://doi.org/10.3390/informatics6010013.

(61) Wang, R. et al. 2021. Revamp: Enhancing Accessible Information Seeking Experience of Online Shopping for Blind or Low Vision Users. Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (Yokohama Japan, May 2021), 1–14.
(62) Wang, Y. et al. 2021. Toward Automatic Audio Description Generation for Accessible Videos. Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (Yokohama Japan, May 2021), 1–12.
(63) Wu, S. et al. 2017. Automatic Alt-text: Computer-generated Image Descriptions for Blind Users on a Social Network Service. Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (Portland Oregon USA, Feb. 2017), 1180–1192.
(64) Wu, Y. et al. 2018. Using Semi-supervised Group Sparse Regression to Improve Web Accessibility Evaluation. Computers Helping People with Special Needs. K. Miesenberger and G. Kouroupetroglou, eds. Springer International Publishing. 52–56.
(65) Yu, C. and Bu, J. 2021. The practice of applying AI to benefit visually impaired people in China. Communications of the ACM. 64, 11 (Nov. 2021), 70–75. DOI:https://doi.org/10.1145/3481623.
(66) Yu, Z. et al. 2020. A Multi-site Collaborative Sampling for Web Accessibility Evaluation. Computers Helping People with Special Needs. K. Miesenberger et al., eds. Springer International Publishing. 329–335.
(67) Zhang, M. et al. 2015. A sampling method based on URL clustering for fast web accessibility evaluation. Frontiers of Information Technology & Electronic Engineering. 16, 6 (Jun. 2015), 449–456. DOI:https://doi.org/10.1631/FITEE.1400377.
(68) Zhang, M. et al. 2015. An optimal sampling method for web accessibility quantitative metric. Proceedings of the 12th International Web for All Conference (Florence Italy, May 2015), 1–4.
(69) Zhang, M. et al. 2017. Active learning for Web accessibility evaluation. Proceedings of the 14th International Web for All Conference (Perth Western Australia Australia, Apr. 2017), 1–9.
(70) Zhang, M.R. et al. 2021. Voicemoji: Emoji Entry Using Voice for Visually Impaired People. Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (Yokohama Japan, May 2021), 1–18.
(71) Zhu, S. et al. 2010. Sasayaki: an augmented voice-based web browsing experience. Proceedings of the 12th international ACM SIGACCESS conference on Computers and accessibility - ASSETS ’10 (Orlando, Florida, USA, 2010), 279.

12 Annex 5

Detailed information about the AI subdomains and digital accessibility application domains.

Table 14: Natural Language Processing and digital accessibility application domains.

NLP and digital accessibility	Total
media accessibility	11
voice browsing	7
accessible communication	6
web accessibility evaluation	5
web browsing	3
AAC	2
image description quality	2
web information retrieval	1
captcha	1
personalization	1
accessible presentation	1
mediated system	1
easy to read	1
information retrieval	1
web accessibility	1
text simplification	1
gesture interaction	1
pictograms	1
accessible learning	1

Table 15: Computer vision and digital accessibility application domains.

Computer vision and digital accessibility	Total
media accessibility	10
web accessibility evaluation	4

accessible presentation	2
voice browsing	1
blind programming	1
assistive technologies	1
information retrieval	1
gesture interaction	1
accessible learning	1
WAI-ARIA	1
dynamic web content	1
accessible communication	1
AAC	1

Table 16: Deep learning and digital accessibility application domains.

Deep learning and digital accessibility	**Total**
media accessibility	6
web accessibility evaluation	3
WAI-ARIA	3
voice browsing	2
accessible communication	1
web accessibility metrics	1
web accessibility	1
AAC	1
open data	1
gesture interaction	1
image description quality	1
translation	1
dynamic content	1

Table 17: Machine learning and digital accessibility application domains.

Machine learning and digital accessibility	Total
web accessibility evaluation	15
WAI-ARIA	7
media accessibility	7
accessible communication	5
sampling	5
web accessibility	3
voice browsing	3
dynamic content	3
web accessibility metrics	2
information retrieval	2
AAC	2
widgets	2
interface adaptation	2
image description quality	2
text simplification	2
captcha	1
mediated system	1
inclusion	1
multimodal interaction	1
blind programming	1
web browsing	1
scoring	1
dynamic web content	1
data records	1
translation	1
accessible learning	1

pictograms	1
open data	1