Advances in Science and Technology Research Journal

Advances in Science and Technology Research Journal, 2026, 20(2), 33–52
https://doi.org/10.12913/22998624/210532
ISSN 2299-8624, License CC-BY 4.0

Received: 2025.07.08
Accepted: 2025.11.15
Published: 2025.12.23

Human-computer interaction in scientific journal websites: A usability and accessibility analysis of human-designed and artificial intelligence – generated user interfaces

Konrad Skolimowski[1], Bartosz Rola[1], Piotr Tokarski[1*], Karolina Rybak[1], Karol Łazaruk[1], Małgorzata Plechawska-Wójcik[1], Mariusz Dzieńkowski[1]

[1] Department of Computer Science, Lublin University of Technology, ul. Nadbystrzycka 38d, 20-618 Lublin, Poland

* Corresponding author's e-mail: p.tokarski@pollub.pl

ABSTRACT

The issue of usability and accessibility in scientific journals is of crucial importance for effective knowledge communication, with a direct impact on user satisfaction. These aspects must be considered at every stage of website development, from needs analysis and planning, through design and implementation, to testing, maintenance, and further development. In the contemporary digital landscape, a discernible trend has emerged for websites to be designed by the principles of Universal Design (UD). Furthermore, tools based on artificial intelligence (AI) are playing an increasingly significant role in their development. The present study involved the execution of two independent experiments with several people overlapping in both groups. In the first study, a comparative analysis was conducted between the existing website of a scientific journal and a prototype manually developed by a designer, adhering to the UD principles. In the second experiment, the existing website was compared with two prototypes generated by different AI-based tools. The evaluation of usability, accessibility, and user satisfaction was conducted utilising a combination of methodologies, including the eye-tracking technique, an author survey, and automated tools designed to assess compliance with WCAG 2.1 AA guidelines.

Keywords: scientific journals portals, user interfaces, usability, accessibility, universal design, AI, eye tracking.

INTRODUCTION

In this era of digital transformation, websites play a crucial role in scientific communication, particularly in facilitating open access to knowledge. Institutional scholarly journal websites are not only repositories of publications, but also comprehensive environments that support the entire editorial process, from article submission and review to the publication and distribution of content. The usability and accessibility of these websites directly impact the effectiveness of collaboration with authors, reviewers, and readers, thereby influencing the prestige and operation of the journal [1].

Contemporary web design is increasingly based on two complementary approaches. The first is UD, which involves creating the interfaces that cater for the broadest possible range of users, regardless of their physical ability, age, or level of digital competence [2]. Secondly, there is a growing trend in the use of AI tools in the interface generation process, particularly generative template builders (e.g., Bolt.new), which have the capacity to create entire layouts from textual prompts, as opposed to adaptive ML-based systems that are focused on long-term personalisation [3, 4].

In the context of academic editorial websites, the approaches above can be used in parallel or alternatively. The popularity of AI-based tools is growing [5, 6]. However, there is a limited number of studies that systematically examine and compare the effectiveness of human-generated interfaces with those generated by AI tools. Furthermore, extant research on the usability and accessibility of websites in educational institutions has principally concentrated on the homepages of universities, academic libraries, or e-learning platforms [7]. In order to evaluate their usability and accessibility, different approaches were applied, such as fuzzy processes [8], ranking methods [9, 10], or eye tracking [11]. Concurrently, academic journals, in their capacity as standalone sites, despite their pivotal function in knowledge dissemination, remain a relatively under-researched domain in the context of UX (user experience) and WCAG (Web Content Accessibility Guidelines) compliance [12–16].

In the contemporary context of user interface design, two key aspects are of paramount importance: usability and accessibility. The term 'usability' is employed to denote the ease with which a system is able to be utilised by the end user, the intuitiveness of the system, its efficiency, and the level of enjoyment experienced by the user. The five usability characteristics identified by Nielsen: ease of learning, efficiency, memorability, low error rate, and user satisfaction, are directly reflected in the elements of the SUS (System Usability Scale) questionnaire, which allows for their practical and quantitative assessment in this study. Conversely, the concept of accessibility entails the design of environments, systems, and devices for individuals with diverse physical, sensory, and cognitive limitations. The objective is to ensure that individuals with diverse abilities can access and utilise these environments, systems, and devices without impediments. Of particular pertinence in this context is WCAG, which establishes standards for websites with regard to digital accessibility. Conversely, the utilisation of AI is proposed by these standards and concomitantly facilitates the process of creating pages in accordance with the guidelines.

This article sought to bridge this research gap by performing two independent experimental studies that concentrated on the same site, the Journal of Computer Sciences Institute (JCSI) journal service. The research study juxtaposed two approaches to designing scholarly journal interfaces: manual design with UD and automatic interface generation using AI tools, also using UD. Both approaches were implemented in accordance with the UD principles. Both case studies constitute a complementary contribution to the area of usability of academic editorial websites.

The first case study (case study I) compared the current JCSI website with a prototype designed manually in accordance with universal design principles. The alterations encompassed the enhancement of visual contrast, the restructuring of navigational elements, the implementation of semantic markup, and the improvement of interface component readability. In the second study (case study II), the current site was compared with two versions generated by different AI tools (Bolt.new and Lovable.dev), which were given the same guidelines primarily related to improving the usability and accessibility of the site.

The two studies employed a sophisticated research methodology, integrating objective methods to measure user behaviour (e.g., eye-tracking techniques), subjective user evaluations (e.g., questionnaires, including System Usability Scale [17] and analytic hierarchy process (AHP) [18]), and automatic accessibility validation in accordance with WCAG 2.1 international standards. This methodological approach facilitated the assessment of two key areas: firstly, the behaviour of users during specific tasks, and secondly, their feelings towards the interfaces under study, as well as their level of technical accessibility.

The objective of the present article was to provide a comparative analysis of the effectiveness and limitations of the two design approaches in the context of scientific interfaces, from the perspectives of usability and technology. The juxtaposition of the research outcomes enables the identification of not only the strengths of each approach but also its limitations and potential scenarios for coexistence.

Despite the growing interest in AI-based solutions, there is a paucity of comprehensive studies comparing the effectiveness of manually designed websites with those generated by AI tools, especially in the context of scientific websites. Currently, there is a notable lack of scientific studies that evaluate AI-generated websites using a combination of objective measurement techniques, such as eye tracking, and subjective user assessments. The focus of research in this field has been predominantly on the studies utilising AI tools in the context of accessibility-oriented design and the usability of websites, rather than on studies focusing on webpages of scientific journals themselves. The present study aimed to bridge this research gap by exploring the potential of integrating AI with eye-tracking methodologies. The data gathered in the experiment are available at [19].

The present study aimed to undertake a comparative analysis of user interfaces for the scientific journal websites that have been designed manually and those that have been generated automatically using AI tools. The novelty of the study lies in the use of an approach combining subjective user evaluation and objective analysis methods, including eye-tracking and automatic validation for WCAG 2.1 compliance, in the context of evaluating AI-generated webpages. To the best of the present author's knowledge, similar comparisons taking into account both ergonomic and technical aspects of interfaces from the perspective of AI and UD solutions have so far been missing in the literature. In view of the considerations above, the following research questions were posited by this study:

1. The main question guiding this study was: To what extent the interfaces generated automatically by AI tools differ in terms of usability and accessibility from the designs created manually in accordance with Universal Design principles?
2. What are the strengths and weaknesses of both approaches, analysed from both a technical perspective (compliance with WCAG) and a subjective perspective (user preferences and satisfaction)?
3. The following research question is proposed: Can these approaches be used in a complementary manner?

BACKGROUND AND RELATED WORK

In the context of web design for scientific and educational institutions, the concepts of usability and accessibility are foundational to the quality of user interface design. Adequate understanding and implementation of these guidelines have been demonstrated to enhance viewer comfort and satisfaction, whilst concomitantly increasing efficiency in tasks such as searching for articles, downloading templates, and submitting articles.

Usability – definitions and evaluation methods

One of the most widely accepted definitions of usability was proposed by Jakob Nielsen [20], who identified five key characteristics of a usable system: ease of learning, efficiency, memorability, low error rate, and user satisfaction. In practice, the evaluation of these characteristics can be carried out using various methods. These include quantitative methods, such as the measurement of task completion time, and qualitative methods, such as the administration of questionnaires and the conduction of observations.

A number of survey tools have been developed for the evaluation of graphical user interfaces (GUIs), which rely on users' subjective opinions. Questionnaires are widely used in usability, ergonomics, and UX research. The objective of this research was to capture the feelings, preferences, and satisfaction levels of people using a given system. The research combined two approaches: eye tracking, which is an objective measurement, and heuristic analysis, which represents a subjective measure dependent on user satisfaction. The most frequently employed instruments in this field are SUS and AHP methods.

SUS [21, 22] is a questionnaire consisting of 10 items that are designed to assess the overall usability of a system. It is one of the most popular tools in the field, and its effectiveness has been repeatedly confirmed in comparative studies. The popularity of the SUS is attributable to three key factors: its simplicity, its low cost of implementation, and its ability to analyse results expeditiously. It is also important to note the existence of extensions to this tool that facilitate a more precise analysis of individual UX components, such as aesthetics, efficiency, or interface predictability.

Meanwhile, AHP [19] is a multi-criteria decision-making method that finds application in UX research. Users make a comparison of different interface variants against a set of defined criteria, such as intuitiveness, graphical consistency, or element labelling. It is possible to determine user preferences and to create a ranking of the evaluated solutions based on the weights assigned to each criterion and the results of the pairwise comparisons. AHP facilitates not only the identification of the interface that is rated the highest, but also the comprehension of the aspects of the interface that determine this rating.

A secondary pivotal technique in the evaluation of usability is eye tracking, which involves the real-time monitoring of a user's eye movements. This methodology facilitates the acquisition of metrics such as the time to first fixation (TTFF) and the number of fixations in designated areas of the interface (Areas of Interest, AOI) [23, 24]. This method is beneficial for comparing alternative versions of an interface, as well as for identifying the problem areas that users may struggle to locate.

Eye-tracking is widely applied in usability research on digital solutions, yet participant groups in such studies are relatively small and heterogeneous, ranging from 5 to 93 individuals [25]. Even with limited samples, combining eye-tracking with qualitative methods has been shown to enhance diagnostic value [26]. Nielsen suggested that five users can reveal up to 80% of usability issues [27], while other authors [28, 29] recommend slightly larger samples of around 10 (±2). These differing views indicate that the optimal sample size depends on system complexity, user diversity, and study goals. Research practices vary from small to medium groups—about 12–20 participants [30–32], 30 participants [33, 34], or 40 participants [35] - with only a few studies extending to larger cohorts of 60 or more [36, 37].

Digital accessibility and WCAG guidelines

The term 'accessibility' is used to denote the capacity of a website to be utilised efficiently by the widest audience, encompassing individuals with diverse disabilities, including visual, motor, cognitive, and hearing impairments. The fundamental standard in this regard is the web content accessibility guidelines – currently version 2.1.

In accordance with European Union and Polish legislation, all public institutions, including universities and scientific journals, are obligated to ensure adherence to these guidelines. WCAG 2.1 is predicated on four principles: perceivability, functionality, understandability, and robustness (POUR). These guidelines are supported by ARIA (Accessible Rich Internet Applications) attributes, which facilitate the comprehension of the functions and structure of the interface for screen reader users. ARIA elements are HTML attributes that provide screen readers with important information about the structure and function of web content. A manual screen reader audit tests whether these ARIA attributes are implemented correctly and actually improve accessibility, so the audit results directly reflect the effectiveness of ARIA usage [38].

The WCAG compliance of websites is analysed using automatic accessibility assessment tools such as Web Accessibility Evaluation Tool (WAVE) [39], axe DevTools, TAW, or AChecker [40]. Despite the demonstrable efficacy of these methods in identifying fundamental errors, numerous studies advocate their integration with expert evaluation and manual testing.

Universal design

Universal design (UD) can be defined as an approach to the creation of products, services, and environments that are usable and accessible to all people, to the greatest extent possible, without the need for adaptation [41]. In the context of websites, this means, for example, the use of contrasting colours, unambiguous labels, responsive layout, or clear typography. Research has demonstrated that the integration of UD results in enhanced accessibility and usability metrics.

The Principles of Universal Design constitute a comprehensive framework intended to guide the development of environments, products, and communication systems that are accessible and usable by all individuals, to the greatest extent possible, without the need for adaptation or specialised solutions [42, 43]. These principles encompass equitable use, which ensures that designs are functional and marketable to individuals with diverse abilities while avoiding any form of stigmatisation. Flexibility in use accommodates a broad spectrum of individual preferences and abilities, whereas simple and intuitive use guarantees that the design is easily comprehensible, irrespective of the user's experience, knowledge, language proficiency, or cognitive capacity. The principle of perceptible information ensures that essential information is communicated effectively, regardless of the user's sensory abilities or environmental context. Tolerance for error aims to minimise hazards and the adverse consequences of accidental or unintended actions, while low physical effort ensures that a design can be used comfortably, efficiently, and with minimal fatigue. Finally, the principle of size and space for approach and use addresses the need for accessible physical dimensions that accommodate individuals of varying body sizes, postures, and mobility levels. Collectively, these principles foster inclusivity, usability, and functional equity in the design process, benefiting a diverse range of users.

Design using artificial intelligence

Recent years have seen an increased utilisation of AI tools within the domain of interface design [12, 44, 45]. For instance, platforms such as Bolt.new or Lovable.dev facilitate the automated generation of website templates in their entirety, based on text descriptions or imported data. Whilst these tools have been demonstrated to reduce design time and improve WCAG compliance through predefined components, concerns have been raised about subjective perception and the suitability of these tools for specific user groups.

It has been demonstrated by several studies that there is a discrepancy between objective measurements (e.g. eye tracking) and user preferences. Despite the greater effectiveness of AI-generated interfaces, users are more likely to prefer manually designed solutions (Case Study II). Conversely, the merits of AI are particularly evident at the technical layer, as evidenced by its capacity to ensure WCAG compliance and enhance operational efficiency.

RESEARCH METHODOLOGY

General approach

The objective of this study was to make a comparison between different approaches to the design of user interfaces of scientific journal websites, with a focus on aspects of usability, accessibility, and subjective user satisfaction. To this end, two independent comparison experiments were conducted – each based on the same Journal of Computer Sciences Institute editorial website, but using different design approaches to create a comparison version.

Table 1 presents a summary of the two case studies used in this research. In case study I (CS1), the current version of the site was evaluated against a manually designed prototype, constructed under the UD principles. In Case Study II (CS2), a current website was compared with two alternative versions. These versions were generated automatically using artificial intelligence tools, namely Bolt.new and Lovable.dev, based on a prompt designed to enhance usability and accessibility. The exact content has been provided in the supplementary materials.

The study group consisted of 76 computer science students, with males representing 80% of the cohort. The limitations arising from the composition of the study groups, namely their high homogeneity and uneven distribution of participants, are addressed in the Limitations of the study section. In both cases, a consistent and multifaceted evaluation methodology was employed, incorporating the following elements:

- The objective of this study was to analyse the effectiveness and intuitiveness of the interface in the context of performing specific tasks. This was achieved through an objective study using eye tracking.
- The present study employs a subjective approach, utilising the SUS and AHP to evaluate user impressions and preferences.
- Automated technical analysis of accessibility was employed, utilising the tools that are compliant with the WCAG 2.1 standard.
- The following evaluation criteria were adopted for the interface comparisons:
- The efficiency of a given process is determined by the time it takes to complete the task, as well as the ease with which the elements required are located on the page.
- Intuitiveness is characterised by the number of fixations, the quality of scanning paths, and how the interface is explored visually.
- The present study aimed to assess the compliance of WCAG 2.1, with a particular focus on the number of errors detected, contrast, and semantic structure.
- The level of user satisfaction and usability is measured using the SUS questionnaire, while the AHP ranking determines the preference.

Table 1. Summary of both case studies

Comparison element	Case study I (CS1) – Manual design (UD)	Case study II (CS2) – AI-Generated design
Type of interface	Manually designed prototype based on Universal Design principles	Automatically generated interfaces using AI tools
Reference version	Current JCSI website	Current JCSI website
Compared variants	One prototype version	Two AI-generated versions
Study group	18 computer science students	58 computer science students
Evaluation methods	Eye tracking, SUS questionnaire, WCAG analysis (WAVE, axe)	Eye tracking, AHP survey, WCAG analysis (WAVE, TAW, AChecker)
Scope of comparison	Efficiency, intuitiveness, satisfaction, and accessibility	Efficiency, intuitiveness, preferences, and accessibility
AI tools used	–	Bolt.new, Lovable.dev
Key difference	Complete control over layout and structure by a human designer	Interfaces generated from text prompts using AI tools

The employment of data triangulation methods (a combination of quantitative and qualitative data) permitted the assessment of the effectiveness of the interfaces, in addition to the comprehension of users' motivations and sentiments towards diverse design approaches.

Both case studies were preceded by pilot studies aimed at evaluating the difficulty level of each task. Subsequently, expert assessments were conducted to ensure comparability of task difficulty, thereby facilitating a reliable analysis of the results.

Tools and techniques

In both experiments, a professional Gazepoint GP3 HD eye tracker, operating at a sampling rate of 150 Hz and an accuracy of 0.5–1°, was utilised to analyse users' visual behaviour. The calibration of the equipment was conducted in accordance with the standard 9-point procedure, and the experimental procedure itself was carried out in a specially designated laboratory at Lublin University of Technology. The laboratory provides optimal conditions for conducting a controlled experiment, including appropriately selected artificial lighting and acoustic insulation from external noise.

The data were collected and processed using iMotions 9.1 software [46], which facilitated the experimental setup, data acquisition, and visualisation of results (heat maps, scan paths, TTFF plots). In each session, the subjects performed a series of tasks on two versions of the same page. These tasks included searching for articles, locating buttons, downloading files, and identifying contact information. The randomisation of the order in which the versions were presented, coupled with the interleaving of their presentation, was implemented to minimise the learning effect.

CS1 employed the SUS, a 10-question scale with a 5-point Likert scale, to ascertain the overall perception of interface usability. In CS2, the analytic hierarchy process method was utilised to compare users' subjective preferences in relation to several evaluation criteria, including aesthetics, intuitiveness, consistency, and marking of elements.

In both cases, evaluations were made immediately after the tasks were completed, and participants were able to view the interfaces simultaneously. This eliminated the need for memory-based comparison of views and increased the reliability of responses.

A set of four widely utilised tools was employed to evaluate adherence to WCAG 2.1 standards:

- WAVE [40] is a software application designed to detect structural errors, contrast errors, and ARIA (Accessible Rich Internet Applications) [47] semantic deficiencies.
- Axe DevTools [47] are to be utilised in conjunction with an error classification system based on the level of significance. The classification system comprises the following categories: critical, serious, moderate, and minor.
- TAW is used for verifying semantic layers and WCAG-compliant interface components.
- AChecker [48] is a verification tool that is based on a selected profile (e.g. WCAG 2.1 AA). It produces a full error report.

Cohen's effect sizes (d) were employed to verify practical differences between particular pages in both case studies. This method had been previously applied in measuring differences between trials in eye-tracking studies [49]. Effect sizes were classified as small, moderate, and large using thresholds of 0.2, 0.5, and 0.8, respectively [50].

The analysis was conducted for the same five sub-pages of the interface in each version, thus facilitating a fair comparison of the number of errors and semantic compliance. The result of this methodology was a substantial body of research material, which facilitated both a quantitative comparison of the page versions and a qualitative interpretation of the results from an end-user perspective. The ensuing chapters will present the detailed results and analysis of both cases. Additional sample results from the automatic WCAG analysis and the prompt used to generate the AI pages have been included in the supplementary materials.

CASE STUDY I: UNIVERSAL DESIGN VS. TRADITIONAL WEBSITE

Study setup

The initial case study (CS1) sought to evaluate the impact of UD principles on the user interface quality of a scientific journal website. The Journal of Computer Sciences Institute website was utilised as a working exemplar, which was then juxtaposed with a newly designed prototype. This prototype was handcrafted by a designer in accordance with the UD and Web Content Accessibility Guidelines 2.1 guidelines.

An identical set of five key sub-pages was present in both versions of the site, including:

- the present issue of the periodical contains a compendium of articles;
- the issue archive is the repository for all documentation relating to the issue;
- the present text concerns an article search engine;
- the following communication is directed towards the editorial team;
- a designated section has been allocated for authors, in which templates and guidelines have been made available for their convenience.

The initial version of the prototype was developed utilising the Bootstrap 5.3.3 framework. The principal alterations made to the prototype were as follows:

- it is recommended that the contrast between text and background be increased, in accordance with WCAG 2.1 AA guidelines;
- the unification of the header hierarchy was achieved through the introduction of larger fonts and increased inter-element spacing. These modifications were made to enhance the quality of typography;
- the addition of ARIA attributes for screen reader usage;
- the redesign of the navigation system involved the introduction of additional access pathways, with concomitant improvements in menu readability.

A total of 18 students enrolled in the Computer Science programme (16 male, 2 female) between the ages of 20 and 23 participated in the study. The participants had no prior experience with the JCSI web service, thus avoiding the influence of interface familiarity effects. The study was conducted within a laboratory setting, employing a Gazepoint GP3 HD eye tracker. The data were collected based on the version of the JSCI webpage from November 2024.

Results

One of the key performance indicators of the interface under investigation was the time to first fixation, i.e. the time that elapses from the initiation of the task until the user looks at the correct element of the page. As demonstrated in Figure 1, for the majority of tasks (particularly tasks 1, 2, and 6), TTFF was considerably reduced for the version designed in accordance with the UD principles.

The boxes in the box plots represent the distribution of values among the participants. The prototype version facilitated the expeditious identification of crucial information by users.

A further metric that was the focus of the investigation was the number of fixations in the target areas. A higher number may be indicative of greater engagement, but also of potential

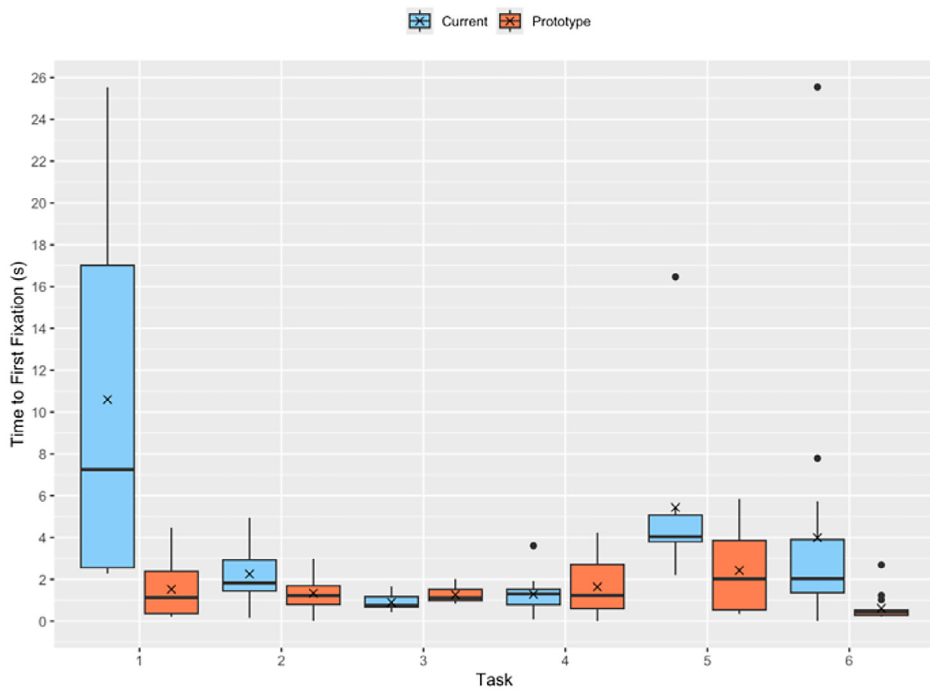

Figure 1. Time to first fixation in areas of interest for the two interface versions (current and prototype) across all tasks (CS1)

difficulty in interpreting the item. As demonstrated in Figure 2, for specific tasks (e.g. tasks 3 and 5), the current version of the page necessitated a higher number of fixations, which may suggest a less intuitive layout.

In the context of the tasks that demand rapid identification of interface functions, the users of the prototype version exhibited a reduced need for visual corrections. A thorough analysis of the heat maps yielded even more definitive conclusions, which can be seen in Figure 3. For instance, in task 6, which involved locating a button to download an article template, the users of the current version of the website exhibited a high degree of disorientation. The areas of fixation were dispersed, and attention was distracted by the elements unrelated to the purpose of the task.

The dispersed distribution of users' gaze indicates a lack of awareness regarding the appropriate location for particular UI elements. The

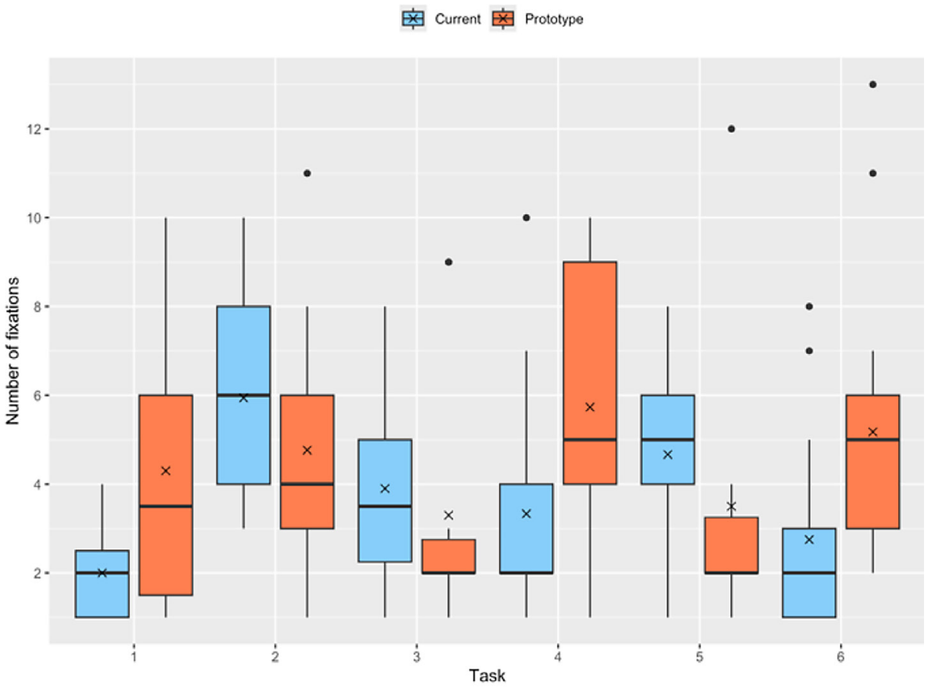

Figure 2. Number of fixations in areas of interest for the two interface versions (current and prototype) across all tasks (CS1)

Figure 3. User gaze heat map for the current website during Task 6: locating the element to download the article template

location of the ‘submit article’ button was not optimised for user comprehension.

In contrast, in the prototype version, the download button was clearly labelled and placed in a logical location, resulting in focused and short fixations (Figure 4).

The areas of focus are indicative of intuitive placement of elements and enhanced navigation. As is demonstrated by the gaze scan paths illustrated in Figures 5 and 6, analogous conclusions can be deduced. In the current version of the interface, the researcher made as many as 52 fixations before locating the correct element. In the UD version, only 18. With regard to the number of fixations in selected areas of interest, presented in Figure 2, it should be noted that although the number of fixations may indicate both difficulty and engagement, in this case, the eye-tracking patterns clearly suggest searching, rather than in-depth exploration, which supports the interpretation in terms of the lower intuitiveness of the current interface. Similar observations apply to tasks 2 and 5, where longer scanning paths with a greater number of fixations can be observed, resulting from difficulties in finding the desired element in the current interface. The extensive and disorderly visual trajectory signifies the illegibility of the page layout.

Figure 4. Heat map for the prototype website during Task 6: locating the element to download the article template

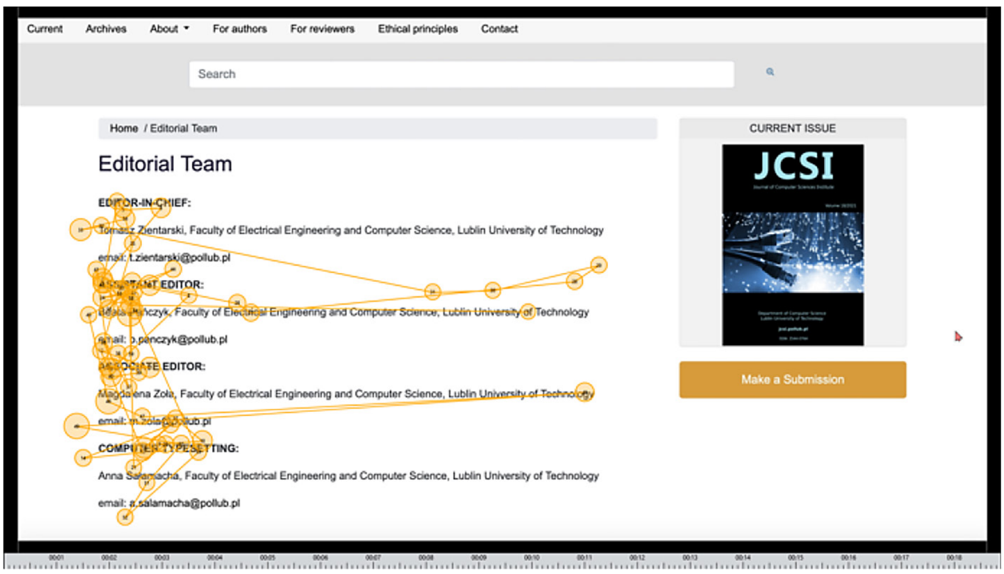

Figure 5. Scan path for the current version during Task 5

The orderly and concise nature of the shorter path indicates a superior arrangement of the information. Following the eye-tracking session, the participants were invited to evaluate both versions of the interface using the System Usability Scale questionnaire (Figure 7). The results were unequivocal:

- the mean average score for the current version was 63 points, which falls below the acceptable threshold of 68 points;
- the mean value for the UD prototype is as follows: The 85-point scale indicates a high level of usability.

The UD version received a higher rating from all respondents. The lowest individual score of the prototype (72 points) exceeded the average score of the current version.

A further element of the analysis was an automated assessment of digital accessibility, performed using WAVE and axe DevTools, which can be seen in Figure 8. The same five sub-pages were analysed in both versions.

In the UD version, no errors or contrast-related issues were detected, whereas the current version of the site yielded an average of 3.2 errors and up to 10.4 contrast errors per subpage.

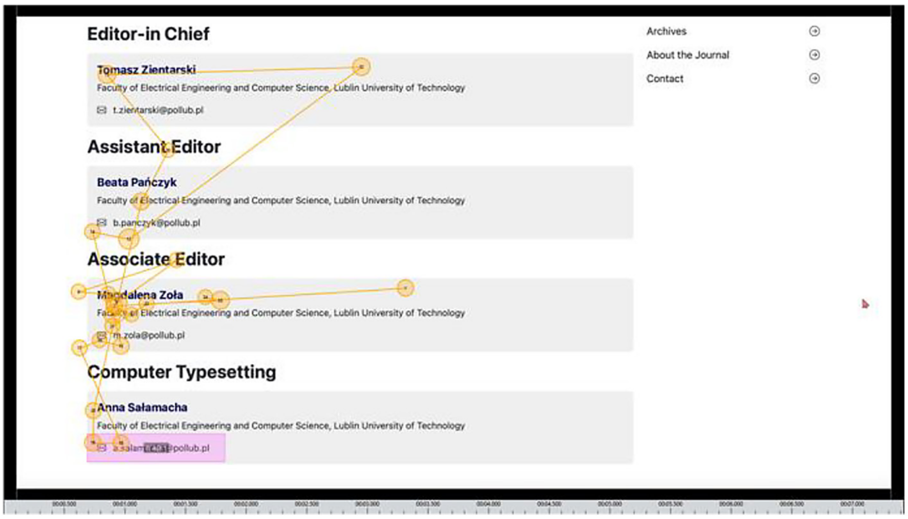

Figure 6. Scan path for the UD version during Task 5

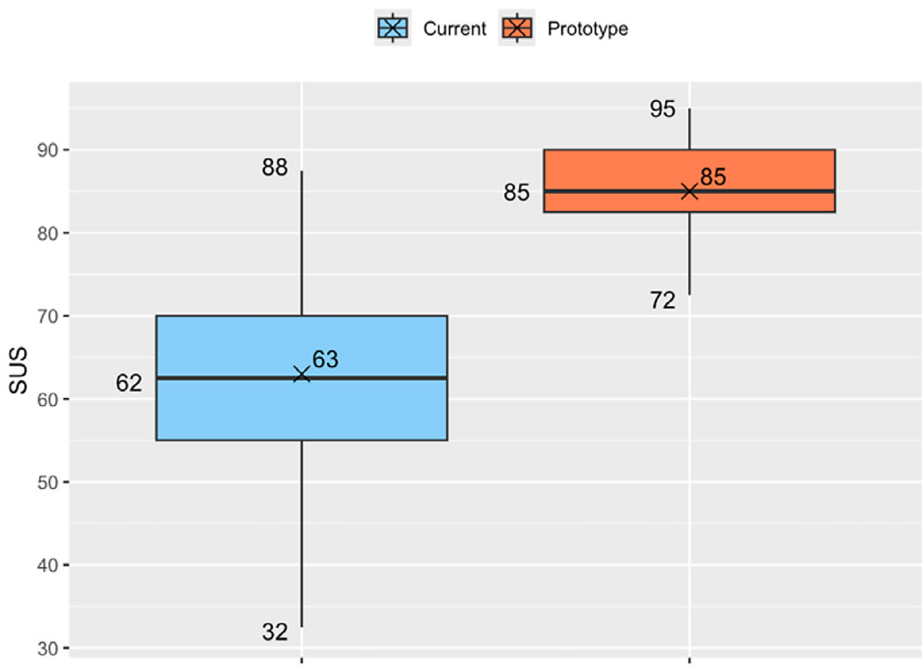

Figure 7. Comparison of mean SUS scores for the two website versions: current and prototype (CS1)

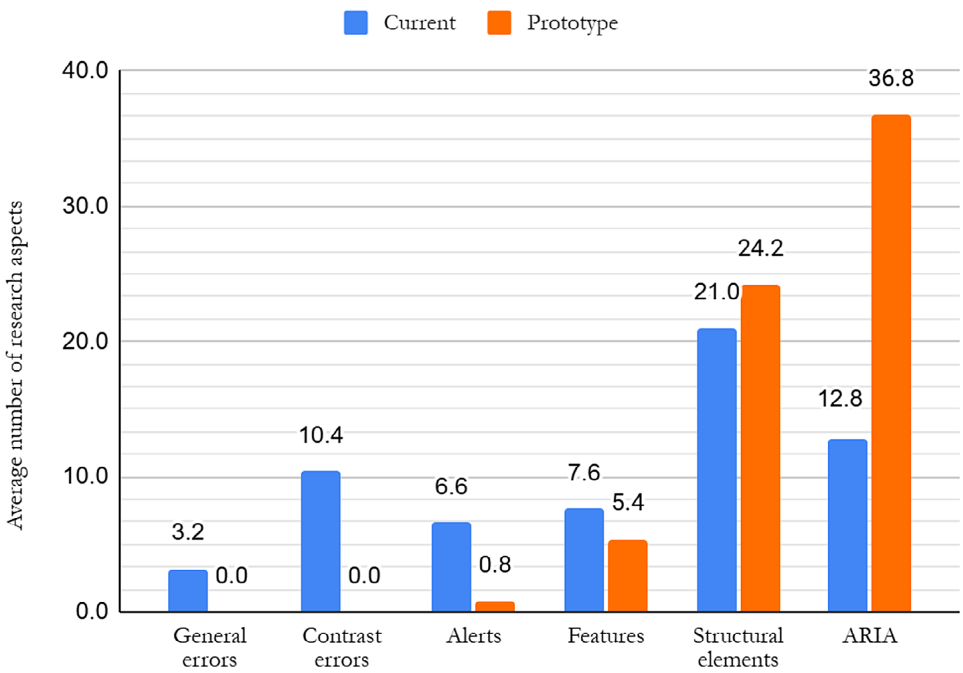

Figure 8. Mean WAVE tool results for the two website versions: current and prototype (CS1)

Furthermore, accessibility alerts occur considerably more frequently in the current version. In the Features category, the present interface incorporates a slightly higher number of elements (7.6) compared to the prototype (5.6). In comparison, in the category of Structural elements, the prototype demonstrates a greater number of elements (24.2). With regard to ARIA elements, which correspond to the results of manual screen-reader audit, the prototype contains an average of 24 elements, in contrast to 12.8 identified in the current version of the site.

The prototype demonstrated superior performance in all evaluation metrics, including errors, contrast, structure, and features.

Furthermore, the axe DevTools analysis revealed an average of 29.6 (± 6.2) accessibility errors per subpage in the current version. It is noteworthy that the prototype version is error-free.

The substantial numerical discrepancy substantiates the adherence of the prototype to WCAG 2.1 guidelines.

Table 2 presents Cohen's effect sizes (d), showing the practical differences between the current and prototype versions, calculated from the averaged results of two metrics: TTFF and NGR. A large effect was found for the TTFF metric.

CS1 analysis showed a clear advantage of the hand-designed interface over the current version in each of the categories analysed:

- higher efficiency in task completion (TTFF, number of fixations);
- greater intuitiveness and predictability of the layout (heatmaps, scan paths);
- significantly higher level of user satisfaction (SUS);
- full compliance with the WCAG 2.1 accessibility requirements.

CASE STUDY II: AI-GENERATED INTERFACES VS. TRADITIONAL WEBSITE

Study setup

The second case study (CS2) aimed to assess the usability and accessibility of the user interfaces generated automatically by artificial intelligence tools. The present study analysed two versions of the Journal of Computer Science Institute

Table 2. Pairwise effect sizes (Cohen's d) for eye-tracking features between interfaces

Feature	Interface comparison	SDpooled	d	Effect size
TTFF	Current vs prototype	2370.34	1.10	large
NGR	Current vs prototype	4.27	-0.06	small

website. One version was generated by the Bolt. new tool and the other by Lovable.dev. In both cases, the reference point was the current version of the JCSI site. The following guidelines were provided to each tool:
- it is imperative that layout and content be preserved;
- this paper sets out the methodology for improving accessibility in accordance with WCAG 2.1;
- the enhancement of visual clarity and the facilitation of intuitive navigation are of paramount importance.

The study involved 58 students of computer science (13 female, 45 male) between the ages of 20 and 23, approximately half of whom had previously been exposed to the current site. The experiment was conducted under controlled conditions using a Gazepoint GP3 HD eye tracker and iMotions 9.x software. Subsequent to the completion of the research tasks, a subjective evaluation was conducted employing the AHP method, and each version of the website was automatically analysed for WCAG compliance using three independent tools: WAVE, TAW, and AChecker. The data were collected based on the version of the JSCI webpage from November 2024.

Results

The eye-tracking analysis of individual tasks (Figure 9) reveals varying results in terms of TTFF for three website versions (JCSI, Bolt, Lovable). For Tasks 1 and 3, the JCSI website demonstrated the highest performance, while Lovable achieved the best results for Tasks 2 and 5. The average TTFF for the JCSI website (current) across all tasks was the shortest among the three solutions, at 2.06 seconds. Among the AI-generated services, Lovable performed better, with an average TTFF of 2.11 seconds, compared to Bolt, which averaged 2.35 seconds. The second metric related to interface performance considered in the study was the number of gaze revisits (NGR). The average number of revisits for all subpages of the three analysed web services for each task is presented in Figure 10. For tasks 4 and 5, the lowest average number of revisits was observed for the Lovable service. In contrast, for tasks 1, 2, and 3, the average values were at a similar level across all three evaluated web services. Considering the averages across all tasks, the Lovable service exhibited the lowest number of revisits (2.01), while the values for JCSI and Bolt were 2.29 and 2.27, respectively.

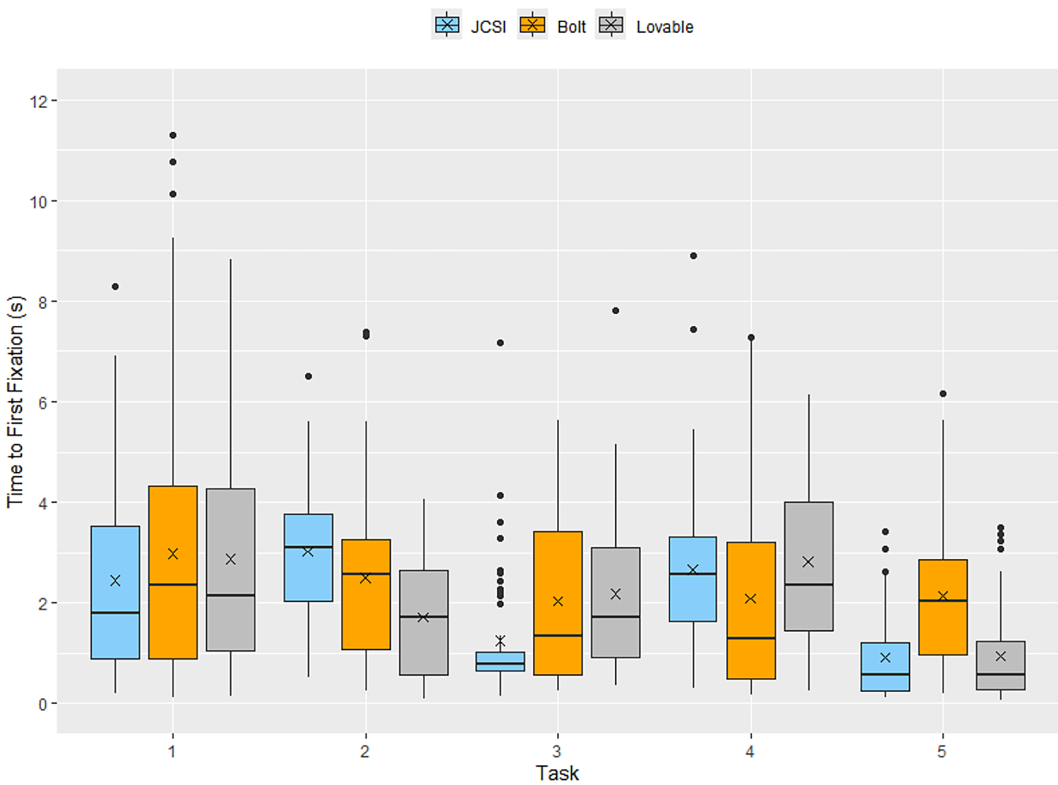

Figure 9. Mean time to first fixation in areas of interest for three website versions (JCSI, Bolt, Lovable) across all tasks (CS2)

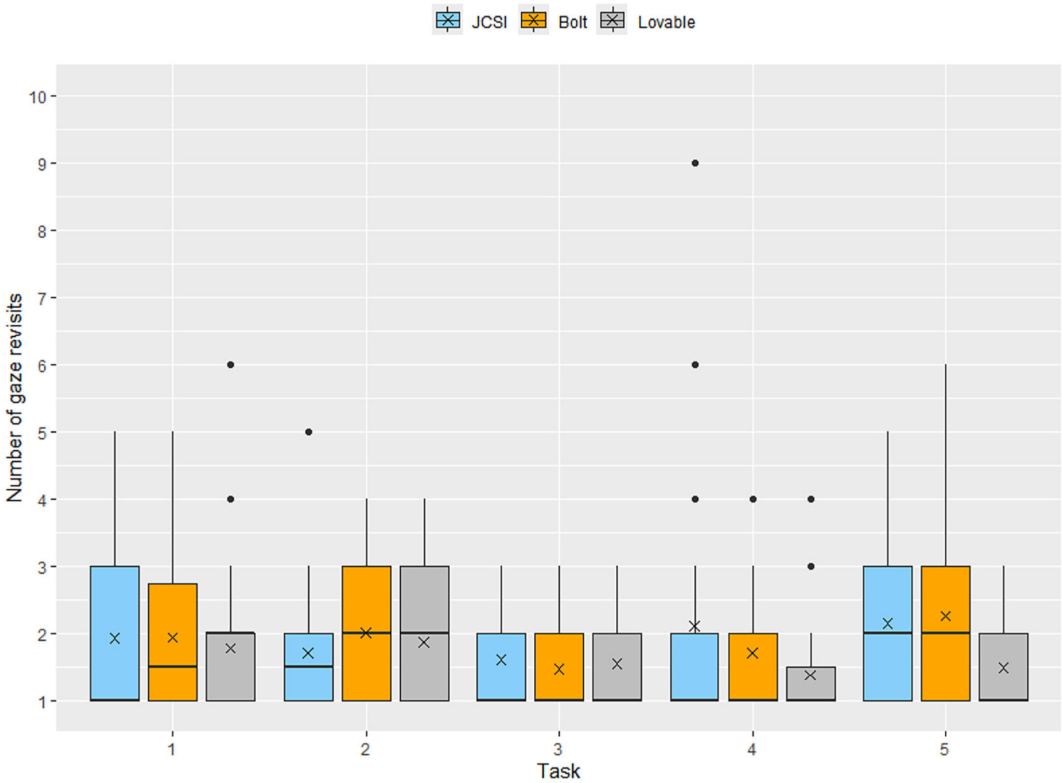

Figure 10. Mean number of gaze revisits to areas of interest for three interface versions (JCSI, Bolt, Lovable) across all tasks (CS2)

The Lovable.dev version consistently achieved the shortest time to first fixation and the fewest gaze revisits to the AOIs, indicating that its layout is highly intuitive. The Bolt.new version exhibited the longest TTFF values and the highest NGR, a phenomenon that may be attributed to the visual complexity of its webpages and indicates potential issues in the efficient retrieval of crucial information.

Figure 11 shows the mean duration of saccades for the three website versions across all tasks. The longest saccade durations for the current JCSI interfaces were observed during tasks 1 and 5. Considering the average saccade durations across all tasks, Bolt had an average of 24.42 ms, Lovable 25.89 ms, and JCSI 26.51 ms. The longer saccades in the current JCSI website may indicate greater cognitive effort required to locate the necessary information.

Table 3 presents Cohen's effect sizes (d), showing the practical differences between JCSI and Bolt interfaces, and between Lovable and JCSI interfaces, calculated from the averaged results of two metrics: TTFF and NGR. A small effect size was observed for the TTFF metric for the Lovable and JCSI interfaces, as well as for the JCSI and Bolt interfaces. Moreover, a medium effect was found for the TTFF metric for both JCSI and Bolt interfaces. A large effect size was observed for the NGR feature in the case of the Lovable and JCSI interfaces.

After the task section, respondents were invited to provide a rating for each version of the website in four categories:

- the intuitive nature of the interface is a key consideration;
- the visual aesthetics of the subject under discussion are of particular relevance in this context;
- the uniformity of graphic elements is of paramount importance;
- the legibility and labelling of components is of paramount importance.
- The evaluation was conducted using the AHP method, as shown in Figure 12.

Among AI-generated interfaces, the Lovable.dev version demonstrated superior performance in three categories: intuitiveness, aesthetics, and consistency. An assessment was conducted to determine the degree to which each version complied with WCAG 2.1 standards. This assessment utilised the WAVE, TAW, and AChecker tools as the primary instruments for evaluation. The AI

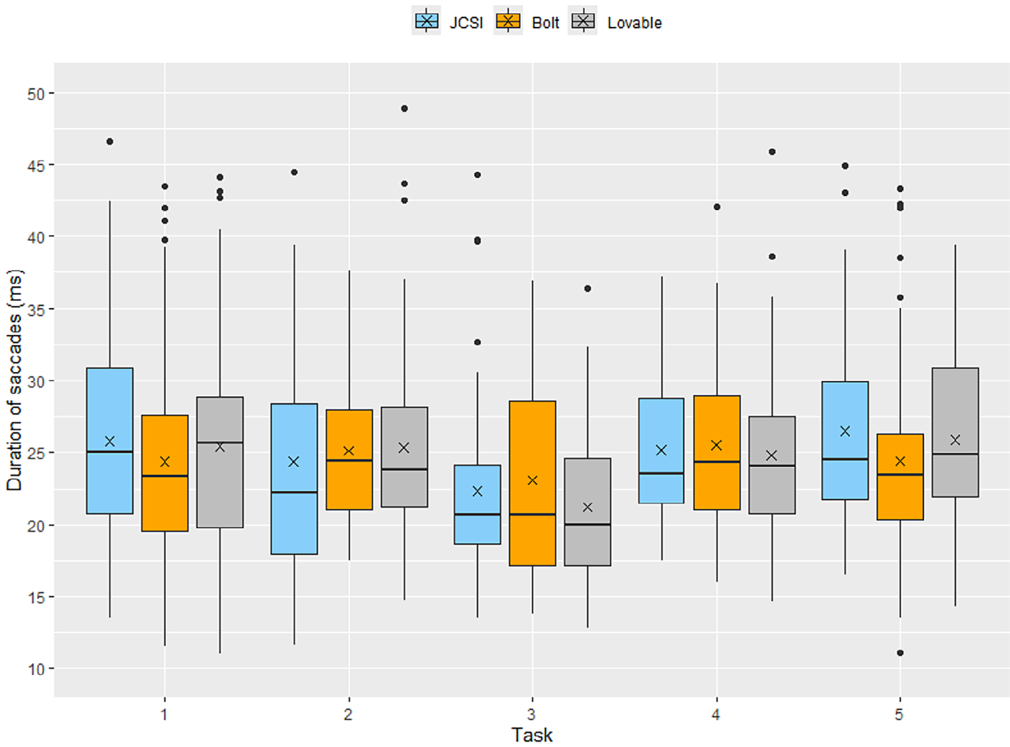

Figure 11. Mean duration of saccades during interface exploration for the three interface versions (JCSI, Bolt, Lovable) across all tasks (CS2)

Table 3. Pairwise effect sizes (Cohen's d) for eye-tracking features between interfaces

Feature	Interface comparison		SDpooled	d	Effect size
TTFF	JCSI	Bolt	635.75	-0.45	medium
	Lovable	JCSI	775.10	0.06	small
NGR	JCSI	Bolt	0.24	0.12	small
	Lovable	JCSI	0.20	-1.48	large

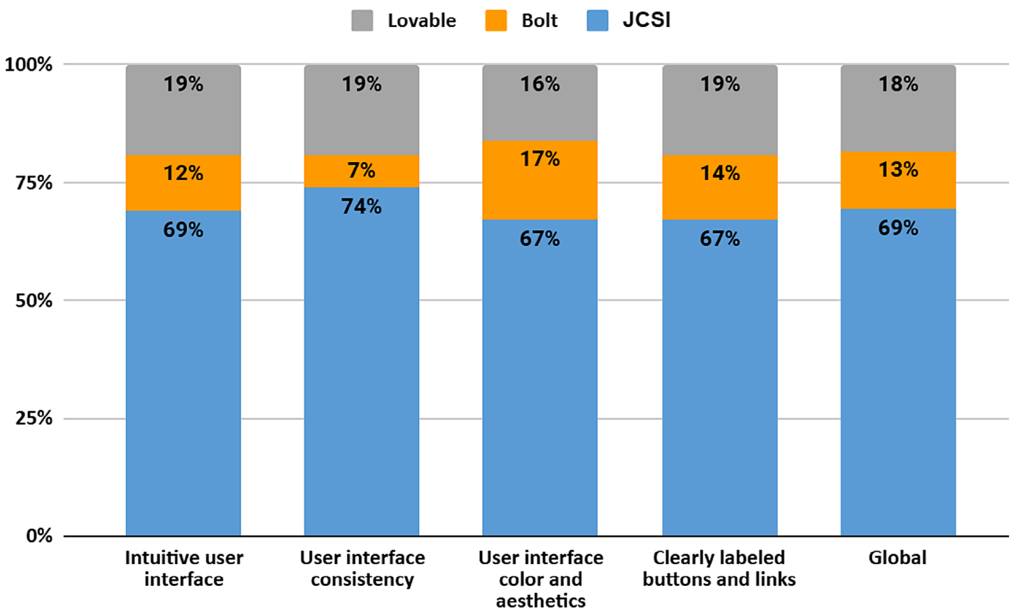

Figure 12. Standardised AHP preferences for four interface evaluation criteria

versions demonstrated a marked performance improvement when compared to the current site, with Lovable.dev achieving the most optimal results in all three tools. The Lovable.dev version was found to generate the lowest number of errors across all tools, particularly in the contrast and document structure categories. The following errors were identified in the current version:

- the presence of low colour contrast is indicated;
- the absence of labelling on buttons and links is a notable issue;
- it is evident that the header structure is not accurate.

Conversely, the AI versions encountered minor issues with (Table 4):

- the process of generating copies of identifiers (IDs) is referred to as 'duplication of identifiers';
- it is evident that the provided description of the links and landmarks is inaccurate.

In addition, ARIA elements were analysed to validate the results of the manual screen-reader audit. It was found that JCSI contained an average of 14.4 elements, in contrast to Bolt and Lovable, which had 8.6 and 5.2 identified elements, respectively. CS2 demonstrated that when utilised effectively, AI tools can facilitate the generation of interfaces:

- it has been determined that the aforementioned option is more intuitive (Lovable.dev);
- the visual aspect has been rendered cleaner and consistent;
- it has been demonstrated that the subject of this study is significantly more WCAG 2.1 compliant than manually created versions that do not have the support of accessibility experts.

Concurrently, the study demonstrated that elevated technical efficiency and compliance do not invariably translate into increased user preferences, with the current interface being regarded as more satisfactory as compared to the AI-generated interfaces. This phenomenon may be attributed to several factors, including the user's familiarity with the website, its visual design, or the trust placed in the website's structure.

DISCUSSION

The experiment presented in the paper is composed of two experiments (CS1 and CS2), and a comparative analysis of the results obtained for the interfaces designed manually (in accordance with the Universal Design principles) and the interfaces generated automatically using AI tools is presented. The juxtaposition of these approaches enables the identification of strengths and weaknesses of each approach, both in terms of objective (measurement) and subjective (feeling) dimensions. Several aspects were analysed in order to ensure consistency and completeness of the research goals.

The efficiency and intuitiveness of the exploration process

In both case studies, an eye-tracking technique was utilised to measure users' visual behaviour. As indicated by the most significant metrics (time to first fixation, number of fixations, and saccade length), the interfaces were consistently found to be both intuitive and straightforward.

In CS1, the prototype designed according to UD principles achieved a shorter TTFF and fewer diffuse fixations than the current version. Heatmap analysis further confirmed that respondents were less visually disoriented and were faster at locating key elements in the prototype version. These results suggest that the changes introduced in the prototype positively impacted the intuitiveness of the interface, particularly in terms of navigation and user orientation within the task space. In CS2, the results are diversified. The JCSI obtained better results in the TTFF metric, whereas in the case of NGR and duration of saccades the AI-generated interfaces performed better. The overall findings indicate that manual creation using UD produces more reliable and effective outcomes than those generated through AI-based methods. Both UD and AI possess the capacity to generate the interfaces that exhibit high cognitive efficiency; however, this potential is only realised when the structure is meticulously designed and consistent

Usability and satisfaction assessment

The subjective feelings of users were measured using different methods. In CS1, the System Usability Scale questionnaire was utilised, whereas in CS2, the analytic hierarchy process method was utilised.

In CS1, the results of the SUS survey revealed that the prototype version of the website, designed according to universal design principles, scored 29 percent higher than the existing website. This difference indicates an improvement in the perception of the website among respondents. The UD prototype achieved an average SUS score of 85, in contrast to the current version, which gained 63 points, which is below the average SUS score of 68 from normative studies. This result suggests a moderate, or even unsatisfactory, user experience. The UD-compliant manual design proved to be a particularly effective solution, resulting in a significant enhancement of the user experience. This assertion is substantiated by the findings of the eye-tracking analysis, which revealed a reduction in the time to first fixation and the optimisation of scanning paths. This study demonstrates that the synergy between AI and eye tracking can serve as the foundation for modern UX evaluation standards [3].

In contrast, the CS2 study revealed that the AI-generated interfaces, especially the version created with Lovable.dev, exhibited high intuitiveness, as indicated by eye-tracking indicators. However, in their subjective evaluation, users demonstrated a preference for the traditional interface over AI-generated interfaces.

Additionally, results did not reveal the influence of prior knowledge of the current journal site, as respondents claiming familiarity with the original version of the interface made similar choices to those with no prior contact with the original site interface.

Compliance with the WCAG 2.1 guidelines

Accessibility analysis revealed the most significant disparities between the approaches. In the case of CS1, the prototype version demonstrated a flawless execution across all tools (WAVE, axe), underscoring the efficacy of integrating manual design principles with accessibility considerations. In the case of CS2, Lovable.dev achieved the most optimal results among the AI versions, with a total of merely 2–4 errors in the various validators. This demonstrates that contemporary AI tools possess the capability to automatically generate the interfaces that are highly WCAG-compliant. The current version demonstrated the least compliance, generating between several and dozens of contrast, structure, and label errors in both cases. Both experiments – the one based on manual design (CS1) and the one using AI tools (CS2) – provided unambiguous data indicating that alternative versions of the website (whether manually designed or automatically generated) outperformed the current interface in terms of WCAG 2.1 compliance.

Table 5 presents the comparison that has been made of WCAG 2.1 compliance between CS1 and CS2, with the average number of errors being calculated for each website. It is evident that both methods – manual design with UD and AI generation – have the potential to result in high WCAG compliance. The efficacy of the AI system is comparable to that of the human designer, provided that the appropriate input parameters are established and final validation is conducted.

General comparison

Table 6 presents a summary comparison of both case studies. General outcomes are presented in the context of such measures as eye-tracking metrics, user satisfaction, accessibility, and WCAG compliance, including ARIA elements and errors. Moreover, aspects such as aesthetic and modern appearance, deployment speed & scalability, as well as the effect size (Cohen's d), were analysed. Results are presented for the best interfaces of each case study. Eye-tracking metrics, such as TTFF, NGR, and the number of fixations, were grouped and analysed in the context of users' effort and efficiency. Both the JCSI prototype for the CS1 and AI-based Lovable.dev for the CS2 proved to be less effort-consuming and more efficient compared to other interfaces. User satisfaction and acceptance were measured using SUS and AHP methods. Results show that the JCSI prototype in CS1 and the JCSI version in CS2 reached the highest user rates.

Table 5. WCAG 2.1 compliance comparison of current, prototype, and AI-generated interfaces

Interface version	Avg. number of errors	Comment
JCSI (CS1 & CS2)	13–21	Numerous contrast issues, missing labels, poor HTML structure
Manual prototype (UD)	0	Fully WCAG-compliant, carefully designed structure
AI – Lovable.dev	2–4	Minimal issues, very high compliance
AI – Bolt.new	5–8	Acceptable structure, but less consistent accessibility

In the case of accessibility, WCAG 2.1 compliance was analysed with particular emphasis on ARIA elements and the number of errors. WCAG 2.1 compliance was found to be better for the JCSI prototype for the CS1 and AI-based Lovable.dev for the CS2. These findings are confirmed by the number of errors, where the JCSI prototype for the CS1 was error-free, whereas AI-based Lovable.dev obtained the fewest number of errors (2–4) for the CS2. However, the number of ARIA elements, corresponding to the manual screen-reader audit, was higher in the case of the prototype in CS1 and the JCSI version in CS2.

In terms of aesthetic and modern appearance, the AI-based interface Lovable.dev shows an advantage, while the JCSI prototype is not considered to have an aesthetic and modern appearance. What is more, both AI-based interfaces demonstrate better deployment speed and scalability in comparison to the JCSI prototype.

The effect size (Cohen's d) for the TTFF eye-tracking measure in comparisons of the two interfaces was generalised for Lovable.dev and Bolt.new. The JCSI prototype for CS1 demonstrated the larger practical differences in contrast to the AI-based interfaces.

The use of eye tracking and automated accessibility analysis tools enabled the acquisition of profound, objective data concerning users' interactions with the interface. The number of fixations in key areas and the quality of scan paths were found to be significant in detecting subtle differences in the perception of interface elements. In contrast, the WAVE, axe DevTools, TAW, and AChecker tools revealed a large number of errors in the current page (mainly related to contrast, incorrect semantics, and missing labels) that were absent from the prototypes. In particular, the interface created according to UD, although not automatically generated, was consciously designed in a WCAG-compliant manner, and it was free of critical errors.

It is important to note that websites, including scientific journal portals, are utilised by a diverse range of users, encompassing individuals with developmental disabilities, as well as those with intellectual, sensory or social disabilities. The manner in which information is presented, the intuitiveness of the interface, and the visual structure must be meticulously tailored to ensure optimal user comfort [51].

In this context, the use of a combination of artificial intelligence and eye-tracking technology as tools to support the design of accessible and tailored interfaces appears to be a highly promising approach. The capacity of AI to analyse the data collected from users with specific impairments, and consequently generate versions of web pages that are better suited to their perception and interaction style, is a noteworthy development. Conversely, eye-tracking facilitates the precise analysis of how individuals with special needs navigate content on a screen, thereby unveiling the challenges that may not be discernible in conventional UX assessments.

LIMITATIONS OF THE STUDY

It is important to note several limitations common to both studies. The sample of the joint research groups comprised 76 computer science students. Nonetheless, the case studies undertaken within this research varied in the size of their participant groups. The research group was characterised by a high degree of homogeneity, with the majority of participants being male computer science students. Such a lack of diversity within the sample may have restricted the heterogeneity of perspectives and, consequently, may limit the generalisability of the

Table 6. Summary comparison of CS1 (UD) and CS2 (AI)

Evaluation criterion	Better in CS1 (Prototype)	Better in CS2 (AI)
Effort (Eye-tracking efficiency)	✓	✓ (Lovable.dev only)
User Satisfaction and Acceptance	✓	✗
Accessibility (WCAG 2.1 compliance)	✓	✓ (Lovable.dev)
ARIA elements	✓	✗
Number of errors	✓	✓ (Lovable.dev)
Aesthetic and modern appearance	✗	✓ (Lovable.dev)
Deployment speed & scalability	✗	✓
Cohen's d (TTFF)	large	medium

findings to broader populations. Sociodemographic characteristics—such as sexual orientation, gender, ethnicity, urbanicity, or cultural background [52] — are inherently complex factors that were not accounted for in the current study; however, it would be beneficial to incorporate these variables in future research designs. Future studies should therefore aim to expand the case study sample by including participants from a broader range of academic disciplines, social and cultural backgrounds, as well as levels of professional or educational experience, in order to enhance both diversity and representativeness. Moreover, there is a discrepancy in the sample sizes between the two case studies. Although each case study was analysed separately, future works will cover more balanced participant groups to enable more extended statistical analysis and reach better generalisability.

In addition, although accessibility formed part of the research discourse, no empirical testing was conducted with disabled participants. Similarly, no manual validation of accessibility was carried out; instead, the study relied exclusively on automated tools for assessment. Within case study 2 (CS2), a notable discrepancy was identified between the outcomes generated by different AI-based tools. This inconsistency highlights the need to establish a standardised process for both the generation and evaluation of results, ensuring the reliability and comparability of future research outputs.

CONCLUSIONS

The present study set out to conduct an integrated analysis of two approaches to user interface design for scientific journal websites. The first approach was manual design following UD principles (CS1), and the second was automatic interface generation using AI-based tools (CS2), also with UD standards. A comparative analysis of the two approaches leads to several key conclusions. Firstly, the application of manual design in accordance with UD principles has been demonstrated to facilitate an enhanced subjective perception of the interface, with respect to UX and satisfaction. However, it should be noted that this approach requires more time, resources and higher design competence. Secondly, interfaces generated by artificial intelligence have been shown to exhibit acceptable results in terms of intuitiveness and visual structure, while demonstrating a high degree of compliance with the technical requirements of the WCAG standard. Concurrently, the approach is distinguished by its limited adaptability and less nuanced aesthetic and stylistic characteristics. It has been demonstrated that, despite the high technical quality, this does not invariably translate into elevated levels of user satisfaction.

It is concluded that both artificial intelligence and user-centred design can fulfil complementary roles in the system design process. The former can function as a tool for the rapid generation of a standards-compliant interface framework, while the latter can be utilised for the refinement of details as well as the infusion of authenticity and distinctiveness into the design of a website. The integration of AI and eye-tracking technology has the potential to yield novel insights and applications in various domains, including web interface design and the evaluation of AI model quality. The value of this approach is especially evident in the environments where efficiency and compliance with usability as well as accessibility standards are paramount.

A critical evaluation of both technologies reveals their respective advantages. Universal design confers comprehensive design authority, facilitates personalisation, fosters profound user comprehension, and ensures context sensitivity. AI tools offer rapid generation of ready-made components, automatic adherence to standards, but limit creativity and customisation. The most promising direction for further research is the combination of the two approaches: the rapid generation of interfaces by artificial intelligence, followed by the manual refinement of these interfaces by a UX/UI specialist. This combination of approaches is intended to ensure consistency, cultural compatibility, and aesthetic authenticity.

REFERENCES

1. Zhao Y., Gong L., Yang W., Zhou Y. How accessibility affects other quality attributes of software? A case study of GitHub. Sci Comput Program. 2024;231:103027. https://doi.org/10.1016/j.scico.2023.103027
2. Vu KPL, Proctor RW, Hung YH. Website design and evaluation. In: Salvendy G, editor. Handbook of Human Factors and Ergonomics. Hoboken, NJ: Wiley; 2021;1016–1036. https://doi.org/10.1002/9781119636113.ch39
3. Lu Y, Lee MKO, Pan SL, Tan CW, Huang K, Wu J, et al. Bridging the gap between UX practitioners'

work practices and AI-enabled design support tools. In: CHI Conf Hum Factors Comput Syst Ext Abstr. 2022;1–7. https://doi.org/10.1145/3491101.3519809

4. Hassani H, Silva ES. The role of ChatGPT in data science: how AI-assisted conversational interfaces are revolutionizing the field. Big Data Cogn Comput. 2023;7(2):62. https://doi.org/10.3390/bdcc7020062
5. Boy AF, Akhyar A, Arif TY, Syahrial S. Development of an artificial intelligence model based on MobileNetV3 for early detection of dental caries using smartphone images: a preliminary study. Adv Sci Technol Res J. 2025;19(4):109–116. https://doi.org/10.12913/22998624/200308
6. Boy AF, Akhyar A, Arif TY, Syahrial S. Artificial intelligence for dental caries detection: a mixup, fine-tuning, and quantization approach on the MobileNetV2 model. J Conserv Dent Endod. 2025;28(8):764–771. https://doi.org/10.4103/JCDE.JCDE_362_25
7. Chow AS, Bridges M, Commander P. The website design and usability of US academic and public libraries. Ref User Serv Q. 2014;53(3):253–265. https://doi.org/10.5860/rusq.53n3.253
8. Muhammad A, Siddique A, Naveed QN, Khaliq U, Aseere AM, Hasan MA, et al. Evaluating usability of academic websites through a fuzzy analytical hierarchical process. Sustainability. 2021;13(4):2040. https://doi.org/10.3390/su13042040
9. Weideman M. Comparative analysis of homepage website visibility and academic rankings for UK universities. Inf Res. 2014;18(4):1–14.
10. Gee LLS, Dasan J, Hasan CHC. Investigating the usability of universities' websites: upgrading visualization preference and system performance. Int J Interact Mob Technol. 2022;16(2). https://doi.org/10.3991/ijim.v16i02.23707
11. Almourad MB, Fekri M, Albahri A, et al. Usability assessment of a university academic portal using eye tracking technology. Procedia Comput Sci. 2023;220:323–330. https://doi.org/10.1016/j.procs.2023.03.042
12. Stige Å, Erichsen C, Haugstvedt H, et al. Artificial intelligence (AI) for user experience (UX) design: a systematic literature review and future research agenda. Inf Technol People. 2024;37(6):2324–2352. https://doi.org/10.1108/ITP-07-2022-0519
13. Baharum A, Halim NSA, Mansor EA, et al. Enhancing UX design through eye-tracking and image processing: practical insights and applications. In: 2024 Int Conf Platform Technol Serv (PlatCon). IEEE; 2024;1–4. https://doi.org/10.1109/PlatCon63925.2024.10830718
14. Hasan TI, Khan R, Abu Bakar S, et al. UI/UX design impact on e-commerce attracting users. Procedia Comput Sci. 2024;245:1075–1082. https://doi.org/10.1016/j.procs.2024.10.336
15. Chatziemmanouil T, Katsanos C. Accessibility Academy: interactive learning of the WCAG 2.1 Web Accessibility Guidelines. In: 2024 IEEE Global Engineering Education Conf (EDUCON). IEEE; 2024;1–7. https://doi.org/10.1109/EDUCON60312.2024.10578915
16. Kirkpatrick A, O'Connor J, Campbell A, Cooper M. Web Content Accessibility Guidelines (WCAG) 2.1. World Wide Web Consortium (W3C); 2018. https://www.w3.org/TR/WCAG21/
17. Fiandhika MF, Kelana B. System Usability Scale validation from the expert perspective. E3S Web Conf. 2024;483:03008. https://doi.org/10.1051/e3sconf/202448303008
18. Munier N, Hontoria E, Jimenez A. Uses and limitations of the AHP method. Cham: Springer Int Publ; 2021. https://doi.org/10.1007/978-3-030-60392-2
19. Skolimowski K, Rola B, Tokarski P, Rybak K, Łazaruk K, Plechawska-Wójcik M, Dzieńkowski M. Human-computer interaction in scientific journal websites: a usability and accessibility analysis of human-designed and AI-generated user interfaces. Adv Sci Technol Res J. 2025;19(4). https://doi.org/10.5281/zenodo.17064358
20. Nielsen J. How to conduct a heuristic evaluation. Fremont, CA: Nielsen Norman Group; 1995.
21. Tullis TS, Stetson JN. A comparison of questionnaires for assessing website usability. In: Proc UPA Usability Professionals' Assoc Conf; 2004 Jun 7–11; Minneapolis, MN, USA. 1–12.
22. Brooke J. SUS: A "quick and dirty" usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland AL, editors. Usability Evaluation in Industry. London: Taylor & Francis; 1996;189–194.
23. Carter BT, Luke SG. Best practices in eye tracking research. Int J Psychophysiol. 2020;155:49–62. https://doi.org/10.1016/j.ijpsycho.2020.05.010
24. Hessels RS, Holleman GA, Cornelissen THW, Hooge ITC. The area-of-interest problem in eye tracking research: a noise-robust solution for face and sparse stimuli. Behav Res Methods. 2016;48:1694–1712. https://doi.org/10.3758/s13428-015-0676-y
25. van den Berg L, Engelsma T, Peute L. Exploration of eye-tracking methodologies in usability testing of digital health technology: a rapid review. In: Digital health and informatics innovations for sustainable health care systems. 2024;1130–1134.
26. Elbabour F, Alhadreti O, Mayhew P. Eye tracking in retrospective think-aloud usability testing: is there added value? J Usability Stud. 2017;12(3):95–110.
27. Nielsen J. Usability testing with 5 users. Jakob Nielsen's Website. 2000 Mar 19 [cited 2025 Aug 20]. Available from: http://www.useit.com/alertbox/20000319.html
28. Hwang W, Salvendy G. Number of people required for usability evaluation: the 10±2 rule. Commun ACM. 2010;53(5):130–133. https://doi.org/10.1145/1735223.1735255

29. Albert B, Tullis T. Measuring the user experience: collecting, analyzing, and presenting UX metrics. 2nd ed. San Francisco: Morgan Kaufmann; 2022.
30. Wegner S, Lohmeyer Q, Wahlen D, Neumann S, Groebli JC, Meboldt M. Value of eye-tracking data for classification of information processing-intensive handling tasks: quasi-experimental study on cognition and user interface design. JMIR Hum Factors. 2020;7(2):e15581. https://doi.org/10.2196/15581
31. Paulus YT, Remijn GB. Usability of various dwell times for eye-gaze-based object selection with eye tracking. Displays. 2021;67:101997. https://doi.org/10.1016/j.displa.2021.101997
32. Ţichindelean M, Ţichindelean MT, Cetină I, Orzan G. A comparative eye tracking study of usability—towards sustainable web design. Sustainability. 2021;13(18):10415. https://doi.org/10.3390/su131810415
33. Borkin M, Bylinskii Z, Kim N, Bainbridge C, Yeh C, Borkin D, Pfister H, Oliva A. Beyond memorability: visualization recognition and recall. IEEE Trans Vis Comput Graph. 2016;22(1):519–528. https://doi.org/10.1109/TVCG.2015.2467732
34. Maudlin LC, McNeal KS, Dinon-Aldridge H, Davis C, Boyles R, Atkins RM. Website usability differences between males and females: an eye-tracking evaluation of a climate decision support system. Weather Clim Soc. 2020;12(1):183–192. https://doi.org/10.1175/WCAS-D-18-0127.1
35. Leonidou P, Constantinides A, Belk M, Fidas C, Pitsillides A. Eye gaze and interaction differences of holistic versus analytic users in image-recognition human interaction proof schemes. In: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Vol. 12788 LNCS. 2021;66–75. https://doi.org/10.1007/978-3-030-77392-2_5
36. Shi Y, Zhang Z, Huang K, Ma W, Tu S. Human–computer interaction based on face feature localization. J Vis Commun Image Represent. 2020;70:102740. https://doi.org/10.1016/j.jvcir.2019.102740
37. Borkin M, Vo A, Bylinskii Z, Isola P, Sunkavalli S, Oliva A, Pfister H, et al. What makes a visualization memorable? IEEE Trans Vis Comput Graph. 2013;19(12):2306–2315. https://doi.org/10.1109/TVCG.2013.234
38. Yu Y, Ryskeldiev B, Tsutsui A, Gillingham M, Wang Y. From cluttered to clear: improving the web accessibility design for screen reader users in e-commerce with generative AI. arXiv e-prints. 2025;arXiv-2502.
39. Martín A, Cechich A, Rossi G. Comparing approaches to web accessibility assessment. In: Handbook of Research on Web Information Systems Quality. Hershey, PA: IGI Global; 2008;181–205.
40. Ahmi A, Mohamad R. Evaluating accessibility of Malaysian public universities websites using AChecker and WAVE. Journal of Information and Communication Technology (JICT) 2016; 15(2):193–214. https://doi.org/10.32890/jict.15.2.2016.6703
41. Kurek K., Skublewska-Paszkowska M., Dzieńkowski M., Powroźnik P. The impact of applying universal design principles on the usability of online accommodation booking websites. Appl Comput Sci. 2024;20(1):56–71. http://dx.doi.org/10.35784/acs-2024-04
42. Dolph E. The developing definition of universal design. J Access Des All. 2021;11(2):178–194. https://doi.org/10.17411/jacces.v11i2.263
43. Law CM, Yi JS, Choi YS, Jacko JA. A systematic examination of universal design resources: part 1, heuristic evaluation. Univ Access Inf Soc. 2008;7:31–54. https://doi.org/10.1007/s10209-007-0100-1
44. He Z, Guo W, Zhang P. Performance prediction, optimal design and operational control of thermal energy storage using artificial intelligence methods. Renew Sustain Energy Rev. 2022;156:111977. https://doi.org/10.1016/j.rser.2021.111977
45. Mustafa B. The impact of artificial intelligence on the graphic design industry. Res Militaris. 2023;13(3):243–255. https://doi.org/10.7176/ADS/104-01
46. iMotions. iMotions 9.1 – Human Behavior Research Software [Internet]. 2025 [cited 2025 Jun 12]. Available from: https://imotions.com
47. Upadhyaya N. Accessibility and internationalization. In: Advanced Front-End Development: Building Scalable and High-Performance Web Applications with React. Berkeley, CA: Apress; 2025. p. 301–316. https://doi.org/10.1007/979-8-8688-1318-4_16
48. Gay GR, Li CQ. AChecker: open, interactive, customizable, web accessibility checking. In: Proc 2010 Int Cross Disciplinary Conf Web Accessibility (W4A). ACM; 2010;1–2. https://doi.org/10.1145/1805986.1806019
49. Clarke A, MacMahon C, Pickering T, Driller M. Assessing the reliability of a novel eye tracking test to measure fatigue in athletes. Sports (Basel). 2025;13(3):71. https://doi.org/10.3390/sports13030071
50. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Routledge; 1988. https://doi.org/10.4324/9780203771587
51. Mróz K, Jonak K, Plechawska-Wójcik M, Siejka A, Zimenkovsky A, Rejdak R. Analysis of visual perception in children using an eye tracker–A pilot study. Adv Sci Technol Res J. 2025;19(2). https://doi.org/10.12913/22998624/195719
52. Jager J, Putnick DL, Bornstein MH. More than just convenient: the scientific merits of homogeneous convenience samples. Monographs of the Society for Research in Child Development. 2017;82(2):13–30.