Uncovering the New Accessibility Crisis in Scholarly PDFs

Publishing Model and Platform Changes Contribute to Declining Scholarly Document Accessibility in the Last Decade

Anukriti Kumar
anukumar@uw.edu
University of Washington
Seattle, WA, USA

Lucy Lu Wang
lucylw@uw.edu
University of Washington
Seattle, WA, USA

ABSTRACT

Most scholarly works are distributed online in PDF format, which can present significant accessibility challenges for blind and low-vision readers. To characterize the scope of this issue, we perform a large-scale analysis of 20K open- and closed-access scholarly PDFs published between 2014–2023 sampled across broad fields of study. We assess the accessibility compliance of these documents based on six criteria: Default Language, Appropriate Nesting, Tagged PDF, Table Headers, Tab Order, and Alt-Text; selected based on prior work and the SIGACCESS Guide for Accessible PDFs [34]. To ensure robustness, we corroborate our findings through automated accessibility checking, manual evaluation of alt text, comparative assessments with an alternate accessibility checker, and manual assessments with screen readers. Our findings reveal that less than 3.2% of tested PDFs satisfy all criteria, while a large majority (74.9%) fail to meet any criteria at all. Worse yet, we observe a concerning drop in PDF accessibility since 2019, largely among open access papers, suggesting that efforts to improve document accessibility have not taken hold and are on a backslide. While investigating factors contributing to this drop, we identify key associations between fields of study, creation platforms used, models of publishing, and PDF accessibility compliance, suggesting that publisher and author choices significantly influence document accessibility. This paper highlights a new crisis in scholarly document accessibility and the need for a multi-faceted approach to address the problem, involving the development of better tools, enhanced author education, and systemic changes in academic publishing practices.

1 INTRODUCTION

Although the proliferation of digital documents in academic publishing has significantly increased the availability of content, the
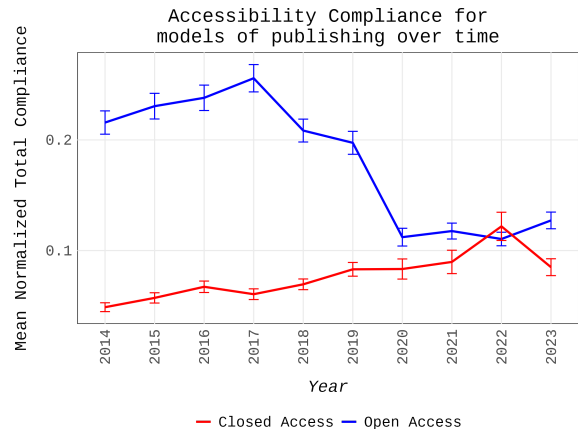
Figure 1: Accessibility compliance for models of publishing over time (2014-2023) with error bars (± standard error). There is a large decline in mean normalized total compliance after 2019 for open access papers. Following this, the compliance values remain stable or increase slightly.

true accessibility of these documents still remains a significant challenge, especially for blind and low-vision (BLV) readers or screen-reader users. Despite advancements in digital publishing technologies, including document creation software, automated accessibility checkers, and evolving standards like the Web Content Accessibility Guidelines (WCAG) and PDF/Universal Accessibility (PDF/UA) guidelines, actual implementations often lag behind these advancements due to significant gaps in the widespread adoption and application of these technologies across the academic landscape. To make PDFs (the most common format for scholarly literature) accessible, they must be annotated with proper reading order, headings, tags, table structure, image alt-text, and more. Unfortunately, creating these annotations is laborious, requiring proprietary tools and a significant amount of knowledge and motivation from authors. Moreover, each PDF variant, even minor revisions, must be annotated anew, posing a continuous barrier to accessibility. As a result, the vast majority of scholarly PDFs are not accessible, leading to high cognitive load and frustration for BLV researchers trying to read these papers.

Poor paper PDF accessibility has been documented in prior work [4, 5, 17, 22, 29], though these studies are limited in terms of generalizability and mechanistic understanding of how technical and publishing decisions contribute to low document accessibility. For

Anukriti Kumar and Lucy Lu Wang

example, most prior work analyzed papers published in fields adjacent to accessible computing—human-computer interaction, disability studies, etc—and do not necessarily generalize to other fields of study. The use of automated checkers to evaluate PDF accessibility focuses on technical compliance with accessibility standards rather than the user experience, which also constrains the generalization of results [5, 10, 17, 22, 29]. Furthermore, existing research does not explore important questions regarding the impact of various document creation software, publishing models (open access vs. closed access), and field of study on document accessibility. Our work aims to fill these gaps by addressing critical research questions: (i) What are the overall status and trends in scholarly document accessibility? (ii) What are the associations between PDF creation platforms, models of publishing (open versus closed access), fields of study, and scholarly PDF accessibility? (iii) To what extent do automated accessibility checkers reflect the performance of screen readers when reading scholarly documents?

To answer these questions, we create a corpus of 20K papers published in the last decade (2014-2023), representing scholarly publications across 19 fields and 766 sub-fields of study. Using this corpus, we characterize the current state of PDF accessibility within academic literature across disciplines and publishing models, as well as accessibility trends over time, providing a comprehensive analysis that has been lacking in existing literature. Our contributions in this work are summarized below:

- We characterize the state of scientific paper PDF accessibility by estimating the degree of adherence to accessibility criteria for our corpus of 20K papers. Our findings highlight that overall compliance levels are very low (only 3.2% satisfy all criteria we measure) and highly variable across fields of study. Additionally, we identify a concerning trend that overall accessibility has been declining since 2019 and remains lower than before. Despite efforts to address accessibility issues, this finding suggests that existing approaches may not be systematically addressing the underlying challenges, and there is a need to understand the factors contributing to this decline.

- Building upon the methods introduced in Wang et al. [37], we study the relationships between creation platforms, publishing models, and fields of study with PDF accessibility, to explain the differences in accessibility compliance we observe. Our analysis reveals significant correlations between the creation platform and PDF accessibility, as well as publishing modality and PDF accessibility. Specifically, we find that documents created by some publishers/publishing software (e.g., Microsoft Word, Adobe In-Design, Elsevier) tend to exhibit higher accessibility compliance, while those created by other software (e.g., LaTeX, Arbortext APP, XPP) show low levels of compliance. We also observe that while open-access papers exhibit significantly higher accessibility compliance than closed-access papers, that a decline in open-access accessibility accounts for much of the overall decline observed since 2019 (Figure 1).

- We assess the effectiveness and limitations of automated accessibility checkers by leveraging secondary assessments with alternate checkers (Adobe versus PAC3) as well as with screen readers (NVDA and VoiceOver). We find that although automated checkers are valuable for large-scale assessments (and

make them possible), relying solely on automated reports may give a false sense of progress, as they can fall short of accurately capturing the nuances of document accessibility. For instance, they do not provide an adequate assessment of alt-text compliance, with many papers that pass the criteria often lacking any usable alt-text. Our findings underscore the continued need for manual intervention from authors and publishers. We also highlight opportunities to improve automated checker capabilities to better align them with the actual accessibility experiences of end-users, particularly those relying on screen readers.

This paper is organized as follows: Following a description of related work in Section 2, Section 3 outlines our data collection and methodology for conducting the analysis of the current state of academic PDF accessibility. In Section 4, we document our analysis results on the accessibility of academic PDFs across various disciplines. Section 5 presents results from our extended qualitative analysis of automated accessibility checkers, offering insights into their reliability. Section 6 discusses the implications of our findings and proposes recommendations for improving PDF accessibility in academic publishing, followed by limitations in Section 7.

2 BACKGROUND AND RELATED WORK

Accessibility is essential for individuals with diverse abilities to fully participate in academic and professional endeavors. For those who are blind and low vision, equitable access to scientific literature can level the playing field, allowing these individuals to pursue education, conduct research, and contribute to scholarly discourse with fewer barriers. Moreover, making digital content, including scholarly literature, accessible is a legal obligation in many jurisdictions, governed by laws such as the Americans with Disabilities Act (ADA) [38] and the European Accessibility Act [1].

2.1 Challenges with PDF Accessibility

PDF documents, while widely used in academic publishing, present significant challenges due to their inherent limitations in accessibility features [17, 24, 28]. Although they support accessibility features like semantic tagging and alternative text for images, the effective use of these features is often limited by authoring tools and a lack of proactive authoring practices [3–5, 37]. Screen reader users often struggle with accessing PDF documents because many authors have not created files to be accessible [4], largely due to three factors: (1) the complexity of the PDF file format, which make it less amenable to certain accessibility features, (2) the dearth of tools, especially non-proprietary tools, for creating accessible PDFs, and (3) the dependency on volunteerism from the community with minimal support or enforcement. These limitations emphasize the PDF format's original intent, to support faithful visual representation of a document for printing, a goal which conflicts with the needs for accessibility.

There are established guidelines and standards, such as the W3C Web Content Accessibility Guidelines (WCAG) [7, 9], and ISO 14289-1 or PDF/UA (Universal Accessibility) [11, 12], that provide comprehensive guidance to authors on making individual elements within PDF documents accessible. These standards also detail the technical implementations necessary to enhance PDF accessibility. However, a significant challenge persists due to the complexity of

these guidelines and the depth of understanding of accessibility issues required to effectively apply them, hindering their widespread adoption [15].

For academic publishing, guidelines and policy changes have been introduced over the past decade to ameliorate some of the issues around scholarly PDF accessibility. Professional organizations such as the Association for Computing Machinery (ACM) have encouraged PDF accessibility through standards and writing guidelines[1]. Influential conferences like the ACM CHI Conference on Human Factors in Computing Systems (CHI) and the ACM SIGACCESS Conference on Computers and Accessibility (ASSETS) have released guidelines and mandates for creating accessible submissions.[2] There has also been a growing trend to outsource some of these accessibility tasks to specialized companies.

Despite having access to these resources and guidelines, Menzies et al. [21] identifies a gap between accessibility guidance and its practical application in academic writing. The authors analyzed content elements such as tables, charts, and images in 330 papers published during 2011-2020 at ASSETS. The analysis focuses on specific design choices that, while maintaining the visual design of the papers, render these elements inaccessible. Ribera et al. [29] conducted a case study on DSAI 2016 (Software Development and Technologies for Enhancing Accessibility and Fighting Infoexclusion), where submitting authors identified barriers to creating accessible proceedings, including the lack of sufficient tooling and lack of awareness of accessibility. The authors recommended creating a new role in the organizing committee dedicated to accessible publishing. In recent years, some publishers (including Science, Nature, PLoS, and others) now provide HTML reading experiences for their papers, which can dramatically mitigate challenges for BLV researchers; the ACM Digital Library[3] provides some publications in HTML format, which is easier to make accessible than PDF [14]. ArXiv has also released a beta version of its HTML Papers project, providing HTML for nearly all newly submitted papers on arXiv along with a number of older ones [13]. These policy changes have led to improvements in localized communities, but have not been widely adopted by academic publishers and conference organizers.

Additionally, Bigham et al. [4] and Brady et al. [5] highlighted the scarcity of tools and dependency on volunteerism as major challenges for creating accessible PDFs. Since 2018, however, a variety of tools and frameworks have been developed to address these issues, focusing on automated solutions [25, 37, 42], artificial intelligence (AI) integration [19, 30, 33], and improved authoring practices [6, 27]. However, the remediation process is still labor-intensive and unintuitive, especially for those not well-versed in accessibility standards [21]. Many content creators are simply unaware of the needs of visually impaired users or the tools available to improve accessibility [17]. In many fields outside of computing, such as Biology and Medicine, versions of record (the final published versions of papers) are produced by publishers from an author's submitted manuscript. This transfer of control over the paper's PDF accessibility from authors to publishers appears to be a step forward. However, publishers rely on specialized providers for formatting and accessibility remediation, a process that incurs significant costs and time, and which may still not lead to accessible PDFs. These options are also not available to everyone: small publishers typically lack the budget for extensive manual or custom processing that larger journals can afford. Similarly, platforms like ArXiv, struggle to implement comprehensive accessibility measures due to the sheer volume and rapid turnover of documents [13].

2.2 Assessing the state of PDF Accessibility

Table 1 lists previous studies that have analyzed PDF accessibility of academic papers and shows how our study compares. Prior work has focused on papers published in venues for accessibility, human-computer interaction, disability studies, and related fields, and many studies are specific to certain publication venues. Brady et al. [5] quantified the accessibility of 1,811 papers from CHI 2010-2016, ASSETS 2014, and W4A, assessing the presence of document tags, headers, and language. They found that compliance improved over time as a response to conference organizers offering to make papers accessible as a service to any author upon request. Lazar et al. [17] conducted a study quantifying accessibility compliance at CHI from 2010 to 2016 as well as ASSETS 2015, confirming the results of Brady et al. [5]. They found that across 5 accessibility criteria, the rate of compliance was less than 30% for CHI papers in each of the 7 years that were studied. The study also analyzed papers from ASSETS 2015, an ACM conference explicitly focused on accessibility, and found that those papers had significantly higher rates of compliance, with over 90% of papers being tagged for correct reading order and no criteria having less than 50% compliance. This finding indicates that community buy-in is an important contributor to paper accessibility.

Nganji [22] examined 200 PDFs of papers published between 2009 and 2013 from four journals in disability studies, revealing a compliance rate of only 15-30% for accessibility features like tagging and alternative text for images. In a more extensive analysis [23], only 15.5% of the PDF documents were found to be tagged, and just 10.5% provided alternative text for images, highlighting a significant gap in adherence to the PDF/UA (Universal Accessibility) standards and WCAG 2.0 guidelines. In a broader survey conducted with 2500 open-access PDF documents from online repositories in Switzerland from 2018-2022 [10], it was revealed that fewer than 11% of documents were found to have minimal accessibility features. This study identified a lack of knowledge and prioritization of PDF accessibility among repository managers, indicating a broader systemic issue that extends beyond individual publishers or conferences to the infrastructure of academic knowledge storage itself. To date, the most extensive study [37] analyzed over 11K PDFs from various fields, also concluding that adherence to accessibility criteria remains low, nearly 2.4% of the documents meeting all accessibility criteria set by researchers, showcasing vast room for improvement in scientific document accessibility.

While these studies provide valuable insights, they tend to be limited in scope and do not cover a wide range of academic fields or track long-term trends. A majority of these works focus on specific content elements without a comprehensive assessment of different aspects of PDF accessibility, such as reading order, logical structure, and the effectiveness of alternative text, which can only be evaluated

---

Prior work	PDFs analyzed	Venues	Year	Accessibility checker
Brady et al. [5]	1,811	CHI, ASSETS and W4A	2011–2014	PDFA Inspector
Lazar et al. [17]	465 + 32	CHI and ASSETS	2014–2015	Adobe Acrobat Action Wizard
Ribera et al. [29]	59	DSAI	2016	Adobe PDF Accessibility Checker 2.0
Nganji [22, 23]	200	*Disability & Society, Journal of Developmental and Physical Disabilities, Journal of Learning Disabilities*, and *Research in Developmental Disabilities*	2009–2013	Adobe PDF Accessibility Checker 1.3
Darvishy et al. [10]	2500	Swiss repositories	2018-2022	Adobe PDF Accessibility Checker
Wang et al. [37]	11,397	Venues across 19 fields of study	2010-2019	Adobe Acrobat Accessibility Plug-in Version 21.001.20145
Our analysis	19,997	Venues across 19 fields and 766 subfields of study	2014–2023	Adobe Acrobat Accessibility Plug-in version 24.002.20687

Table 1: Prior work has investigated PDF accessibility for papers published in specific venues such as CHI, ASSETS, W4A, DSAI, or various disability journals. Several of these works were conducted manually, and were limited to a small number of papers, while more thorough analyses were conducted for CHI and ASSETS, two conference venues focused on accessibility and HCI, and open access papers in a Swiss repository. The analysis conducted by Wang et al. [37] is over broad fields of study and is closest to ours. Our study expands on this prior work to investigate accessibility trends over 19997 PDFs sampled from across different fields of study in the last decade.

manually or using screen readers. Additionally, only a few studies have investigated trends over time, crucial for evaluating the impact of efforts to improve accessibility. In light of these gaps, our study aims to conduct a large-scale, multi-field analysis of scientific PDF accessibility over the past decade, offering insights into trends and factors influencing PDF accessibility in scholarly communication.

3 DATA & METHODS

To capture and better characterize the scope and severity of problems around academic PDF accessibility, we conduct an analysis over a representative sample of scholarly documents using industry-standard accessibility checker tools. This section provides a thorough overview of the methodology employed to perform this analysis. We hope these methods will guide more accurate monitoring of PDF accessibility in the future.

3.1 Corpus Construction

We constructed and analyzed a corpus of 20K papers published in the last decade (2014-2023), stratified across domains of study. We began by sampling documents from OpenAlex,[4] a large, non-profit-operated resource that indexes the metadata of over 250M scholarly works from 250k sources [26]. Initially, we sampled 60K papers including both open and closed-access articles, as indexed by OpenAlex using the 'is_oa' boolean attribute which classifies a paper as open access if there is a freely accessible URL for the full text, without login requirements. For domains, we used the 19 top-level fields of study (e.g., Biology, Computer Science, Sociology) and their associated 766 subfields as defined by Microsoft Academic

Graph [32, 35], which is adopted by OpenAlex. For each subfield of study, we randomly sampled papers from the top 20 venues as defined by aggregate citation count over the last decade.

We downloaded open access PDFs using URLs provided by OpenAlex. For closed access papers, we identified PDFs by matching OpenAlex metadata using a publicly available scholarly document search API and downloaded PDFs from behind paywalls for inclusion in the analysis. We were able to download PDFs for 35K of the 60K papers. While the initial mixture of 60K papers were approximately 2/3 open access and 1/3 closed access, the 35K associated PDFs contained a mixture of 21.5K (61%) open access and 13.8K (39%) closed access papers.[5]

We processed these 35K PDFs through the PyPDF library to remove malformed PDFs; 30K of 35K PDFs could be opened successfully by PyPDF. Finally, we used the Adobe accessibility checker to generate accessibility reports for all remaining PDFs. Of the 30K PDFs, an additional 14.3% could not be processed by the Adobe checker (we could not generate an accessibility report). The accessibility checker most commonly fails because the PDF file is password protected or the PDF file is corrupt. In both of these cases, the PDF could be considered inaccessible to the user. We excluded these PDFs from subsequent analysis. Finally, we re-balanced our dataset by subsampling 20K PDFs with relatively equal representation across subfields, years, and open- versus closed-access papers. While preprocessing and resampling introduces bias, we emphasize that all of these steps tend to identify and remove less accessible

---

[5]This is somewhat counter-intuitive, since we assume that PDFs of open access papers would be easier to find. We believe this shift in distribution is due to several factors, such as (i) noise in the open access URLs provided by OpenAlex and (ii) open access papers not actually being made available by publishers [41].

papers, and any bias introduced is likely to lead to over-estimation rather than under-estimation of accessibility compliance.

This selection process resulted in a dataset sourced from 872 unique publication venues, with representation ranging from 27 to 89 venues (average of 58) per top-level field of study, placing Art at the lower end and Computer Science at the higher end of this spectrum. Additionally, our dataset has an almost equal representation of open and closed access papers, with approximately 10K papers associated with each model of publication. The bulk of our dataset is comprised of published papers, in contrast to preprints or other non-peer-reviewed manuscripts. Selected publication venues are reputable journals, such as *The Lancet* or *Neurology* for Medicine, *The Astrophysical Journal* and *Physical Review Letters* for Physics, and various IEEE and ACM journals for Computer Science.

We extracted metadata for each PDF document in our corpus, focusing on descriptors related to its creation process. To accomplish this, we employed the PyPDF2 library to read the metadata fields '/Creator' and '/Producer', which are populated by the software used to create each PDF. We reviewed all unique PDF creation platforms associated with more than 20 PDFs in our dataset and mapped them to standardized categories. For example, platform names containing `latex`, `pdftex`, `tex live`, `tex`, `vtex pdf`, `xetex` were mapped to the broader category "LaTeX", while names containing `microsoft`, `for word`, `word` and other variants were mapped to the Microsoft Word cluster. We realize that not all Microsoft Word versions, LaTeX distributions, or other versions of the creation platform within a cluster are equal, but this normalization allows us to generalize over these clusters. We use this metadata to analyze the associations between different PDF creation platforms and the accessibility of the resulting PDF document. We identify 24 total clusters, and present statistics associated with the 7 largest clusters, after which there is a steep falloff in cluster size.

3.2 Checking PDF Accessibility

Consistent with prior work [17, 22, 29], we elect primarily to analyze the PDFs in our corpus using the Adobe Acrobat Pro PDF accessibility checker. Though this tool is proprietary and requires a paid license, it is the most comprehensive industry-standard checker available. It is also conducive to large-scale analysis, in that the checker supports the ability to process thousands of PDFs in bulk.

In addition to Adobe, we also evaluated several leading non-proprietary options, including AxesPDF,[6] PDFBox,[7] PDF Inspector,[8] PDFix Lite,[9] PAVE,[10] and PAC3.[11] Our evaluation criteria focused on the comprehensiveness of accessibility checks, ease of integration into our workflow, widespread usage and recognition in the academic and/or publishing communities, and adherence to the Web Content Accessibility Guidelines (WCAG) 2.1 and PDF/Universal Accessibility (PDF/UA) standards. We noted that AxesPDF and PDFBox did not consistently extract accessibility information from PDFs, even when we found these criteria to be met. While PDFA

Inspector is used in prior studies [5], it only analyzes three criteria, whereas our interests extend to other essential accessibility features such as the presence of alt text on figures. PDFix Lite, despite being freely available, lacked the depth of analysis found in more comprehensive tools, especially in evaluating accessibility features like reading order and alt-text evaluation. PAVE, promising for its web-based accessibility checks, was impractical due to its file size limitations, lengthy processing times, and generic issue descriptions without precise location details. PAC3 emerged as a strong contender with its detailed compliance reporting and free availability. However, its adoption is limited as there is no support available for large-scale analysis. Despite this, PAC3's in-depth reporting on standards such as WCAG and PDF/UA compliance made it an invaluable tool for secondary assessment. Hence, we utilized it to further validate our findings and address any limitations of the Adobe Acrobat Accessibility Checker.

For each PDF, we used the Adobe accessibility checker to generate an accessibility report. We access the checker through the user interface of a licensed version of Adobe Acrobat Pro, and use the Action Wizard feature for bulk processing. Each PDF took 10 seconds on average to process and produce a corresponding accessibility report, allowing us to scale up our analysis to tens of thousands of papers. We saved these reports from the checker in HTML format for subsequent analysis.

Each report contains a total of 32 accessibility criteria, marked as "Passed," "Failed," or "Needs manual check."[12] We focus on the following six criteria:

- Default language: The document has a specified reading language.
- Tagged PDF: The document is tagged to specify the correct reading order.
- Tab order: The document is tagged such that the tab order aligns with its structure, thereby facilitating navigation within the document.
- Appropriate nesting: Consistent use of heading levels or tags without skipping levels, reflecting accurate document hierarchy.
- Alt-text: Figures have alternate text.
- Table headers: Tables have headers.

Five of these six criteria were evaluated in prior work [17, 37]. We add the *Appropriate Nesting* criterion since others have emphasized the importance of structural coherence in enhancing document accessibility [10]. For these criteria, we compute and report the following:

- *Criteria Compliance*: for each of the 6 criteria, whether a paper passed;
- *Total Compliance*: the sum number of accessibility criteria met by a paper (e.g., if a paper meets 3 out of 6 criteria, Total Compliance is 3);
- *Normalized Total Compliance*: the proportion of the 6 criteria which are satisfied by a paper (this yields a value between 0 and 1, where 0 indicates the paper met none of the criteria and 1 means the paper met all criteria);

Criterion	Percentage of papers (n=19997)
Default language	17.3%
Tagged PDF	12.6%
Tab order	6.8%
Appropriate Nesting	15.9%
Alt-text	8.5%
Table headers	13.4%
Adobe-6 Compliance	3.2%

Table 2: Percent of papers in our dataset of 19997 PDFs that satisfy each criterion, along with Adobe-6 Compliance.

- *Adobe-6 Compliance*: a binary judgment of whether a paper has met all 6 criteria (1 if all 6 criteria are met, 0 if at least one unmet)

For papers containing no tables and/or figures, we observe that the Adobe checker can still yield either a pass or fail judgment for the Table header and Alt text criteria. Specifically, when objects in the PDF are *not* tagged, the checker tends to fail these criteria, even if the paper does not include tables or figures. Conversely, when objects in the PDF *are* tagged and the PDF is accessible, the checker often passes these criteria, even in the absence of tables or figures. Additionally, we note that if tables are not appropriately tagged as tables (and interpreted as paragraphs instead), the Table headers criteria may also pass.

4 ANALYSIS RESULTS

4.1 Proportion of Papers with Accessible PDFs

Accessibility compliance over all papers is low. Table 2 shows the percent of papers meeting each of the six criteria, as well as the Adobe-6 Compliance rate associated with the sample of papers. Figure 2 shows that the vast majority of papers do not meet any of the six accessibility criteria (14970 papers, 74.86% do not meet any criteria) and very few (634 papers, 3.2%) meet all six. Of the PDFs meeting 1 criterion, the most commonly met criterion is Default Language (1750 of 1802, 97.1%). Of the PDFs meeting 5 criteria, the most common *missing* criterion is Alt-text (364 of 581, 62.6%). In fact, only 1691 PDFs (8.5%) in the whole dataset passed the alt-text check for figures. This is intuitive as Alt-text is the only criterion that typically requires author input to achieve, while the other five criteria can be derived from the document or automatically inferred, depending on the software used to generate the PDF. We also observed a very small number of papers (168) meeting 2 criteria (Total Compliance = 2). Among these, 3 combinations frequently occurred together (Appropriate nesting with Tagged PDF, Table Headers, or Alt Text), accounting for approximately 77% of these papers, while combinations of Appropriate nesting with more than one of the other criteria occurred more frequently, leading to more papers that meet 3 criteria. Refer to Appendix A for correlations between different accessibility criteria.

As shown in Figure 3, all fields have an Adobe-6 Compliance of less than 7%. The fields with the highest rates of compliance are Environmental Science (6.6%), Medicine (4.9%), Psychology (4.5%),
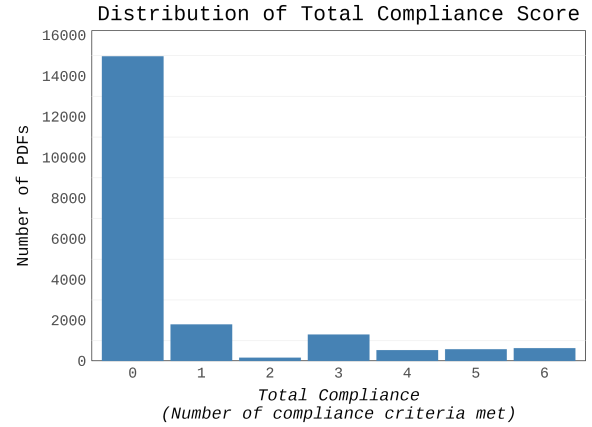

Figure 2: The distribution of numbers of PDFs in our dataset that meet our defined accessibility compliance criteria. A large majority (14970, 74.9%) meet 0 out of 6 accessibility criteria. Of those meeting 1 criterion (Total Compliance = 1), the most commonly met criterion is Default Language (1750 of 1802, 97.1%). Of those meeting 5 criteria (Total Compliance = 5), the most common missing criterion is Alt-text (384 of 581, 66.1%). A small number of PDFs (168) meet only 2 criteria (Total Compliance = 2). Among these, 3 combinations frequently occurred together (Appropriate nesting with Tagged PDF, Table Headers, or Alt Text), accounting for approximately 77% of the papers. This suggests a stronger association among these combinations.
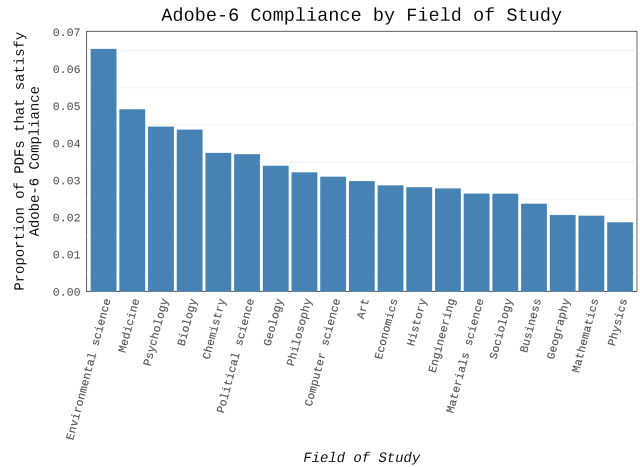
Figure 3: Proportion of papers per field of study that meet all six accessibility criteria defined by Adobe-6 Compliance. Environmental Science (6.6%), Medicine (4.9%), Psychology (4.5%), and Biology (4.4%) have the highest rates of Adobe-6 Compliance while the fields with the lowest rates of compliance are Physics (1.8%), Mathematics (2.1%), and Geography (2.1%). None of these fields had more than 6.6% of PDFs satisfying all the six criteria.
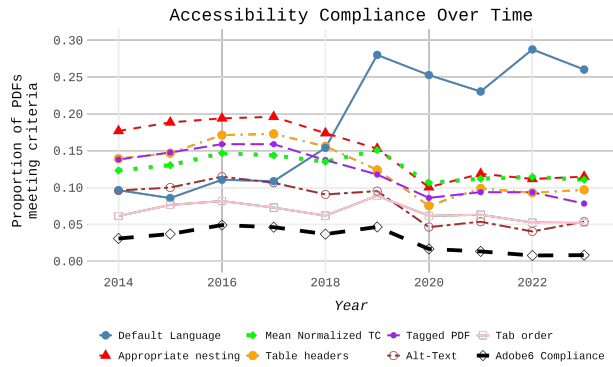
Figure 4: Accessibility compliance over time (2014-2023). While compliance rates seemed to improve slightly over time up through 2017, we observe a large decrease in all criteria but Default Language since then. There is also a significant decline in the mean normalized total compliance and Adobe-6 compliance after 2019. This is a concerning trend in the wrong direction that we attribute to a lack of improvement in accessibility awareness as well as changes in PDF creation platforms and models of publishing.

and Biology (4.4%) while the fields with the lowest rates of compliance are Physics (1.8%), Mathematics (2.1%), and Geography (2.1%). Fields associated with higher compliance tend to align with life sciences and health-related disciplines. Conversely, fields with lower compliance rates are a mixture of physical sciences and mathematical disciplines, business, and social science. Lower accessibility in math-adjacent disciplines could be attributed to the prevalence of complex equations, figures, and specialized formatting in their literature, which are challenging to make compliant without significant effort or specific knowledge of accessibility standards. Also, this variance in compliance rates across different fields may result from different field-specific strategies employed, including document editing and creation platforms, and we explore these associations in Section 4.3.

4.2 Trends in Paper Accessibility Over Time

We show changes in compliance for all criteria over time in Figure 4. With the exception of Default Language, all accessibility criteria demonstrate decreasing compliance rates over the past decade, with a notable drop following 2019. Although rates of decrease varied, the relative ranking of criteria compliance remained mostly consistent; in order from most to least compliant, these are Appropriate Nesting, Table Headers, Tagged PDF, Tab Order, and Alt-Text.

Compliance with the Default Language criterion has seen the most rapid increase, from around 10% in 2014 to over 25% in 2023. This may be due to changes in PDF generation defaults in various creation platforms, making it easier for authors or publishers to comply with this particular criterion. Though this improvement is positive, it is important to note that ensuring a default language is set is the easiest of the six criteria to bring into compliance, and arguably the least valuable in terms of improving the accessible
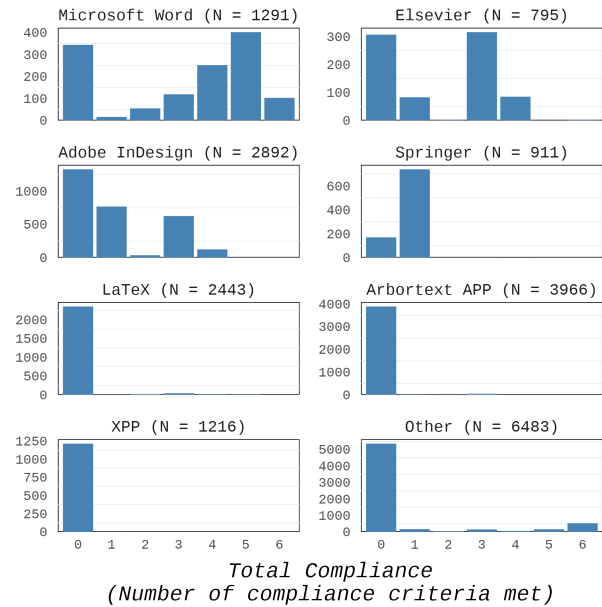

Figure 5: Histograms showing the distribution of Total Compliance scores for each of the top 7 PDF creation platforms, and an 'Other' category that groups the remaining platforms, ordered by decreasing mean Total Compliance. Microsoft Word stands out as producing PDFs with significantly higher Total Compliance than other platforms. Three of the most common PDF creation platform clusters, Arbortext APP, XPP, and LaTeX, produce PDFs with low Total Compliance, with the majority of PDFs at 0 compliance.

reading experience. The criteria showing the lowest rates of compliance are Tab Order and Alt-Text. These findings are particularly concerning because Alt-Text is the only criterion among the six that requires direct input from authors, suggesting a continued lack of engagement with accessibility considerations among authors. It also further supports the theory that any observed improvements can be partially attributed to automatic enhancements made by creation platforms or publisher-level changes rather than a genuine increase in accessibility awareness and compliance among content creators. Furthermore, the general decline in compliance across most criteria, except for Default Language, also resulted in a downward trend in Adobe-6 Compliance. This trend reinforces the notion that despite technological advancements and potential shifts in publisher policies, significant gaps in accessibility remain, underscoring the need for continued efforts to raise awareness and improve compliance with all accessibility criteria, not just the simplest ones.

4.3 Association Between PDF Creation Platform and Paper Accessibility

For analysis, we compare the seven most commonly observed creation platform clusters in our dataset, grouping all others into a

cluster called `Other`. The most popular PDF creators are Arbortext APP, Adobe InDesign, LaTeX, Microsoft Word, XPP, Springer, and Elsevier. Among these, Arbortext APP is used predominantly for technical documentation due to its precision in layout control, while Adobe InDesign is preferred for its advanced layout capabilities across both digital and print media. LaTeX is highly esteemed in academia for its ability to handle complex mathematical formulas and structures effectively. Microsoft Word is widely utilized for its versatility in document editing and formatting across various settings. XPP, or Xyvision Production Publisher, provides advanced automation in creating complex documents, essential in professional publishing environments. Springer and Elsevier are not software but well-known publishing companies that often employ their own proprietary tools or standardized templates to create and typeset documents according to their specific publishing standards. The "Other" category aggregates papers created by the remaining 17 clusters of creation platforms, each with counts of less than 30, as well as those created with unknown PDF creation platforms.

Figure 5 shows histograms of the Total Compliance score for PDFs in the seven most common PDF creation platform clusters. While the vast majority of papers do not meet any accessibility criteria, it is clear that Microsoft Word produces the most accessible PDFs, followed by Elsevier, Adobe InDesign, and Springer. To determine the significance of this difference, we apply the Kruskal-Wallis H-test [16], a non-parametric method for analysis of variance that can be applied to non-normally distributed data. With the PDF creation platform clusters as the sample groups and the Total Compliance as the measurements for the groups, we compute a Kruskal-Wallis H statistic of 6378.8 ($p < 0.001$). This indicates a significant difference in the distribution of Total Compliance scores between the seven most common PDF creation platforms.

In Appendix C, we provide plots of correlations between the proportion of PDFs typeset using different creation platforms per field of study and the corresponding mean normalized Total Compliance rates for those fields. Higher rates of Microsoft Word usage are statistically correlated with higher mean normalized Total Compliance ($r = 0.66$, $p < 0.01$). Conversely, higher rates of LaTeX usage are statistically correlated with lower mean normalized Total Compliance ($r = -0.74$, $p < 0.001$). There is also a positive correlation between PDFs created with header Elsevier and higher accessibility compliance ($r = 0.53$, $p < 0.05$). Correlation scores for Adobe InDesign, Springer, and XPP are non-significant (r = 0.16, 0.12, and -0.06 respectively, all non-significant).

We plot changes in the usage of each of the seven PDF creation platforms over time in Figure 6. Early in the decade, papers were predominantly created using Arbortext APP and Adobe InDesign. Recently, however, there has been a noticeable growth in the usage of LaTeX, Microsoft Word, and Springer. Conversely, Arbortext APP, Elsevier, and XPP are witnessing a gradual decline in usage, suggesting they are being replaced in PDF headers, becoming obsolete, or are less preferred compared to their alternatives. These shifts in preferences reflect broader changes in the publishing industry, driven by evolving standards and technological advancements. Such changes are often influenced by endorsements from leading institutions and journals, contributing to significant trends in document accessibility. Refer to Appendix B to examine trends of creation platforms used in open versus closed publishing models
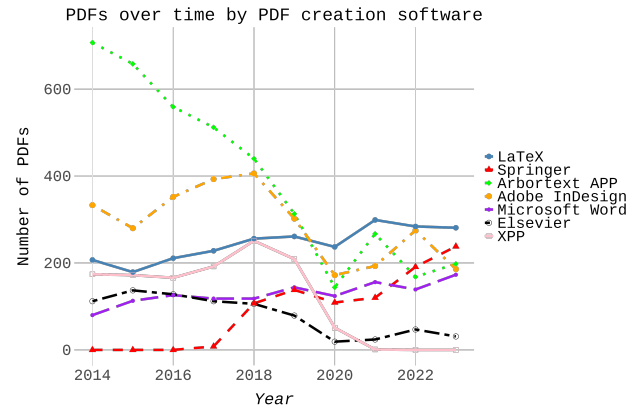

Figure 6: The number of PDFs created by the seven most common PDF-creating platforms over time. Overall, platforms such as LaTeX, Microsoft Word, and Springer are increasing in popularity over time. Conversely, the popularity of Arbortext APP, Adobe InDesign, Elsevier, and XPP has declined significantly.
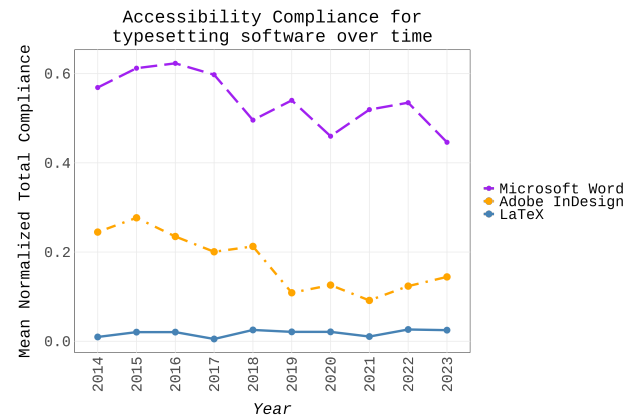

Figure 7: Mean normalized accessibility compliance by select PDF creation platforms over time (2014-2023). While accessibility compliance for PDFs created using LaTeX remains relatively stable (and low), there is a significant decline in mean normalized total compliance for PDFs created with Microsoft Word and Adobe InDesign after 2017.

over time. Splitting by open versus closed access shows large differences, with platforms like Adobe InDesign and LaTeX favored among open access papers along with the rise of Microsoft Word, while platforms like Arbortext APP were heavily used in the earlier part of the decade among closed access papers and dominate yet along with LaTeX.

To further assess the influence of major PDF creation platforms (Microsoft Word, Adobe InDesign, and LaTeX) on accessibility compliance over time, we plot the mean normalized compliance of papers created using these three platforms over time in Figure 7.

Accessibility Compliance over
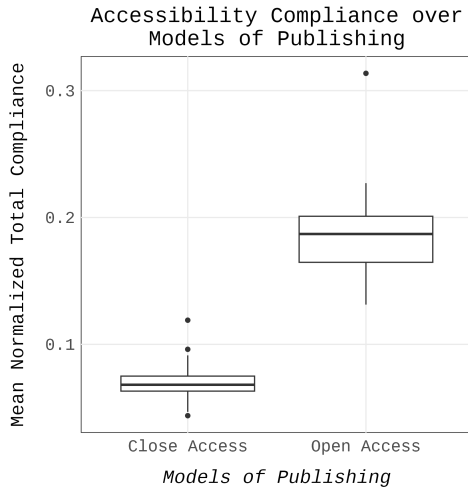Models of Publishing

Figure 8: Accessibility compliance for the two models of publishing (open vs. closed). The median of mean normalized total compliance across all fields of study for open-access papers is 0.18 (IQR = 0.04), significantly higher than the 0.07 (IQR = 0.01) observed for closed-access papers. This demonstrates that open-access papers have higher accessibility compliance compared to closed access ones. As per the Mann-Whitney U test, there is a significant difference in the distribution of mean normalized total compliance scores between the two access types (Z = 24.76, p < 0.0001).

While accessibility compliance for PDFs created with LaTeX remains relatively stable (and inaccessible), a decline in compliance is observed for PDFs created using Microsoft Word and Adobe InDesign, especially for the period of time from 2017 to 2020. This decline may be partially attributed to updates in these platforms that prioritized visual and functional enhancements without adequate integration of accessibility standards. For example, the 2019 updates to Microsoft Office introduced AI-driven tools,[13] designed to improve user productivity did not include necessary accessibility features like tagging or alt text. Furthermore, the adoption of these new features without corresponding increases in accessibility awareness or training likely contributed to the observed drop in compliance. This trend aligns with the overall patterns in compliance scores over time, as detailed in Figure 4.

4.4 Association Between Models of Publishing and Paper Accessibility

Our dataset contains near equal representation of open (10.1K) and closed (9.9K) access models of publishing. Our hypothesis was that open access papers would demonstrate higher accessibility compliance compared to closed access papers due to the broader visibility and inherent motivation for wider dissemination in the open access model. To assess this hypothesis, we studied accessibility compliance across these two access types, as illustrated in

---

[13] See https://www.microsoft.com/en-us/microsoft/blog/2019/powerpoint-ai-upgrade-designer and https://venturebeat.com/ai/features-office-2019/

Figure 8. The median of accessibility compliance over fields of study for open-access papers was 0.18 (IQR = 0.04), significantly higher than the 0.07 (IQR = 0.01) observed for closed-access papers.

To determine the significance of this difference, we applied the Mann-Whitney U test, a non-parametric method suitable for comparing the distributions of two independent groups with non-normally distributed data. With the publishing access type (open vs. closed) as the sample groups and the accessibility compliance as the measurements for these groups, we found a significant difference in the distribution of mean normalized total compliance scores between the two access types (Z = 24.76, p < 0.0001).

Figure 1 shows compliance changes by publication model (open vs. closed access) over time. The plot reveals a large decline in mean normalized total compliance for open access papers after 2019, followed by relatively stable or slightly increasing compliance values. These trends suggest that the decline in accessibility compliance observed in Figure 4 can largely be attributed to a decline in compliance among open access publications. This finding supports our initial hypothesis and highlights that publishing model has a strong influence on PDF accessibility compliance. While improvements in accessibility compliance are still needed across the board, the emerging crisis among open access publications begs further attention and marks an unwelcome departure from observed improvements in scholarly PDF accessibility documented in prior work.

5 ASSESSING THE LIMITATIONS OF AUTOMATED ACCESSIBILITY CHECKERS

While we rely on automated accessibility checkers to enable large-scale analysis, we acknowledge they are imperfect instruments. We therefore conduct additional analysis to qualify the limitations of these methods, through manual evaluation of alt text, comparative assessments with an alternate accessibility checker, and manual assessments with screen readers.

5.1 Verifying alt-text quality

We further process PDFs that "Passed" the alt-text criteria and extract the corresponding author-written alt-texts. To extract alt-text, we employ the method described in Chintalapati et al. [8], which uses the Adobe Acrobat Pro PDF to HTML conversion utility to convert the sample of documents into HTML, from which we access the alt-text associated with each figure. We successfully converted and extracted alt-texts from 1536 of 1691 PDFs that "Passed" the Alt-text criteria. Given that significant information content is required of figure alt-text to satisfy BLV user needs, we analyzed these alt-texts to determine whether they contain meaningful descriptions of figure content. Although we initially intended to use frameworks from prior work [18, 20, 39] to assess alt-text quality, we found that many alt-texts extracted from our data sample were at or below the lowest evaluative levels of these frameworks. As a result, we manually categorized them into five types: (i) a filepath, (ii) unspecific (words like 'Image' or 'Figure' that are likely auto-generated), (iii) caption (duplicate of the figure caption), (iv) cursory (too short to provide a meaningful understanding of the image), and (v) satisfactory (contains information about the visual aspects of the image).

On analyzing these, we found that 95% of these papers (1461) contained nondescript alt-text such as 'Image,' 'Figure,' or 'Logo.' Approximately 3.2% (49) papers had the file path or URL of the image source in their alt-texts, and 1.4% (21) had minimal descriptions such as 'A picture containing graphical user interface', 'A close-up of a map', and 'A picture containing surface chart', which may not offer sufficient information to understand the visual content of the image. Only 5 PDFs (less than 0.4%) that passed the Alt-text criterion (already only 8.5% among our total sample of 19,997 PDFs) actually contained useful and meaningful alt-text, offering detailed descriptions of figure content. Extrapolating to our entire sample, this equates to around 0.03% of papers meeting the stricter criteria of having "meaningful" alt-text.

5.2 Comparative assessment with Adobe Checker and PAC3

We extend the evaluation of PDF accessibility beyond Adobe's checker by incorporating a secondary assessment with the PAC3 tool. Because the tool has no API or bulk processing feature, assessment is manual and we are forced to limit our analysis to a smaller sample of PDFs. We subsample 133 papers, stratified along the 19 top-level fields of study across total compliance values 0-6 as determined by the Adobe checker. For each of these PDFs, we manually conduct a check using PAC3 and save the report for analysis. Unlike Adobe, which reports element-wise compliance, PAC3 reports are organized based on the WCAG and PDF/UA standards. The criteria reported by these checkers are not equivalent, so we manually identify correspondences to our six assessed criteria in the PAC3 output (Table 3) and extract these values for analysis.

We compare the outputs of these two checkers: Adobe and PAC3, side by side along with the original PDF. Our primary tool for this analysis was Adobe Acrobat Reader, supplemented by manual testing with screen readers (VoiceOver), especially to assess the practical navigational experience with the keyboard. For the Tagged PDF, Table Headers, and Appropriate Nesting criteria, we utilized the Reader to inspect the presence and hierarchy of PDF tags within the tag tree. We further used it to check for the presence of Alt-text for images. We also verified the specification of the Default Language by examining the document properties. Additionally, we evaluated the Tab Order criterion using primarily VoiceOver to verify if the tab order paralleled the document structure. These results are presented in Table 3.[14]

Our comparative analysis of the Adobe and PAC3 checkers revealed discrepancies in compliance values for specific criteria, highlighting potential disagreements between the checkers and their impact on the conclusions drawn from automated assessments. For example, our analysis reveals that the Adobe Checker is more likely to yield false positive classifications for the Table Headers criterion (saying this criterion is not met) compared to PAC3. Conversely, PAC3 appears more lenient, often overestimating compliance, especially for the Appropriate Nesting and Table Headers criteria. For instance, PAC3 passed the majority of documents (more than 80%) in these criteria, overlooking significant accessibility issues.

This could lead to a false sense of progress regarding the state of document accessibility. These findings underscore the limitations of relying solely on automated tools for PDF accessibility evaluation and emphasize the necessity for manual reviews to accurately evaluate PDF accessibility.

5.3 Manual assessments with screen readers

Automated accessibility checker tools provide a binary 'Passed' or 'Failed' judgement while evaluating documents against a predefined set of standards such as WCAG and PDF/UA, but often fall short in assessing a document's practical and functional coherence. We explore the limitations of automated checkers across some of these criteria by conducting manual assessments with screen readers. We read 133 PDFs (sampled across fields and total compliance rates) using NVDA and VoiceOver, and describe key differences between checker results and the qualitative experience of using screen readers to read these documents:

(1) **Tagged PDF**: Although automated tools can verify the presence of tags in a PDF, they do not assess the correctness and appropriateness of these tags (e.g., whether a tag used as a list actually corresponds to list items), which often requires human judgment. These tools also fail to interpret the reading sequence or identify issues caused by visually overlaid content, structured differently in the tag tree. This can result in documents where the tagged order conflicts with intuitive visual cues, making navigation difficult for screen reader users.

(2) **Table Headers and Appropriate nesting**: Automated tools can check for the presence of table header tags, but they cannot determine if these headers are meaningful or if they are correctly associated with the corresponding data cells through attributes like scope or headers. Additionally, these tools often fail with complex tables, such as those with multi-level headers or irregular header arrangements, and thus, require manual verification. Likewise, automated checkers can flag improper heading hierarchies and validate against syntactic rules but fail to assess whether the nesting reflects the intended logical structure of the content. As a result, users might find it challenging to understand these relationships and context within the document.

(3) **Tab order**: While automated tools can check if tab order is specified, they cannot determine if the order makes logical sense. This misalignment often results in a disjointed reading experience, making it challenging for users to navigate the content efficiently or correctly.

(4) **Default Language**: These checkers cannot assess whether the detected language accurately reflects the document's content. For instance, a page could be marked as English while containing primarily French text. They are also not capable of detecting changes in language in a multilingual context. This oversight can lead to scenarios where even if a document is technically compliant, it can be practically misleading for users when the default language does not match the content.

(5) **Alt-text**: While these checkers are capable of detecting the presence of alternative text for images and other non-text elements, they fail to assess the descriptiveness and contextual relevance

---

[14]The false positive and false negative rates reported in Table 3 were measured on 152 papers, which up-samples compliant PDFs, stratified by Adobe compliance score. These rates are not representative of false positive and false negative rates over our entire corpus though serve to illustrate commonly occurring limitations of these tools.

Criterion	Corresponding Guidelines	Adobe Checker	PAC3
Default Language	PDF/UA; WCAG 2.1 guideline 3.1.1 (Language of Page)	Effective; false positives are infrequent due to integration with Adobe Acrobat, which handles metadata well (0.78% false positives)	Incorrectly flags documents as missing the language tag, possibly due to less sophisticated recognition of metadata (17.05% false positives)
Appropriate Nesting	PDF/UA; WCAG 2.1 guideline 1.3.1 (Info and Relationships)	Identifies heading structures accurately, but sometimes fails to detect nested structures in documents with complex layouts (4.8% false positives)	Tends to over-report compliance, suggesting leniency in recognizing heading structures (75% false negatives)
Tagged PDF	PDF/UA; WCAG 2.1 does not specifically address PDF tags, but relevant as per guideline 4.1.2 (Name, Role, Value)	Marks documents as compliant, despite having several untagged elements (11.1% false negatives); overlooks untagged nested elements due to their complex structure, especially tables (1.1% false negatives)	More rigorous in assessing tagged elements; however, incorrectly interprets whitespace or decorative elements as needing tags (91.8% false positives)
Table Headers	PDF/UA; WCAG 2.1 guideline 1.3.1 (Info and Relationships)	Misidentifies non-table elements as tables and flags well-structured but complex tables, despite having header tags (34.2% false positives)	Usually marks tables as compliant and misidentifies styled text as table headers (85.7% false negatives)
Tab Order	PDF/UA does not explicitly address tab order but requires that the document's navigation be logical and consistent; WCAG 2.1 guideline 2.4.3 (Focus Order)	Mostly accurate and checks tab order to match document structure, but does not assess element-level reading order within pages, which overlooks finer navigation issues	Focuses on keyboard navigation by checking the focus order rather than tab order, which could miss issues related to visual document flow
Alt-Text	PDF/UA; WCAG 2.1 guideline 1.1.1 (Non-text Content)	Checks for the presence of alt-text but does not assess the quality or appropriateness of the descriptions, which could lead to non-descriptive alt-text passing the check. Sometimes, incorrectly flags decorative images that do not require alt-text (10% false positives)	Similarly checks for alt-text presence without assessing quality; also incorrectly flags decorative images and links that do not require alt-text (10% false positives)

Table 3: Comparative analysis of the Adobe Checker and PAC3, detailing their performance differences when evaluating PDF accessibility. 'X% false positives' refers to the percent of papers reviewed where the tool incorrectly identified a criterion as non-compliant; the denominator is the number of actually compliant documents. Conversely, 'X% false negatives' refers to the percent of papers reviewed where the tool overlooks actual accessibility issues, incorrectly marking the document as compliant; the denominator is the number of non-compliant documents.

of this text. For example, an alt-text that says "image" or "logo" does not convey meaningful information about the image's content, rendering them inaccessible to screen reader users. Our manual review of alt-texts in Section 5.1 reveals that although Adobe's checker identified them as compliant, only a small fraction were genuinely descriptive and meaningful.

Along all criteria, automated checkers fail to capture readability issues that occur when reading these documents with screen readers. Automated checkers, while useful for initial broad assessments, are not a substitute for human evaluation, and are limited in interpreting the quality and practicality of compliance. We ask the reader to bear this in mind when interpreting our analysis (in that we believe the actual screen reader accessibility of scholarly documents to be even lower than what our analysis reports), or when considering the use of automated accessibility checkers as the sole instrument for validating accessibility. To truly support the accessibility needs of all users, especially those relying on screen readers, a more comprehensive approach that combines both automated and manual review techniques is required. This will help uncover and address disparities between technical compliance and practical accessibility, ensuring more inclusive access to digital content.

6 DISCUSSION AND RECOMMENDATIONS

Our analysis reveals a stark landscape of PDF accessibility within academic publishing. We discovered that a large majority of papers in our sample (74.9%) had zero compliance with accessibility criteria across the spectrum of academic fields, with only a small fraction of papers (3.2%) fully compliant. Among the assessed criteria, Default Language was the most commonly met, reflecting the minimal effort required for compliance in this area. Conversely, Alt-Text for figures, which necessitates deliberate author input, was the criterion least likely to be met. This underscores a broader issue: the gap between the potential for accessibility and its actual implementation by authors and publishers.

A concerning trend emerged when analyzing the data over time. Before 2019, PDF accessibility was relatively stable or even improving gradually. However, afterwards, we observed a notable and unexpected decline in accessibility compliance, marking a significant shift in the landscape of scholarly PDF accessibility. What went wrong? Our findings suggest a correlation between open-access papers and higher accessibility compliance, possibly due to the motivation for broader dissemination associated with the open publishing model, encouraging authors to implement accessibility standards more diligently. However, it is among open access papers that we observe the largest drop in accessibility compliance around

2019. We attribute this to several interrelated developments in the publishing ecosystem.

The period after 2019 witnessed a significant increase in the volume of academic publications, especially open access publications. This was driven in part by Plan S,[15] which mandates that scientific publications from publicly funded research be published in compliant Open Access journals or platforms by 2021. The push towards open access may have introduced financial pressures on publishers to reduce costs, which could have resulted in cuts to resources allocated for ensuring accessibility. Additionally, the shift towards more digital and rapid publishing methods like preprints and online-first publications prioritized speed and broad access to information over compliance with detailed accessibility standards. This trend was further accelerated by the COVID-19 pandemic, which heightened the demand for rapid communication of research findings crucial to public health responses. Indeed, around 16% of papers in our dataset published after 2019 include COVID-related keywords. However, accessibility changes among COVID-related papers alone do not account for the severity of the drop in accessibility compliance, which implies more systemic and perhaps persistent changes in scholarly publishing.

We also uncovered a strong association between the platform used to create a document and its accessibility compliance. Microsoft Word was found to produce the most accessibility-compliant PDFs, while those produced by XPP, Arbortext APP, and LaTeX had the lowest compliance, underscoring the significant impact of creation platform choice on accessibility. Microsoft has made investments in enhancing the accessibility of their Office 365 Suite,[16] integrating features like live captions and subtitles in PowerPoint in 2020, which reflects a growing awareness and prioritization of accessibility concerns during document creation. However, the 2019 update of Microsoft Office and subsequent updates introduced AI-driven tools,[17] aimed at boosting user productivity and enhancing user experience, such as AI in Word offering writing suggestions and PowerPoint suggesting design layouts. While these innovations offered significant benefits, they did not fully integrate necessary accessibility features, which may have contributed to the observed decline in accessibility compliance. For instance, advanced formatting features disrupt the tab order necessary for screen readers. We encourage developers of typesetting and publishing software to prioritize accessibility in their development process to address these issues.

While technological advancements are crucial, we caution that not all aspects of PDF accessibility can be automated through software alone, and their limitations must be recognized. Software may be able to effectively infer reading order and detect structural elements like headings and table headers, which are feasible tasks for visual layout-enhanced language models [31, 40] or vision-language models like GPT-4o [2]. However, certain aspects of PDF accessibility require a deeper level of understanding and contextual decision-making that models alone may struggle to provide. For instance, creating meaningful alt-text for figures demands an understanding of the author's intent and the document context,

which is significantly more challenging to automate [33]. Authors therefore continue to play an indispensable role in ensuring the accessibility of documents.

Despite the potential for technology to support accessibility efforts, we have not observed any improvements in accessibility compliance, likely because accessibility concerns are considered marginal, and are outside of the awareness of most publishing authors and researchers. Significant changes in the authorial and publication processes are needed to change this status quo, and to increase the accessibility of scientific papers for BLV readers and screen reader users. Outside of the accessibility and HCI publishing communities, the likelihood of rapid improvement is lower, and automated or crowdsourcing solutions may be needed to improve the accessibility of publications in these other domains. Moreover, the inherent complexity of certain academic materials, such as scientific papers that include complex equations and formats, presents additional challenges that require specialized knowledge and tools to make them accessible. Training and guidelines for authors on creating accessible content, especially in technically dense fields, as well as advocacy for the importance of accessibility, are crucial components to success.

PDFs have been repeatedly called out as being inaccessible, not only for screen readers, but broadly for reading, especially on mobile and other devices with small screen sizes [24]. Dissociating publishing from PDFs continues to be a good goal for the future. In recent years, alternative publication formats have risen in popularity, such as eLife's dual publication in PDF and HTML,[18] the interactive HTML papers at distill.pub,[19] or the ACM Digital Library's very own dual publication (PDF and HTML) process,[20] which is now available for many of the ACM's computing conferences and journals. Additionally, arXiv has introduced HTML formats for nearly all of its newly submitted papers [13]. Similarly, projects like Paper to HTML [36] can transform scientific papers into more accessible HTML versions, improving readability on mobile devices and accessibility via screen readers. We have no doubt that more such viable alternatives to PDF will arise, and we encourage the community to explore these options when making publishing decisions.

Hence, the substantial variability in accessibility compliance across different fields of study, modes of publication, and creation software highlights the multifaceted challenges and opportunities in improving PDF accessibility in academic publishing. Our results underscore the need for concerted efforts from authors, publishers, software developers, and academic communities to prioritize and enhance the accessibility of scholarly work. By adopting a holistic approach involving technology, human effort, and systemic changes in publishing practices, we can transform accessibility from an afterthought to a fundamental component of scholarly output.

7 LIMITATIONS

We acknowledge the limitations of our analysis methods. First, our study employed automated accessibility checkers, enabling us to conduct large-scale analysis of over 20K PDFs, but it inherently

---

[15]Please visit https://www.coalition-s.org/ for more detailed information on Plan S
[16]https://www.microsoft.com/en-us/accessibility/microsoft-365
[17]See https://www.microsoft.com/en-us/microsoft/blog/2019/powerpoint-ai-upgrade-designer and https://venturebeat.com/ai/features-office-2019/

[18]https://reviewer.elifesciences.org/author-guide/post
[19]https://distill.pub/
[20]https://www.acm.org/publications/authors/submissions

limited the depth of our accessibility assessments. The Adobe accessibility checker, used predominantly in our research, provides binary outputs (Passed/Failed) that do not fully capture the nuances of document accessibility, particularly in the evaluation of the quality and meaningfulness of alt-text. Where possible, we have performed smaller scale manual inspection to corroborate our main analysis results. Second, since we did not want to rely on a single accessibility checker, which may have biased our results towards the specific standards and capabilities of the tool, we supplemented our analysis with PAC3 for a secondary assessment on a subset of 152 papers. However, the limited scope of this analysis prevents us from generalizing its findings, underscoring the need for more sophisticated assessment tools that allow for batch processing, integrate a broader range of accessibility criteria, and offer more granular analysis. Addressing these technological and resource limitations in future work could enable more expansive and robust research, potentially leading to more detailed and actionable insights.

We conducted manual assessments with screen readers (NVDA and VoiceOver) on a stratified sample of 133 PDFs, which revealed discrepancies between automated assessments and the actual user experience. Ideally, these manual assessments should have been conducted with the participation of end-users who rely on these technologies daily. Due to resource and time constraints, we were unable to involve end-users directly in this phase, and instead, conducted these ourselves to supplement our existing large-scale analyses. Future studies could enhance the validity of these findings by incorporating end user testing, which would provide invaluable insights into the real-world accessibility of academic PDFs.

Another potential limitation is that there may exist multiple versions of each paper, and one of these versions may be more accessible but not indexed by our data source, OpenAlex. We analyzed the most readily available versions as indexed by OpenAlex, which may not always be the most accessible. This could potentially impact our findings, as we cannot guarantee the capturing of the best possible accessibility score for each paper. Future studies should consider ways to identify and analyze all available versions of a paper to provide a more accurate assessment of its accessibility compliance. Moreover, our sample primarily contained journal publications, with fewer publication types such as preprints, book chapters, reports, and other educational resources. Expanding the scope of future research to these types of documents could yield a more comprehensive understanding of the current state of accessibility across the entire spectrum of academic publishing.

The observed low accessibility compliance across different fields can be attributed to various decisions made by authors, publishers, societies, and conference organizers, which are often influenced by each field's unique practices and standards. As a result, it is challenging to derive field-specific recommendations. Future research should examine field-specific practices and their impact on accessibility to enable the development of more targeted interventions.

8 CONCLUSION

Our study on scholarly PDF accessibility highlights the significant challenges faced by blind and low-vision readers and screen reader users in accessing academic literature. Our analysis across various disciplines reveals low compliance with accessibility criteria, as well

as a concerning drop in compliance led by open access publications in recent years. These disparities are influenced by factors such as PDF creation platforms and publishing models, emphasizing the need to improve both the available tools and awareness and practices within the academic community. We advocate for collaborative efforts among authors, publishers, and software developers to prioritize accessibility in academic publishing. Furthermore, our manual assessments underscore the importance of supplementing automated checker results with comprehensive evaluations to ensure a more accurate measure of accessibility. As the digital landscape evolves, the academic community should consider moving beyond PDFs to more accessible formats to ensure that all scholarly work is truly accessible to everyone, including those with disabilities. This shift not only aligns with ethical standards but also broadens the dissemination and impact of scholarly work.

ACKNOWLEDGMENTS

REFERENCES

[1] [n. d.]. European accessibility act. https://ec.europa.eu/social/main.jsp?catId=1202&langId=en
[2] [n. d.]. Hello GPT-4o. https://openai.com/index/hello-gpt-4o/
[3] Katrin Angerbauer, Nils Rodrigues, Rene Cutura, Seyda Öney, Nelusa Pathmanathan, Cristina Morariu, Daniel Weiskopf, and Michael Sedlmair. 2022. Accessibility for Color Vision Deficiencies: Challenges and Findings of a Large Scale Study on Paper Figures. *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (2022). https://doi.org/10.1145/3491102.3502133
[4] Jeffrey P. Bigham, E. Brady, Cole Gleason, Anhong Guo, and D. Shamma. 2016. An Uninteresting Tour Through Why Our Research Papers Aren't Accessible. *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (2016).
[5] E. Brady, Y. Zhong, and Jeffrey P. Bigham. 2015. Creating accessible PDFs for conference proceedings. *Proceedings of the 12th Web for All Conference* (2015).
[6] S. Brinn, C. Cameron, D. Fielding, C. Frankston, A. Fromme, P. Huang, M. Nazzaro, S. Orphan, S. Sigurdsson, R. Tay, M. Yang, and Q. Zhou. 2022. A framework for improving the accessibility of research papers on arxiv.org. (2022). https://doi.org/10.48550/arxiv.2212.07286
[7] B. Caldwell, M. Cooper, Loretta Guarino Reid, and G. Vanderheiden. 2008. Web Content Accessibility Guidelines (WCAG) 2.0.
[8] Sanjana Shivani Chintalapati, Jonathan Bragg, and Lucy Lu Wang. 2022. A Dataset of Alt Texts from HCI Publications: Analyses and Uses Towards Producing More Descriptive Alt Texts of Data Visualizations in Scientific Papers. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '22)*. ACM. https://doi.org/10.1145/3517428.3544796
[9] W. Chisholm, G. Vanderheiden, and Ian Jacobs. 2001. Web content accessibility guidelines 1.0. *Interactions* 8 (2001), 35–54.
[10] Alireza Darvishy, Rolf Sethe, Ines Engler, Oriane Pierres, and Juliet Manning. 2023. The state of scientific PDF accessibility in repositories: A survey in Switzerland. *Learned Publishing* 36, 4 (2023), 577–584.
[11] O. Drummer. 2012. PDF/UA (ISO 14289-1) – applying WCAG 2.0 principles to the world of PDF documents. *Proceedings of Computers Helping People with Special Needs* (2012), 587–594. https://doi.org/10.1007/978-3-642-31522-0_89
[12] O. Drummer and M. Erle. 2012. pdf/ua – a new era for document accessibility: understanding, managing and implementing the iso standard pdf/ua (universal accessibility): introduction to the special thematic session. (2012), 585–586. https://doi.org/10.1007/978-3-642-31522-0_88
[13] Charles Frankston, Jonathan Godfrey, Shamsi Brinn, Alison Hofer, and Mark Nazzaro. 2024. HTML papers on arXiv - why it is important, and how we made it happen. *ArXiv* abs/2402.08954 (2024).
[14] Miquel Térmens i Graells, M. B. Cerrejón, M. D. Boladeras, D. Murillo, P. Asensio, and Mireia Ribera Turró. 2007. Estudio de la accesibilidad de los documentos científicos en soporte digital. *Revista Espanola De Documentacion Cientifica* 31 (2007), 552–572.
[15] B. Kelly, D. Sloan, L. Phipps, H. Petrie, and F. Hamilton. 2005. Forcing standardization or accommodating diversity?: a framework for applying the WCAG in the real world. *Proceedings of the 2005 International Cross-Disciplinary Workshop on Web Accessibility (W4A)* (2005), 46–54. https://doi.org/10.1145/1061811.1061820

[16] W. Kruskal and W. A. Wallis. 1952. Use of Ranks in One-Criterion Variance Analysis. *J. Amer. Statist. Assoc.* 47 (1952), 583–621.

[17] J. Lazar, E. Churchill, T. Grossman, G. V. D. Veer, Philippe A. Palanque, J. Morris, and Jennifer Mankoff. 2017. Making the field of computing more inclusive. *Commun. ACM* 60 (2017), 50 – 59.

[18] Alan Lundgard and Arvind Satyanarayan. 2022. Accessible Visualization via Natural Language Descriptions: A Four-Level Model of Semantic Content. *IEEE Transactions on Visualization and Computer Graphics* 28 (2022), 1073–1083.

[19] Kelly Mack, Edward Cutrell, Bongshin Lee, and Meredith Ringel Morris. 2021. Designing Tools for High-Quality Alt Text Authoring. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '21)*. Article 23, 14 pages. https://doi.org/10.1145/3441852.3471207

[20] Kelly M. Mack, Edward Cutrell, Bongshin Lee, and Meredith Ringel Morris. 2021. Designing Tools for High-Quality Alt Text Authoring. *The 23rd International ACM SIGACCESS Conference on Computers and Accessibility* (2021).

[21] R. Menzies, G. Tigwell, and M. Crabb. 2022. author reflections on creating accessible academic papers. *Acm Transactions on accessible Computing* 15 (2022), 1–36. Issue 4. https://doi.org/10.1145/3546195

[22] J. Nganji. 2015. The Portable Document Format (PDF) accessibility practice of four journal publishers. *Library & Information Science Research* 37 (2015), 254–262.

[23] J. Nganji. 2018. An assessment of the accessibility of PDF versions of selected journal articles published in a WCAG 2.0 era (2014–2018). *Learned Publishing* 31 (2018), 391–401. Issue 4. https://doi.org/10.1002/leap.1197

[24] Jakob Nielsen and Anna Kaley. 2020. PDF: Still Unfit for Human Consumption, 20 Years Later. https://www.nngroup.com/articles/pdf-unfit-for-human-consumption/. Accessed: 2021-01-31.

[25] Debashish Pradhan, Tripti Rajput, Aravind Jembu Rajkumar, Jonathan Lazar, Rajiv Jain, Vlad I. Morariu, and Varun Manjunatha. 2022. Development and Evaluation of a Tool for Assisting Content Creators in Making PDF Files More Accessible. *ACM Trans. Access. Comput.* 15, 1, Article 3 (mar 2022), 52 pages. https://doi.org/10.1145/3507661

[26] Jason Priem, Heather Piwowar, and Richard Orr. 2022. OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. arXiv:2205.01833 [cs.DL] https://arxiv.org/abs/2205.01833

[27] A. J. Rajkumar, J. Lazar, J. B. Jordan, Alireza Darvishy, and H. Hutter. 2020. PDF Accessibility of Research Papers: What Tools are Needed for Assessment and Remediation?. In *HICSS*.

[28] M. Ribera. 2008. Are PDF Documents Accessible? *Information Technology and Libraries* 27 (2008), 25–43. Issue 3. https://doi.org/10.6017/ital.v27i3.3246

[29] M. Ribera, R. Pozzobon, and S. Sayago. 2019. Publishing accessible proceedings: the DSAI 2016 case study. *Universal Access in the Information Society* (2019), 1–13.

[30] F. Schmitt-Koopmann, E. Huang, and A. Darvishy. 2022. Accessible PDFs: Applying Artificial Intelligence for Automated Remediation of STEM PDFs. *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility* (2022), 1–6. https://doi.org/10.1145/3517428.3550407

[31] Zejiang Shen, Kyle Lo, Lucy Lu Wang, Bailey Kuehl, Daniel S. Weld, and Doug Downey. 2022. VILA: Improving Structured Content Extraction from Scientific PDFs Using Visual Layout Groups. *Transactions of the Association for Computational Linguistics (TACL)* (2022).

[32] Zhihong Shen, Hao Ma, and Kuansan Wang. 2018. A Web-scale system for scientific knowledge exploration. In *Proceedings of ACL 2018, System Demonstrations*. Association for Computational Linguistics, Melbourne, Australia, 87–92. https://doi.org/10.18653/v1/P18-4015

[33] Nikhil Singh, Lucy Lu Wang, and Jonathan Bragg. 2024. FigurA11y: AI Assistance for Writing Scientific Alt Text. In *Proceedings of the 29th International Conference on Intelligent User Interfaces (IUI '24)*. Association for Computing Machinery, New York, NY, USA, 886–906. https://doi.org/10.1145/3640543.3645212

[34] Shari Trewin. 2014. Accessible PDF Author Guide. https://www.sigaccess.org/welcome-to-sigaccess/resources/accessible-pdf-author-guide/

[35] Kuansan Wang, Zhihong Shen, Chiyuan Huang, Chieh-Han Wu, Darrin Eide, Yuxiao Dong, Junjie Qian, Anshul Kanakia, Alvin Chen, and Richard Rogahn. 2019. A Review of Microsoft Academic Services for Science of Science Studies. *Frontiers in Big Data* 2 (2019).

[36] Lucy Lu Wang, Jonathan Bragg, and Daniel S. Weld. 2023. Paper to HTML: A Publicly Available Web Tool for Converting Scientific Pdfs into Accessible HTML. *SIGACCESS Access. Comput.* 134, Article 1 (jan 2023), 1 pages. https://doi.org/10.1145/3582298.3582299

[37] Lucy Lu Wang, Isabel Cachola, Jonathan Bragg, Evie (Yu-Yen) Cheng, Chelsea Hess Haupt, Matt Latzke, Bailey Kuehl, Madeleine van Zuylen, Linda M. Wagner, and Daniel S. Weld. 2021. Improving the Accessibility of Scientific Documents: Current State, User Needs, and a System Solution to Enhance Scientific PDF Accessibility for Blind and Low Vision Users. *ArXiv* abs/2105.00076 (2021). https://api.semanticscholar.org/CorpusID:233481964

[38] P. G. Watson. 1990. The americans with disabilities act: more rights for people with disabilities. *Rehabilitation Nursing* 15 (1990), 325–328. Issue 6. https://doi.org/10.1002/j.2048-7940.1990.tb01505.x

[39] Candace Williams, Lilian de Greef, Ed Harris, Leah Findlater, Amy Pavel, and Cynthia Bennett. 2022. Toward supporting quality alt text in computing publications. In *Proceedings of the 19th International Web for All Conference* (Lyon, France) *(W4A '22)*. Association for Computing Machinery, New York, NY, USA, Article 20, 12 pages. https://doi.org/10.1145/3493612.3520449

[40] Yang Xu, Yiheng Xu, Tengchao Lv, Lei Cui, Furu Wei, Guoxin Wang, Yijuan Lu, Dinei A. F. Florêncio, Cha Zhang, Wanxiang Che, Min Zhang, and Lidong Zhou. 2021. LayoutLMv2: Multi-modal Pre-training for Visually-Rich Document Understanding. *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing* abs/2012.14740 (2021).

[41] Y. Yamada, A. Nishikawa-Pacher, and J. A. T. d. Silva. 2023. Is it open access if registration is required to obtain scientific content? *European Science Editing* 49 (2023). https://doi.org/10.3897/ese.2023.e98101

[42] Shaban Zulfiqar, Safa Arooj, Umar Hayat, Suleman Shahid, and Asim Karim. 2020. Automated Generation of Accessible PDF. In *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '20)*. Association for Computing Machinery, New York, NY, USA, Article 91, 3 pages. https://doi.org/10.1145/3373625.3418045

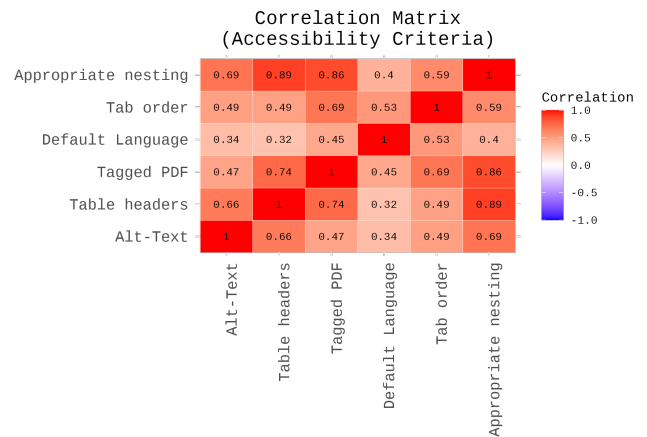## A CORRELATIONS BETWEEN ACCESSIBILITY CRITERIA

Figure 9: Pearson correlation coefficients between accessibility criteria. We observe strong positive correlations between several criteria, notably between Appropriate Nesting and Table headers ($r = 0.89$), Appropriate Nesting and Tagged PDF ($r = 0.86$), and Tagged PDF and Table Headers ($r = 0.74$).

In Figure 9, we plot a heatmap that displays the Pearson correlation coefficients for various accessibility criteria used in evaluating PDFs. We observe strong correlations between Alt-Text and Table Headers (0.66), Alt-Text and Appropriate Nesting (0.69), Table Headers and Tagged PDF (0.74), Tagged PDF and Appropriate Nesting (0.86), and Table Headers and Appropriate Nesting (0.89). Additionally, we note moderate correlations between Tagged PDF and Default Language (0.45), as well as Default Language and Tab Order (0.53), indicating these combinations frequently appear together in papers. Other correlations range from lower to moderate, as depicted by varying shades from lighter red to white on the heatmap.

B USAGE OF PDF CREATION PLATFORMS IN DIFFERENT PUBLISHING MODELS

To examine the trends in PDF creation platform usage within different publishing models, we plot the usage of the seven major
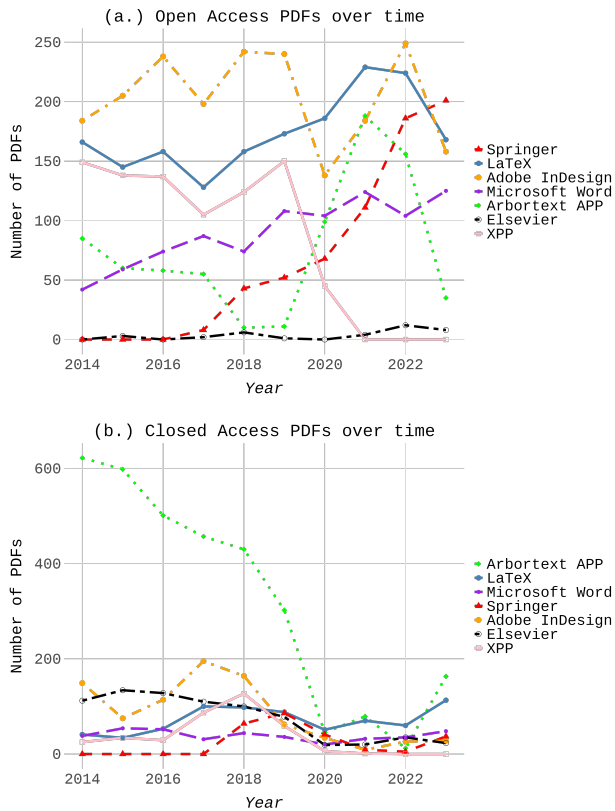
Figure 10: The number of (a) Open Access and (b) Closed Access PDFs created by the seven most common PDF-creation platforms over time.

$p < 0.01$, 95% CI shown) with higher compliance in fields like Medicine, Environmental Science, Business, Political Science, Sociology, Economics, and Psychology, where its usage exceeds 15%. Elsevier also shows a strong positive correlation ($r = 0.53$, $p < 0.05$, 95% CI shown) with higher compliance, especially in Political science and Business. Conversely, LaTeX demonstrates a strong negative correlation ($r = -0.74$, $p < 0.001$, 95% CI shown) with lower compliance, particularly in fields such as Mathematics, Physics, and Computer Science, where it is used similarly extensively. Correlations for other creation platforms (Adobe InDesign, Springer, and XPP) are non-significant ($r = 0.16$, 0.12, and -0.06 respectively, all non-significant).

PDF creation platforms for open and closed-access papers over time in Figure 10. For open-access papers, Adobe InDesign, LaTeX, and XPP were initially the most utilized platforms. Over the years, we observe a significant increase in the use of Springer and Microsoft Word, while LaTeX's usage has continued to grow steadily. In contrast, XPP's usage has significantly declined, becoming almost negligible for open-access publications. Post-2019, there is a notable increase in the usage of Arbortext APP, corresponding with a decline in the use of Adobe InDesign. For closed-access papers, Arbortext APP was the dominant platform early in the decade but experienced a sharp decline subsequently. The use of other creation platforms has either been stable or has shown a gradual decline, suggesting that these platforms may be becoming obsolete or are less preferred compared to newer alternatives.

C CORRELATIONS BETWEEN CREATION PLATFORMS AND PDF ACCESSIBILITY

We plot the proportion of usage of seven major PDF creation platforms per field of study and the corresponding mean normalized Total Compliance rates for those fields, in Figure 11. There is a strong positive correlation between the proportion of PDFs created using Microsoft Word and their accessibility compliance ($r = 0.66$,
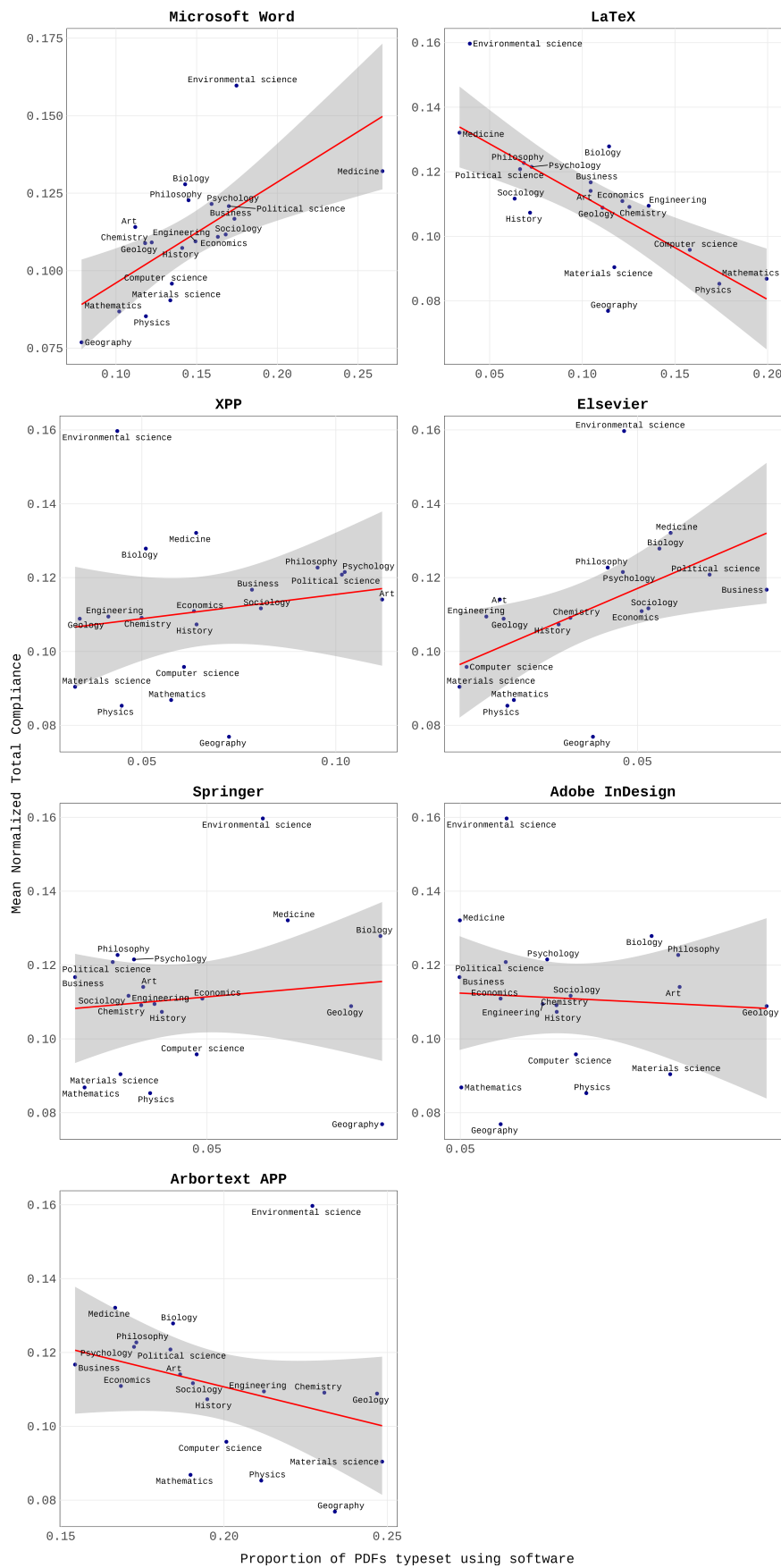
Figure 11: Correlation between PDF Creation Software Usage and Accessibility Compliance by Field of Study.