ChartChecker: A User-Centred Approach to Support the Understanding of Misleading Charts

Tom Biselli
Technical University of Darmstadt
Science and Technology for Peace and Security (PEASEC)
Darmstadt, Germany
biselli@peasec.tu-darmstadt.de

Katrin Hartwig
Technical University of Darmstadt
Science and Technology for Peace and Security (PEASEC)
Darmstadt, Germany
hartwig@peasec.tu-darmstadt.de

Niklas Kneissl
Technical University of Darmstadt
Science and Technology for Peace and Security (PEASEC)
Darmstadt, Germany
kneissl@peasec.tu-darmstadt.de

Louis Pouliot
Technical University of Darmstadt
Science and Technology for Peace and Security (PEASEC)
Darmstadt, Germany
pouliot@peasec.tu-darmstadt.de

Christian Reuter
Technical University of Darmstadt
Science and Technology for Peace and Security (PEASEC)
Darmstadt, Germany
reuter@peasec.tu-darmstadt.de

Abstract

Misinformation through data visualisation is particularly dangerous because charts are often perceived as objective data representations. While past efforts to counter misinformation have focused on text and, to some extent, images and video, developing user-centred strategies to combat misleading charts remains an unresolved challenge. This study presents a conceptual approach through ChartChecker, a browser-plugin that aims to automatically extract line and bar chart data and detect potentially misleading features such as non-linear axis scales. A participatory design approach was used to develop a user-centred interface to provide transparent, comprehensible information about potentially misleading features in charts. Finally, a think-aloud study (N = 15) with ChartChecker revealed overall satisfaction with the tools' user interface, comprehensibility, functionality, and usefulness. The results are discussed in terms of improving user engagement, increasing transparency and optimising tools designed to counter misleading information in charts, leading to overarching design implications for user-centred strategies for the visual domain.

CCS Concepts

• **Human-centered computing → Empirical studies in HCI**; **Social media**; *Empirical studies in collaborative and social computing*.

Keywords

misinformation, disinformation, fake news, user-centred intervention, charts, data visualisation, transparency

DIS '25, Funchal, Portugal

ACM ISBN 979-8-4007-1485-6/25/07
https://doi.org/10.1145/3715336.3735784

ACM Reference Format:
Tom Biselli, Katrin Hartwig, Niklas Kneissl, Louis Pouliot, and Christian Reuter. 2025. ChartChecker: A User-Centred Approach to Support the Understanding of Misleading Charts. In *Designing Interactive Systems Conference (DIS '25), July 05–09, 2025, Funchal, Portugal.* ACM, New York, NY, USA, 28 pages. https://doi.org/10.1145/3715336.3735784

1 Introduction

Global crises such as the COVID-19 pandemic and the War in Ukraine have exacerbated the spread of online misinformation, posing significant challenges due to the algorithmic amplification of false information over factual news [109, 112]. Manual fact-checking and content moderation are resource-intensive, highlighting the need for advanced methods to mitigate this problem. In particular, misinformation conveyed through data visualisation poses unique risks because charts are often viewed as objective representations of information. Previous studies suggest that charts significantly influence user perception [47, 61, 69], yet tools for detecting misinformation in charts lag behind other modalities. Following the approach of other researchers [4, 25, 67, 114], we use the term *misinformation* as an umbrella term that includes both intentionally misleading information, commonly referred to as *disinformation* or, less commonly in academic contexts, *fake news*, and unintentionally misleading information. Misinformation can take many forms, including false news, conspiracy theories, and inaccurate reporting [117]. While charts aim to simplify information, poor design choices can mislead viewers. Well-established design principles, such as clear titles, labelled axes, and appropriate scale, guide good practice [116]. Common problems include manipulation of vertical scales, selective data presentation and incorrect axis labelling, colour reversals, and generally inappropriate chart types [69, 116].

Human-computer interaction (HCI) research addresses misinformation through a variety of approaches, including exploring the factors that drive its spread [25], investigating how social media users engage in the sharing of misinformation [68], and focusing on the user-centred design of digital interventions, such as warnings or corrections, to counter misinformation [44]. Despite some progress, the field is still in need of novel approaches. While chart classification and detection of textual misinformation are relatively well-studied challenging research domains, automatically analysing and transparently informing about a potentially misleading representation of information in charts remains relatively unexplored. Thus, we complement and extend HCI research by taking a user-centred perspective and introduce *ChartChecker* to support users in dealing with misleading information in charts. We chose a participatory design approach to develop a comprehensible user interface of *ChartChecker*, a browser extension to detect misleading features in charts and foster users' critical reflection on deceptive charts. The design was informed by $N = 10$ participant interviews and evaluated in a think-aloud study ($N = 15$). To allow for a more realistic scenario and to not only rely on simulations, we further evaluated the feasibility of the technical data extraction and misleading feature detection in charts. Therefore, we extended previous work on a browser extension that detects misleading features in line charts [32] by implementing the additional type of bar charts and extending the approach to several more misleading features such as non-linear axis scales, multiple axes, or inconsistent tick intervals in addition to truncated or inverted axes. Our study revealed that users desired a direct comparison between the original and recommended charts, along with a transparent list of detected misleading features. ChartChecker was generally well-received, especially for its transparent explanations and the introduction of a recommended chart with a more neutral depiction of information. While the tool has the potential to correct user misunderstandings and help transfer insights to new, unseen charts, it is important to note that ChartChecker provides guidance but not the absolute truth about the "best" way to present chart information. Thus, our contributions are twofold: (C1) providing a user-centred approach to transparently support users in assessing potentially misleading charts, and (C2) advancing the automatic detection of misleading features in line charts and bar charts. Through these contributions, ChartChecker aims to empower users by providing transparent explanations and interactive guidance to encourage critical thinking. By prioritising user-centricity, this work addresses a relatively under-explored area in misinformation research, where most efforts have been focused on modalities other than the chart domain [43–45] or on technically-focused solutions [24, 32].

2 Related Work

This work contributes to designing and evaluating user-centred interventions that assist users' informed navigation of misinformation in charts. We discuss related work on online misinformation and particularly misinformation in charts as a significant challenge people face (see Section 2.1). We shed light on user-centred digital misinformation interventions as one piece of the puzzle to tackle the effects of misinformation (see Section 2.2), and delve into the concept of visualisation literacy (see Section 2.3). Finally, we emphasise the resulting research gaps and present our research questions to address them (see Section 2.4).

2.1 Online Misinformation and Misinformation in Charts

HCI research has started to investigate online misinformation from multiple perspectives, including motivations for misinformation creation and user perceptions of misinformation [25], sharing behaviour under various conditions [68], and digital interventions to counter the effects of misinformation [44]. A majority of these studies tackles the phenomenon with a particular focus on mainly text-based content or typical social media posts combining text and images as commonly found on X (formerly Twitter) [101] or Facebook [104]. In addition, isolated studies investigate misinformation in short videos on TikTok [43, 83], in images on Instagram [113], or in voice messages [45, 76]. The inclusion of graphical or multimodal content in misinformation has an impact on users' beliefs and sharing behaviour, i.e., multimodal misinformation is considered slightly more credible [42], and content with graphic elements is shared more frequently as it captures more attention than text-online content [81].

While graphical content can be misleading on many levels, charts are a specific subcategory characterised by the inclusion of numbers and statistics [89]. As a useful and common means of presenting information, they can easily be presented in a misleading way and are an effective means of spreading misinformation [61, 116]. While this is not a novel problem at all [47], it became more widespread with online content's information overload — a development that strongly motivated our work. Recent research in HCI and related disciplines highlights misinformation in charts, affecting both social media users and readers of online and offline (news) articles and blogs [116]. Distortions and resulting misinterpretations may result from a lack of expertise; however, they can also be intentionally applied to advance a specific agenda [89].

Empirical studies have demonstrated how distorting techniques in charts are applied [89], giving insights that guide our contribution. For instance, Cairo [21] describes multiple ways in which charts can be misleading: by poor design, displaying dubious or insufficient data, concealing uncertainty, or suggesting misleading patterns. Further, Lo et al. [73] developed a taxonomy of misleading elements in visualisations. Research indicates that deceptive techniques include, for instance, truncated axes, disproportionate sizes, inverted angles [89], cherry picking [37, 73, 78], or exaggerated or misaligned titles [59, 61]. Lisnic et al. [69] analysed X data from the COVID-19 pandemic and studied the elements of misinformation in visualisations. They manually annotated 9,958 relevant posts and found that over 90% of those posts that contained interpretations of visualised data also contained reasoning errors, mostly relating to cherry-picked data and causal inference. About 12% of the overall charts contained violations of chart design conventions. Scholars have started empirically evaluating how people are deceived by these tactics under specific conditions, for instance, when deceptive visualisations are combined with accurate text [87]. Their empirical findings confirmed the misleading effects of deceptive data visualisations even when accompanied by factual text, emphasising the need to address the issue in misinformation research.

2.2 User-centred Misinformation Interventions

Mitigating the detrimental effects of misinformation can be reached from multiple perspectives, including facilitating critical journalism, media literacy training at school, or technology-driven solutions like automatic detection [50] or user-centred interventions with a direct influence on end users via information presentation or withholding [44]. The automated detection of misinformation in charts currently presents a number of challenges, and there has been only limited research conducted into the development of support tools for users in this regard despite the existence of promising theoretical frameworks for technical interventions [46].

While various complementing countermeasures exist, HCI research suggests digital user-centred misinformation interventions as one piece of the puzzle to combat misinformation online [44]. Taking a human-centred perspective on technology-driven solutions, they pursue different goals, such as reducing the overall dissemination of misinformation, understanding and reducing misinformation-sharing intentions, or encouraging critical thinking. Interventions can encompass, for example, corrections (e.g., "naturally" occurring due to other users in the comment sections on social media or active corrections by officials and algorithms) [6, 11], collaborative capture the flag competitions [110], or labels to mark content as problematic [7]. Research particularly emphasises users' needs for transparency and comprehensibility in technology-driven credibility assessments [56, 107]. This has partly been addressed in indicator-based interventions displaying cues within the misleading content [12, 43, 103]. The heterogeneous research landscape on misinformation interventions has mostly focused on textual content, with recent approaches tackling visual content as images and videos as well [43, 102].

In the context of misleading charts, scholars have started to explore digital interventions as countermeasures as well. Research on design conventions for data visualisation constitutes a knowledge foundation, providing insights into how conventions are violated and how these violations can be corrected [116]. Ge et al. [37] identified eleven common misleading chart features and tested their impact on interpretation with 497 participants. Users struggled most with manipulated scales, overplotting, and data omission, especially those with lower visualisation literacy. Applying knowledge on violations of design conventions to an empirical study, Wijnker et al. [116] investigated the effectiveness of correction methods for debunking bar charts with manipulated vertical axes, revealing promising effects of presenting an accurate alternative, giving visual cues and text-based cues to activate graph literacy, and text-based warnings.

While the work of Wijnker et al. [116] already proposes some interventions that facilitate informed credibility assessment, this approach is also pursued by Fan et al. [32]. They present a valuable approach in the form of a web app framework that accepts line chart images, semi-automatically extracts data using *WebPlotDigitizer* [75], and checks for three misleading features (i.e., y-axis truncation, inverted y-axis, and misleading aspect ratio). The intervention then reconstructs an annotated version of the chart, still containing the misleading features but pointing them out to users and displaying a corrected version of the original chart [32]. However, including transparent, user-centred explanations of misleading features to address users' needs for comprehensibility [56] is still open for future research. Recent work complements their initial approach by providing further ways of translating chart images into textual information. Some of these operate on rule-based systems [2], others incorporate deep learning-based methods [5], or combine machine learning and rule-based methods to extract key data from charts, such as *ChartOCR* [74]. The latter, a partially rule-based method, analyses multiple chart types (line, bar and pie charts) to enable successful optical character recognition (OCR). Indeed, Fan et al. [32] emphasise the necessity to include other types of charts (e.g., bar charts as very dominant and frequently used visualisation) — a suggestion we draw on in this work. A technological response to misinformation in charts is also evaluated from other perspectives. Lo et al. [72] focus on the effectiveness of different approaches to persuade the designers of charts to adjust their visualisations in favour of less misleading alternatives. They emphasise that while current detection tools like the one by Fan et al. [32] can be helpful for this purpose, a more substantial research focus on communication aspects of identified misleading features is necessary to enhance the overall impact. *VizLinter* by Chen et al. [24], for instance, proposes a framework that helps users to detect flaws in visualisations via a visualisation linter and automatically corrects them. This has been tackled by misinformation research regarding other modalities like videos [43], where an informed credibility assessment is aimed at by encouraging media literacy. This still needs further explorations for charts. In that context, research indicates visualisation literacy as a critical concept to be aimed at when developing user-centred interventions [22].

2.3 Facilitating Visualisation Literacy

Visualisation literacy is defined as *"the ability to confidently use a given data visualisation to translate questions specified in the data domain into visual queries in the visual domain, as well as interpreting visual patterns in the visual domain as properties in the data domain."* [16]. Camba et al. [22] argue that the ability to identify deceptive visualisations is a core element of visualisation literacy that must be trained through learning approaches and interventions, with the effect being stronger the more active the intervention. This relates to indicator-based misinformation interventions referring to the concept of media literacy that have recently been explored for other modalities — particularly text [44]. In that context, user-centred approaches aim to enhance the overall efficiency of digital countermeasures by targeting psychological factors to reduce feelings of reactance and by (implicitly) facilitating users' media literacy skills.

Due to their graphical form and reference to numbers or statistics, misinformation in charts differs strongly from text-based misinformation. Features of textual misinformation (e.g., emotional language or message ambiguity [66]) are only partially viable in the case of misinformation through charts as a portion of their information is not carried in the text but the chart itself. Still, text in visualisations plays a significant role. For instance, research indicates that the recall of a visualisation's message is more frequently aligned with the titles than with the visualisation itself [59], and biased wording of text in visualisations has been explored as having an impact on the perception of the visualisation [106]. Data is often ascribed a certain *veracity* and *neutrality* that is not justified in reality, as the collection of data is always accompanied by (human) decisions about what is collected and what is not. In this respect, data itself is inherently biased [39] and visualisations can be understood as representations of power rather than of knowledge [57, 63]. Hence, design choices can lead to deceptions, especially if viewers are uncritical and unreflected towards data visualisations [49].

Intervening with corrective design principles and strengthening user knowledge about data visualisations can be an important contribution of HCI to combat misinformation. Digital interventions aiming at an increased visualisation literacy have partly been investigated for charts on conceptual levels [95] and as full implementations for individual contexts like misleading line charts [32]. However, a strong user-centred focus is still to be explored, and a broader expansion to relevant visualisations, such as bar charts, is still to be developed [32]. This entails perspectives on the communication of misleading features via suitable user interfaces and corresponding measures of perceived comprehensibility and usefulness of interventions, picking up on HCI research on indicator-based interventions for other modalities like videos [43] and transferring insights to the specific context of charts with its unique potentials to mislead.

2.4 Research Gaps

Our study advances HCI research on user-centred digital interventions to facilitate informed navigation of misleading charts. To do this, we extend the state of research (1) from a user-centred perspective with a particular focus on the user interface of a digital intervention that has been designed applying transparent and comprehensible indicators for misleading charts with explanations and a user-centred visualisation of the recommended chart, as well as its thorough evaluation in a user study. We further advance the state of research (2) extending prior foundational work on the detection of misleading features in charts [32] from a technical perspective, including the detection of additional misleading features, fast chart data extraction, and coverage of bar charts. This work addresses research gaps in the following areas:

1st gap: Design of user-centred misinformation interventions for charts. Other scholars in HCI and related disciplines have emphasised the necessity of user-centred digital misinformation interventions that consider users' needs for comprehensibility [56, 86] and allow for informed navigation of misinformation content [43]. These have partly been investigated for textual content [77], and short-videos [43]. However, little is known about how to technically address visually presented misinformation in charts [116]. In that context, related studies have started to look into interventions for charts aiming at an increased visualisation literacy on a conceptual level [95] and regarding implementations of automatic detection and basic visualisation of features in the specific context of misleading line charts [32].

2nd gap: Automatic detection of misleading charts. Automatic misinformation detection has been largely researched for the context of textual content [50]. This includes not only approaches for automatic filtering of misinformation, often using machine learning, but also the explicit detection of comprehensible misleading features in text [77]. However, research indicates a particular relevance of misinformation in charts as well, as charts constitute an effective means for spreading misinformation [61, 116]. In that context, insights into how distortion techniques are applied [89] and can be detected have been tackled [32] but still necessitate a more exhaustive and user-centred investigation for highly relevant visualisations such as bar charts, as well as efficient implementations.

Our work aims to complement these findings with a user-centred perspective on the design and evaluation of an indicator-based intervention to address users' needs for comprehensibility and facilitate informed credibility assessment in charts. The consideration of related studies and the combination of the resulting gaps lead to the following overarching research questions of this study:

- RQ1: *How can a user-centred, indicator-based approach support users in navigating misleading information in charts?*
- RQ2: *How can existing approaches be improved to enhance the automatic detection of misleading features in charts?*

3 Design and Implementation

The overarching goal of this work is a user-centred design and evaluation of the user interface of ChartChecker, an annotation tool to facilitate navigation of misleading bar charts and line charts. Before focusing on the user interface, we assessed the feasibility of the underlying technology-driven chart annotation (see Section 3.1). Therefore, we built on prior chart annotation approaches, specifically on the foundational browser extension developed by Fan et al. [32], which detects three misleading features in line charts and presents alternative visualisations. Our technical enhancements primarily aimed at exploring the feasibility of integrating automated data extraction through 'ChartOCR' [74], enabling the analysis of bar charts alongside line charts and expanding the detection to seven misleading features. As the main contribution of our work, we then developed the user-centred, indicator-based interface, informed by interviews (N = 10) in a participatory design approach (see Section 3.2) and evaluated it in a think-aloud study (N = 15) (see Section 4), to ensure a comprehensible and usable presentation of the detection results. Participatory design encompasses research methods that actively and iteratively involve end users as stakeholders (in our case: participants of diverse socio-demographic backgrounds such as age, gender, and education) in the design process to ensure that outcomes are user-centred. The concept has been widely applied and discussed in the context of misinformation research for interactive system design as well [100]. A conceptual overview of the study design is visualised in Figure 1.

3.1 Backend: Feasibility of Automated Data Extraction and Feature Detection

From a user perspective, the browser plugin is accessed by right-clicking on an image of a chart to start the analysis. Then, the relevant data is extracted using 'ChartOCR' (see Section 3.1.1), and once successful, the chart and text data are sent to the backend. Subsequently, a detection process of a list of seven misleading features is run in sequence using Python (see Section 3.1.2). The list of detected misleading features, along with the chart and text data, is sent back to the frontend, where two charts — the original misleading chart and the proposed chart without potentially misleading

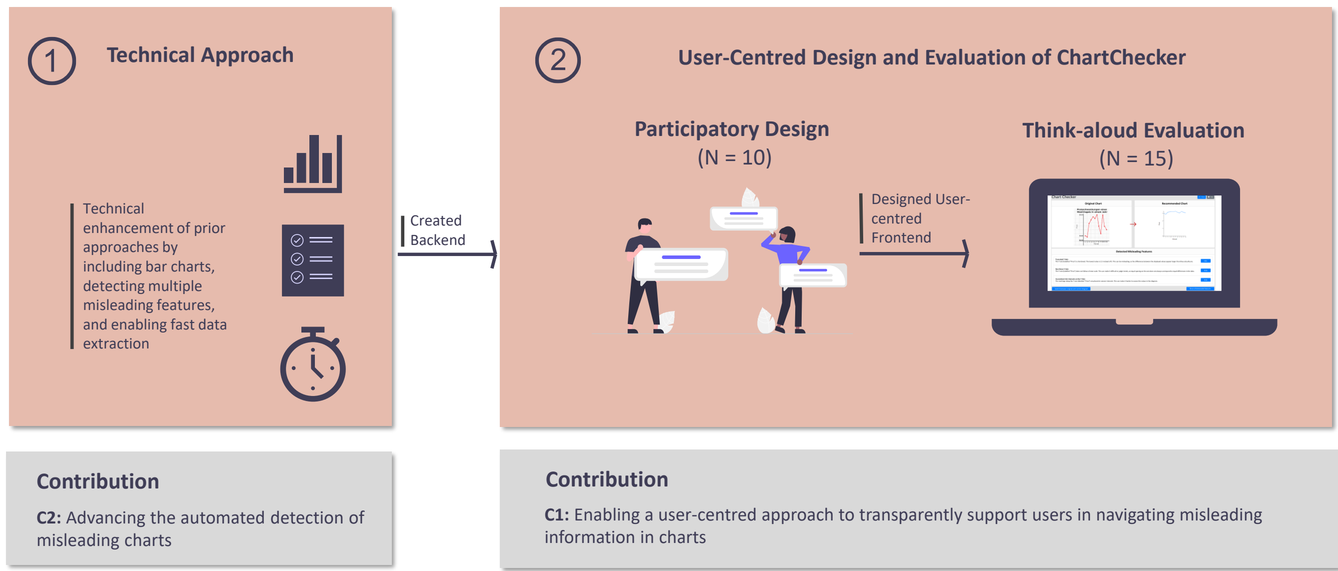


Figure 1: Conceptual overview of the study design

features — are generated using the D3[1] library. An overview of ChartChecker's pipeline is shown in Figure 3.

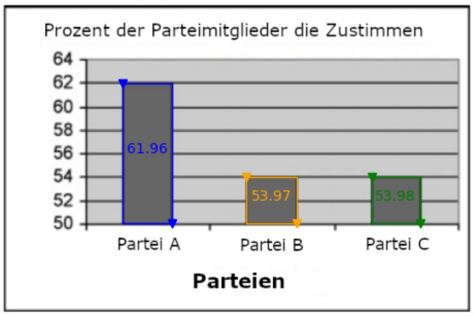
3.1.1 OCR Approach. To reliably and quickly extract data from bar charts as well as line charts, we have integrated elements from 'ChartOCR' [74]. Previous approaches already support the annotation of line charts [32], which, according to Battle et al. [8], account for about 21% of charts. To broaden coverage, bar charts were added to the analysable set alongside line charts as they are among the three most popular chart types besides pie charts [8]. Bar charts are particularly relevant and suitable for analysis because, like line charts, they share similar vulnerabilities to misleading features such as truncated axes, non-linear scales, inconsistent tick intervals and missing labels. ChartOCR uses a combination of rule-based and ML-based approaches to extract data from visualisations. It uses a total of 5 modified CornerNet [62] neural networks (NNs) with different tasks. One is for chart classification and *key point* extraction, where key points are important points depending on the chart type. For bar charts, key points are the individual bars at the top left and bottom right; for line charts, they are the pivot points of the line. In addition, a separate NN and separate rule-based methods are used for each chart type to collect and process more type-specific information. In the case of multiple lines in a line chart, the last NN divides them into separate lines. Examples of the output of our data extraction methods can be seen in Figure 2 (Appendix). With this data extraction approach, we successfully extracted text rotated at a 90° angle, which is commonly used for y-axis labels. Data extraction time took 1.344 seconds for line charts and 0.689 seconds for bar charts (measured on five charts of each type on the backend without transmission times).

In addition to runtime, we qualitatively analysed the accuracy of the new methods by examining the extraction results of diverse charts and found that our version of ChartOCR works with high accuracy, especially when extracting key points in bar charts without errors. However, we found the performance on line charts to be lower than expected, with only four out of seven charts being completely accurate when tested. While the general direction of the lines was accurately extracted, the specific prescribed values, for example, were partly incorrect due to difficulties in extracting y-axis values (see Figure 2 for examples of data extraction results). Overall, text extraction remained a challenge regarding the generated results. Our approach uses the 'pytesseract'[2] library for OCR which extracted y-axis labels correctly only 36% of the time across all line and bar charts tested. Although more advanced OCR algorithms, such as those offered by commercial platforms such as Microsoft Azure OCR exist, their integration was not feasible within the resource constraints of this study. In addition, pytesseract was chosen for its accessibility, which supports the overarching goal of developing an open source, scalable solution. Future work, however, could explore other OCR apporaches or apply post-OCR-correction on the pytesseract-based OCR results to mitigate OCR errors while maintaining the open source approach. Beyond OCR improvements, fine-tuning the feature detection pipeline (section 3.1.2) is another way to mitigate the detrimental effects of noisy inputs, reducing the bottleneck potential for the OCR step. Due to the partial limitations of data extraction, we addressed these cases in the think-aloud study by pre-loading the relevant .csv files with accurate values. This approach allowed us to separate the technical challenges from the user-centred interface, ensuring that we could focus on evaluating the wider potential of our approach in the think-aloud study, with a particular focus on the user-centred interface and users' needs and preferences.

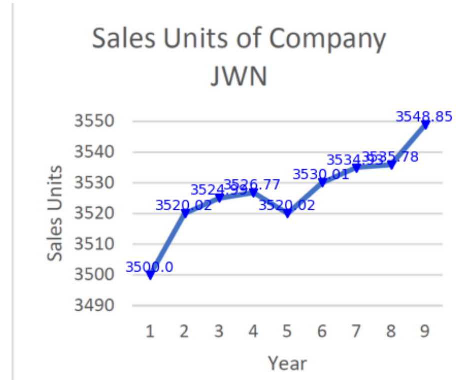
3.1.2 Feature Detection. A large survey of over 1000 chart images by Lo et al. [73] has revealed the most common potentially misleading features. Fan et al. [32] already covered truncated and inverted

[1] d3js.org

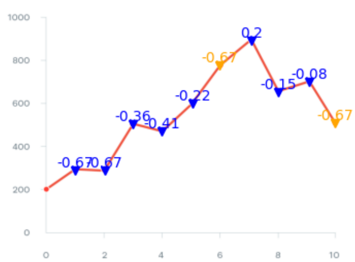
[2] https://pypi.org/project/pytesseract/



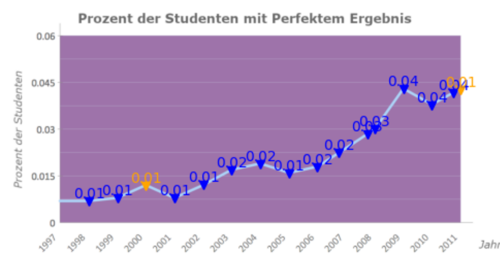
(a) Correctly extracted bar data.



(b) Correctly extracted line data.



(c) The y-axis labels could not be read correctly resulting in incorrect graph values.



(d) Some vertices along the line are being classified as members of a nonexistent second line.

Figure 2: Visualised results of the chart data extraction using our modified version of ChartOCR

y-axes, as well as a misleading aspect ratio (AR), which the survey found covers about 19.5% of potentially misleading features in charts. Our research expands the scope of potentially misleading features covered to approximately 50% by implementing the detection of several additional potentially misleading features, including multiple axes, non-linear axis scales, inconsistent tick intervals, and missing labels (see Table 1 for an overview). ChartChecker analyses whether potentially misleading features such as non-linear axis scaling are present by checking the distances between axis labels and detecting irregularities, such as when ticks are unevenly distributed. Corresponding explanations for the potentially misleading features that are shown to users can be found in Figure 5. As Lo et al. [73] noted, charts classified as potentially misleading contained an average of 1.33 misleading features, so our tool allows multiple dimensions to be detected and displayed simultaneously.

It is important to note that the heuristics used to classify a chart as potentially misleading are based solely on the presence of specific design features that are typically considered problematic. While certain deviations from standard design conventions, such as truncated axes, may be contextually justified – for example, to highlight meaningful fluctuations in particular data ranges – ChartChecker does not currently take these contextual factors into account. The heuristics used by ChartChecker are designed to flag features that violate standard design principles. As such, the tool operates within the constraints of established design conventions and aims to transparently inform users of potential problems, while leaving the final decision to the user.

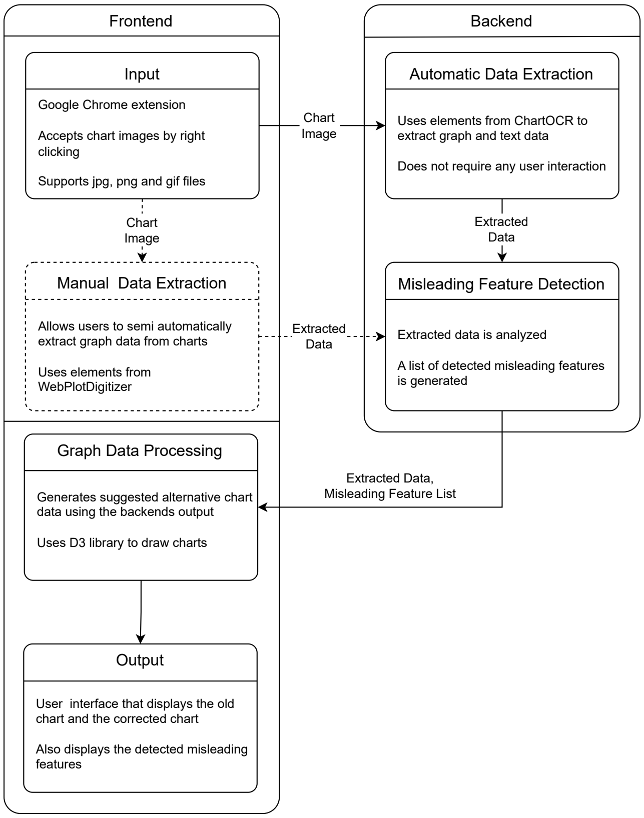


Figure 3: The pipeline of the main elements of ChartChecker. The novel automatic data extraction step replaces the manual data extraction by Fan et al. [32], as shown by the dotted lines.

Previous detections	New detections
Truncated Y-Axis	Multiple axes
Inverted Y-Axis	Non-linear axis scales
Misleading aspect ratio	Inconsistent tick intervals
	Missing labels (includes chart title, axis titles, axis labels, legend titles and legend labels)

Table 1: Shows which misleading features have been previously detected [32] and which new detection algorithms were enabled with ChartChecker's approach.

3.2 Frontend: Design and Implementation

This section details the frontend development, focusing on the design and how a participatory design approach informed the final user interface.

3.2.1 Prototype Design. The frontend is written in HTML/CSS and JavaScript, using the popular bootstrap[3] framework for standardised UI elements, and the D3 library[4] for drawing the charts. For the user interface, we were guided by 13 design principles that are inspired by common guidelines and heuristics for interaction design [34, 54] and were adjusted to the specific context of charts. The guidelines entail, for instance, avoiding unnecessary complexity, minimising cognitive load, and keeping the users' needs in mind (Table 8 (Appendix)). To design a user-centred UI for our study, we first created six prototype UIs to explore variations in layout, features, and presentation. Each variation aimed to balance clarity, cognitive load, and usability while adhering to the design guidelines (Table 8). The prototypes were developed collaboratively by the authors to integrate different perspectives and ensure a comprehensive exploration of design alternatives. Specifically, they varied in the (1) order of the original chart, corrected chart, and list of detected features, (2) whether or not a control chart was included to allow users to verify the accuracy of the OCR-based data extraction, and (3) the inclusion of several additional features such as a sharing function or the ability to individually show and hide specific detected features (see Table 10 (Appendix) for all potential features).

3.2.2 Evaluation and User Feedback. We generally used a participatory design approach (see Section 4.1.4 for a discussion of ethics) and interviewed N = 10 participants from the local (German) community, drawn from the researchers' personal and professional networks (see Table 5 for demographic information). Previous research has demonstrated that a sample size as small as N = 5 can be sufficient to uncover usability-related issues in qualitative studies, yielding meaningful insights [82, 111]. In fact, smaller sample sizes are not uncommon in HCI research, particularly when exploring novel topics [20]. While the empirical focus was on the think-aloud study, the primary objective here was to pre-test different versions and evaluate user preferences before the think-aloud study, with the goal of enhancing the user-centredness of the UI rather than just incorporating developer preferences.

Participants were shown the six UI prototypes as screenshots and asked to rate them on a Likert scale (one to six) in terms of *comprehensibility*, *intuitiveness*, *visual appearance*, *usability* and *overall impression.* They were also asked to rank the six UIs from the one they most liked to the one they liked least (see Table 9). We also presented them with features based on the design principles and asked them to rank them in order of importance on a ten-point scale to prioritise the implementation (see Table 7). Overarching design implications derived from the participatory design approach and the subsequent think-aloud study can be found in section 5.2.

Evaluation Results. Overall, participants gave the highest score to Prototype 1, which included an improved chart and a list of detected misleading features, and considered it the best in terms of intuitiveness, visual appeal and overall impression. However, on average, it only achieved a mid-range ranking compared to the other prototypes. This highlighted the need to integrate positively rated elements from other prototypes in order to develop an improved user interface. In terms of individual features, based on the ratings, the following elements were of particular importance to users: (1) *control charts*, (2) a view that displays all deceptive features, and a *description and explanation* of these deceptions, (3) the ability to *show and hide each deceptive feature* individually, and (4) a *share button.*

[3] https://getbootstrap.com/
[4] https://d3js.org/

Qualitative Feedback. Finally, participants were given the opportunity to provide open comments and ideas, which were collected and implemented when mentioned more than once. These included adding a *red arrow*, a *hide button* for detected deceptions, renaming the improved chart as *recommended*, and rewriting the textual descriptions with *simpler wording* to accommodate users with limited prior knowledge. In terms of visual improvements, we reverted to the *more neutral colour scheme* and other standard bootstrap elements such as title appearance and font type, as participants criticised the colour scheme of the prototypes.

The final UI was created by integrating positively rated features into the overall best rated prototype. This result was then further refined by adding features and improving aspects, based on the participants' open feedback. The end result can be seen in Figure 4. It consists of three main elements: the original image, the recommended chart and the detected misleading features.

4 Evaluation

One of the main goals of our work was to gain in-depth insight into user perceptions of the tool when interpreting visual information. Therefore, we conducted a qualitative think-aloud study [65] to gain rich insights into user perceptions during real-time use, rather than retrospectively [97]. The think-aloud method is already well established in the field of misinformation research [36, 38, 43] and software usability testing [3, 60, 84].

4.1 Method

4.1.1 Study Structure. The study began with a brief introductory segment that collected demographic information and assessed participants' affinity for technology interaction (ATI-S) [115] (administered via SoSciSurvey[5]). A remote connection was then established using Chrome Remote Desktop[6], allowing participants to interact with the tool remotely. Participants were first given a brief introduction to the topic of online visualisation annotation and the procedure for the think-aloud part. A short simulation of the think-aloud method was demonstrated and participants were generally instructed to "think aloud" while using ChartChecker. A chart without misleading features was then presented to demonstrate the use of the tool. This was followed by the think-aloud session in which participants were presented with the first of seven stimuli (see Section 4.1.2 for a description of stimuli) and repeated the following steps for each of the stimuli:

(1) The participant verbalises their interpretation of the chart and its data without using ChartChecker.
(2) The participant activates the automatic analysis feature to access the comparison interface and narrates their thoughts as they navigate through this interface.
(3) The participant is asked to reflect on how their interpretation of the chart and its data has changed.

Upon completion of all stimuli, participants completed the System Usability Scale [18], followed by a series of final interview questions. These questions targeted the four main themes explored during the think-aloud sessions and were designed to elicit comprehensive final opinions on the overarching study questions (see Section C.3 for all questions asked):

- Is the tool's interface intuitive enough to allow for easy ***user interaction***?
- How do participants rate the ***comprehensibility*** and ***functionality*** of the tool's texts and generated charts?
- How do participants perceive the tool's ***usefulness*** and would they use or recommend it?

A diagram of the study's structure can be found in Figure 7 (Appendix). All think-aloud sessions were conducted via Zoom[7] and recorded locally for transcription. One session lasted for about 60 minutes.

The study therefore took a qualitative approach, prioritising depth and user-centred insights into how participants interact with and perceive the tool. Given the lack of user-centred strategies to support users in navigating misleading information in charts, we chose to focus on an in-depth think-aloud methodology at an earlier stage of development, rather than conducting a large-scale, advanced evaluation of quantitative effects. Including pre- and post-test assessments or group differences between an experimental or control group as a quantitative measure of effectiveness was therefore beyond the scope of this study, but would valuable for future research to address.

4.1.2 Stimuli. To address a variety of scenarios while ensuring sufficient time for detailed engagement during the think-aloud approach, a set of seven stimuli was used. It consisted of several real-world examples from news articles, news shows or social media that contained misleading features and was supplemented with synthetically constructed misleading charts. The number of stimuli was limited to allow deep and thorough engagement with each one, facilitating the collection of rich insights through our think-aloud approach. Two researchers put together a diverse set of stimuli covering various topics from different domains that were not currently debated so that participants were unlikely to be familiar with them and ensured that the content was thematically balanced. Care was also taken to remove specific political implications to reduce potential bias (e.g. by changing the heading to a neutral one). The stimuli can be seen in Figure 8, Figure 9 and Figure 10 (while the original ChartChecker screenshots can be found in Figure 11 to Figure 17) and the contained misleading features are summarised in Table 11 (Appendix). Figure 8a and Figure 8b are positive examples that do not contain any misleading features and were included to allow for a more realistic usage of the tool where not all analysed charts are misleading. While the same potentially misleading dimensions apply to line and bar charts, we have aimed to ensure diversity in stimuli mainly by the inclusion of various topics and including a variety of misleading features. Thus, all other visualisations contain one or more misleading features: *misleading aspect ratio* (Figure 9a); *inverted y-axis* (Figure 9b); *non-linear x-axis*, *non-linear y-axis*, and *missing axis labels* (Figure 10a); *truncated y-axis*, *inconsistent tick placement*, and *non-linear y-axis* (Figure 10b); *truncated y-axis* (Figure 10c). This allows for a broad yet non-exhaustive set of

[5]soscisurvey.de
[6]remotedesktop.google.com
[7]zoom.us

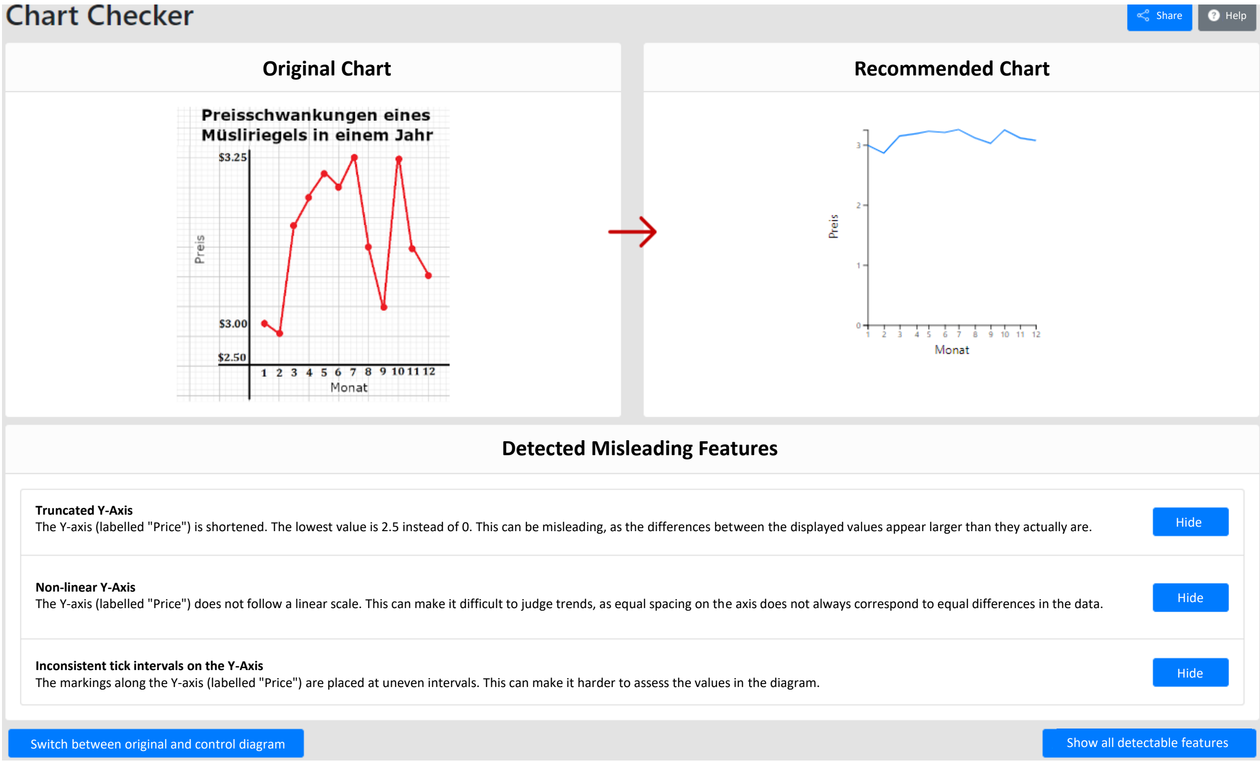


Figure 4: The main view (English translation) of the new user interface we created based on the results of the participatory design approach.

examples of misleading charts, acknowledging that other additional misleading features may not be represented in these instances.

4.1.3 Participants. For our user study, we recruited $N = 15$ participants via Prolific[8]. After conducting 15 interviews, thematic saturation was achieved, where few new insights or perspectives emerged. Consequently, participant recruitment was concluded. Several studies have demonstrated that Prolific is a reliable platform capable of collecting high-quality data [1, 91]. Since misleading charts occur across various platforms and affect users with diverse demographic characteristics (e.g., age, gender, education), we chose to include a broad and varied pool of participants rather than limiting our study to a specific group (e.g., college students) both in the participatory design interviews and the think-aloud study. Inclusion criteria were German language proficiency and residence in Germany, Austria or Switzerland. This was done to mitigate potential Prolific participant recruitment limitations by reducing the risk of including users in the study who claim to understand German but whose level of proficiency is not sufficient to engage deeply with the German version of ChartChecker. Additionally, focusing on this cohort helps mitigate potential biases arising from differing technological infrastructures and regulatory environments that could unintentionally influence the study results. The associated limitations in generalisability are discussed in the limitations section. In addition, the study description included technical requirements for access to Zoom and a Chromium browser. There were no explicit exclusion criteria. Eight participants were male, six were female and one was non-binary. They were between 23 and 53 years old (*mean* = 34), the majority of them held either a university degree ($N = 6$) or a high school diploma ($N = 5$) and the remaining participants had completed vocational training ($N = 1$) or held a secondary school degree or lower ($N = 3$). All participants had at least basic technical skills, as evidenced by their use of an online platform such as Prolific. Based on the Affinity for Technology Interaction (ATI) scale [35], participants had an average score of 4.1 ($SD = 1.35$) (scale range from (1) to (6)). Compared to previous studies conducted by [35], this measure indicates a relatively high technical affinity, falling just below their highest-performing study.

4.1.4 Ethics. The think-aloud study, just as the interviews conducted to inform the UI development, complied with the requirements of our university's ethics committee, including ensuring the anonymity of participants and minimising distress or harm. We collected limited personal information (age, gender, education) and did not collect sensitive data (e.g. ethnicity, religion, health). Participants gave informed consent and were able to withdraw from the study at any time. Questionnaire data was collected on the platform *SoSci Survey*, whose servers are located in Germany and who store the data in accordance with the GDPR[9]. All study data, including transcripts, were secured and processed on university servers in

[8] app.prolific.com/

[9] https://www.soscisurvey.de/en/privacy/

accordance with GDPR data protection regulations. All participants were paid an hourly rate of €14.

4.1.5 Analysis. All sessions were recorded and locally transcribed using using *Whisper*[10], followed by thorough manual editing. Participants' responses were anonymised. The transcripts were then manually annotated by two researchers using *MAXQDA*[11]. We used thematic analysis, a method commonly used in such studies [17, 41], to identify and interpret patterns within qualitative data [17]. Major themes were identified using a mix of pre-established codes that were determined prior to analysis and additional codes that were added iteratively as new themes emerged during the coding process. Pre-defined codes included the main first-level target variables of user interaction, comprehensibility, functionality and usefulness with subcategories for corresponding arbitrary positive, neutral and negative ratings. Specific instances of these ratings, along with additional themes, were added iteratively during the coding process. Two independent coders used the codebook (see Section C.4) to analyse the data, with disagreements resolved by discussion to reach consensus. Inter-coder reliability, measured by Cohen's kappa, indicated substantial agreement ($k = 0.63$).

4.2 Results

Based on the responses to the System Usability Scale (SUS) [18] administered after the think-aloud sessions, the participants were generally satisfied with the usability of ChartChecker. The tool achieved an average SUS score of 83, indicating good overall usability. However, some participants highlighted the need for a more comprehensive introduction to the system. The evaluation of the think-aloud procedure and the subsequent interview of each participant provides a more detailed understanding of these user perceptions. We systematise our findings based on the categories *User Interaction, Comprehensibility, Functionality* and *Usefulness* established in subsection 4.1. Our core contributions and findings are further summarised in Table 3.

4.2.1 User Interaction.

Interaction With UI Elements. Participants demonstrated general confidence in terms of interaction with the tool, and six of them even initiated the automatic analysis independently without the moderator's standard guidance. The *share* function was well received by all seven participants who interacted with it, though some participants expected it to enable sharing results on social media. In fact, however, the share function generates an image of the analysis UI for the user to copy or download manually. The *help* function was very clear to the participants that interacted with it, they understood its purpose and finished the tutorial correctly as intended. Yet, we noted that only 40% ($N = 6$) of participants took note of the function, which potentially affected participants' ability to understand the purpose of other UI parts. The most commonly misunderstood part of the UI was the *show all detectable features* button and the amount of text associated with it. (Figure 5). Multiple participants wrongly mistook this for a more detailed list of misleading features the tool found in the chart. However, the list actually contained all potentially misleading features that ChartChecker can detect. Similarly, the *hide* buttons that hide specific misleading feature corrections from the generated chart were rarely used.

When asked for additional feedback, multiple participants highlighted their appreciation for the colour scheme and simplicity of the UI. Some also commended the straightforwardness of the tools workflow: *"I found the usability and user interface good. Tools that could be used on this page were highlighted in color and yes, I would say that it was generally quite simple."* (P7). Two participants critically remarked, however, that they thought this may only have been the case because of the introduction they were given in the beginning, reinforcing the need for a thorough tutorial.

Generated Chart. The newly generated charts without misleading features received positive feedback, highlighting the simplicity *"Yes, I also find it more pleasant in terms of colour, [...]."* (P5 about Figure 8b) and the displaying of the data *"[...] it is definitely better and at least displays the data much, much more accurately."* (P5 about Figure 10b). Participants particularly appreciated the direct comparison between the original chart and the revised version without the misleading features, noting that it provided a clearer and more accurate presentation of the information.

> *"Exactly, and in the first diagram, there is the difference between the approval between party A and the other parties. It looks very large, but in this proposed diagram you can see that, yes, there is already significantly more in favour of party A, but it's not as massive a difference as was shown in the original diagram due to the shortened scale."* (P2)

Some more ambiguous feedback was due to misunderstandings caused by the misleading nature of the original chart itself, which initially created confusion about the new chart. However, once clarified, participants showed a clear preference for the new chart without misleading features.

> *"So I'm, okay, I'm completely confused by this chart now. Oh, that was zero at the top. Ah, who does that? That's stupid. Okay, so I'm revising my judgement, overall, the right [newly generated chart without misleading features] is better."* (P9)

In summary, the tool's interface allows for intuitive user interaction, as reflected by the participants' confidence in interacting with most of its features and positive feedback on the simplicity of the workflow. The findings emphasise the need for improved tutorial guidance and addressing specific challenges such as the "show all detectable features" page.

4.2.2 Comprehensibility and Functionality.

Misleading Feature Descriptions. Since ChartChecker aims to help users identify misleading features in charts, our study carefully monitored signs of confusion and misinterpretation. Thereby, our observations centred on perceptions of the misleading feature descriptions and the comprehension of the chart itself.

Overall, the misleading feature descriptions were largely understood correctly (48 coded instances) and often made users further analyse the original and recommended charts, facilitating desired critical reflections. The majority of the (few) instances of not understanding misleading features could be attributed to the term

[10]https://github.com/openai/whisper
[11]maxqda.com

"linear" with which some participants were not familiar with and which in one case even led to loss of interest due to the complexity: *"Non-linear y-axis, what does that mean? Non-linear scale? Ah, okay. Well, that seems too complicated for my technical understanding. That's okay, okay, yes, not interested."* (P2) Besides, further enquiries revealed that comprehensibility was negatively affected by the amount of text displayed in the descriptions.

The large number of correct understanding was evident for all other misleading features. Even the *nonlinear axes* feature was well understood by many participants: *"Ah yes, okay, right, because in the original, the distance between 0 and 1 is the same as between 1 and 10, that can't be quite right and it's simply corrected so that the curve flattens out a little."* (P12) In addition, other features such as the *truncated axis* feature were well understood and helped to better understand the information within the chart: *"[...] the y-axis, okay, yes, is shortened [...] You should display from 0 to 100, this can be misleading, as the differences to those displayed appear larger than they actually are."* (P1)

Overall feedback highlighted the value of providing transparent information on misleading features in general: *"But especially with the explanations and such, I found it very understandable and I don't normally find it easy to deal with such issues."* (P2) and *"Even a layman could very quickly see which features were identified. I really liked that."* (P15). On 32 occasions, participants confirmed that ChartChecker identified features that they considered *relevant* to a better understanding of the original charts. However, while 80% of participants were overall satisfied with the functionality, there were also some participants who disagreed with the tool's misleading feature detection. Three participants did not perceive the misleading designs of the original charts as problematic and found the tool's recommendations not helpful or even less accurate: *"And yet I have to say that I can judge the individual price fluctuations per month better with the original diagram than with the proposed ones. It seems a little less accurate to me [...]."* (P15).

Chart Comprehension. After the participants ran the automatic analysis and operated the chart comparison interface, we asked them how their interpretation of data represented in the chart changed. Responses of the think-aloud study can be divided into the three categories *improved*, *unchanged*, and *worsened* understanding. We recorded 50 cases of improved understanding, 15 cases of unchanged understanding, and 8 cases of worsened understanding. The subsequent interview of the participants further emphasised that the tool was successful in improving users' perceived understanding of the diagram, revealing that the majority (87%) liked the diagrams generated. Indeed, the novel recommended chart was the feature that received the most praise during think-aloud sessions. The simplicity compared to the original chart was most valued: *"But compared to the previous diagram, it is much more pleasant to read and interpret, [...]."* (P8). In the cases where participants stated that their understanding of the initially misleading chart did not change, this was mainly due to having recognised all the misleading features before using the tool. However, users generally approved of the tool's ability to annotate misleading features, including as confirmation or support when they had identified misleading features on their own:

Nr.	Suggestions	Count
1	Add titles to generated charts.	9
2	Increase amount of ticks on the axes.	6
3	Add analysis button beside visualisations in the browser.	4
4	Feedback function that allows the user to relay feedback to the developers.	3
5	Add loading animation during the analysis time.	3
6	Display the chart coordinates when hovered over.	3
7	Change the position of the misleading feature descriptions.	2
8	Create additional context with automatic research.	2
9	Add horizontal lines in the background of the chart.	2

Table 2: Every reoccurring feature request for future improvements in descending order of frequency.

> *"I already noticed that the titles were missing, but now that I've seen the misleading features I was confirmed once again."* (P7)

Worsened understanding can mainly be explained by the loss of information within the chart generation process

> *"So I noticed that I find the proposed diagram a bit more confusing because I found it easier to see on the left-hand side that the x-axis is about years. And because the heading is missing [...]. But there's really nothing on the right. "* (P6)

This emphasises the importance of an improved OCR method that achieves higher reliability when extracting the textual data contained within the chart. Particularly regarding Figure 9a, two participants stated that a few inaccuracies in the extracted data made them lose the trust they had placed in the tool.

In the last stage of our study, we asked the participants whether they had any suggestions on how to improve the tool or what features they thought were missing. Table 2 lists the nine most frequently proposed improvements for future work, including, for instance, adding titles to the charts and increasing the amount of ticks on the axes.

In summary, looking at the comprehensibility and functionality, the study achieved promising results, particularly with a good understanding of most of the misleading feature descriptions that facilitate critical reflections of the charts' content and users valuing the simplified, annotated charts. However, complex terms (e.g., non-linear), verbose text and missing titles occasionally caused confusion and limited trust in the tool.

4.2.3 Usefulness. The impact of ChartChecker is arguably most sustainable when it supports users to improve their own visualisation literacy and enables them to debunk misleading charts autonomously. In the course of our user study we noticed five instances where users stated that they recognised previously annotated misleading features in new charts. Some participants reverted to potentially misleading features which had been identified in a prior analysis for assessing novel stimuli: *"[...] the X-Axis also looked good, in linear intervals with labels, so you can't say anything."* (P14).

In addition, a key aim is to clarify misunderstandings and help correct previous misconceptions about charts. Using ChartChecker, 28 cases of successful corrections were identified and coded, for example regarding apparently huge price fluctuations through a non-linear y axis:

> *"I hadn't really noticed that before, but now that I see the hint and look at it again, I didn't notice before that the gap between $2.50 and $3 is relatively small, the gap between $3 and $3.25 is huge and that's why the fluctuations look like that."* (P6)

Similarly, for the truncated x-axis on party approval depicted in bar charts (Figure 10c), ChartChecker helped to interpret the information more accurately.: *"Above all, party A's lead no longer looks as huge as that of parties C and B. Ah, if you look closely, that's actually true. It's misleading because it's actually only 54 per cent to 62 per cent."* (P9)

Overall, the perceived usefulness of the tool was rated positively in most cases. One participant summarised their experience: *"It definitely helped me to understand the diagrams better, to look at them from a different angle or perspective and yes, I would have said it was very helpful."* (P7). When asked whether they would use a version of the ChartChecker tool in the future, 93% of our participants stated that they would be interested. On the one hand, a benefit is to identify misleading features to protect themselves from being misled: *"Yes, that might be useful if you want to practice and familiarise yourself with how to effectively view and compare such charts and avoid or identify pitfalls."* (P2). Moreover, the same amount of participants (93%) stated that they would recommend the tool to family and friends, although some with more reluctance: *"Difficult. Overall, yes, but you need a certain basic understanding of how to read a diagram and what is actually shown. You also need an awareness of the fact that statistics are not perfect, of course, but always have flaws in terms of who took part in the survey."* (P9). This feedback highlights the need for more user guidance, particularly for those with lower levels of visualisation literacy.

In summary, regarding the usefulness of ChartChecker, the study demonstrates promising qualitative results, with participants applying gained insights on misleading features from prior analyses to novel charts, users successfully clarifying misconceptions of misleading charts when confronted with the tool, and positive reporting of perceived helpfulness and envisioned benefits. Better user guidance was emphasised as necessary for future improvements.

5 Discussion

In this study, we developed and evaluated ChartChecker as an approach that supports users in dealing with potentially misleading information in charts. Our contributions include (C1) enabling a user-centred approach to transparently support users in dealing with misleading information in charts and (C2) the technical advancement and evaluation of feasibility of an underlying automatic detection of misleading charts (see Table 3). With these contributions, this work builds on previous research that has primarily focused on data extraction approaches [74, 75] and has only recently begun to address how users can identify deception in line charts [32]. In particular, previous work has not incorporated participatory design or in-depth qualitative think-aloud evaluations to thoroughly explore user perspectives. For example, ChartOCR was designed to extract raw data from visualisations [74], while WebPlotDigitizer provides an interface to view and potentially correct extracted data [75]. However, these tools do not support users in making sense of the raw data, nor do they highlight potential deceptions – an essential goal of ChartChecker. Finally, while the valuable work of Fan et al. [32] points to deception in line charts, it lacks transparent explanations of misleading features in a user-centred interface and does not include participatory design and think-aloud evaluation to address users' needs for comprehensibility.

Technical enhancements building on prior work [32] focused on enabling fast data extraction and extending coverage to a wider range of chart types, including bar charts, as well as automatically detecting several relevant misleading features, such as non-linear axis scales (see Section 5.1). On the user-centred side, a participatory design process guided the development of the user interface, ensuring a transparent and comprehensible presentation of misleading features. ChartChecker was thoroughly evaluated in a think-aloud study, assessing its interface, functionality, comprehensibility and perceived usefulness (see Section 5.2). Based on the insights from the participatory design approach and the think-aloud study, we derive overarching design implications to guide the development of user-centred countermeasures for misleading information in the visual domain (see Table 4).

5.1 How Can Existing Approaches Be Improved to Enhance the Automatic Detection of Misleading Features in Charts?

Our work extends previous work on the detection of misleading features in charts. On the one hand, while previous approaches focused primarily on a narrower set of misleading features (e.g., truncated or inverted axes) [32], we extended this by adding several critical new dimensions, such as non-linear axis scales, inconsistent tick intervals, and multiple axes, and missing labels. These extensions build on previous research which suggested that these features were both relevant and common [73] and thus improve the detection rate to cover around 50% of the most common misinforming features, a notable increase on the 19.5% originally enabled by Fan et al. [32]. In addition, building on previous work on the combination of machine learning and rule-based methods for extracting chart data, the *ChartOCR* [74] approach was used to facilitate fast data extraction from both line and bar charts. This development extended the detection of misleading features to bar charts, a key and widely used chart type, building on previous work focused on line charts [32]. This approach also facilitated the extraction of text rotated by 90°, as is common for y-axis labels. However, challenges remained in achieving fully accurate extraction of all relevant chart data. While key point extraction performed well for bar charts, the accuracy for line charts still had room for improvement. While the primary contribution of this work is the user-centred approach using participatory design and a think-aloud study, our aim here was to develop and assess the feasibility of a conceptual technical solution. Rather than relying on full simulations or assuming an

Core Contributions	Core Findings
C1: Enabling a user-centred approach to transparently support users in navigating misleading information in charts	F1: A participatory design approach for the UI highlighted the desire for a direct comparison between the original and recommended charts, along with a transparent list of detected misleading features. F2: ChartChecker was well-received, with users expressing high satisfaction in both the think-aloud study and SUS scores. F3: Users praised ChartChecker's transparent explanations and presentation of potentially misleading features as key reasons for their positive reception. F4: ChartChecker has the potential to correct user misunderstandings and support the transfer of insights to new, unseen charts. F5: ChartChecker can provide guidance but not the absolute truth, and at times, participants preferred the original, potentially misleading chart over the recommended version.
C2: Advancing the automated detection of misleading charts [32]	F6: The detection of misleading features was expanded from three to seven, allowing for a more comprehensive analysis of charts. F7: A novel OCR approach to extract chart information automatically was implemented. F8: The detection was expanded to incorporate bar charts, effectively broadening the range of relevant chart types.

Table 3: Core contributions and findings of the user-centredness and technical development of the tool.

automatic detection process, as is common in misinformation studies [43, 44], we aimed to test how the technical foundation of our user-centred interface works could work in practice.

5.2 How Can a User-Centred, Indicator-Based Approach Support Users in Navigating Misleading Information in Charts?

The primary contribution of this work lies in its user-centred approach, providing transparent, indicator-based guidance on potentially misleading representations in charts. Previous research has highlighted the importance of improving users' ability to identify misleading visualisations through user-centred interventions [22]. Therefore, this indicator-based method, which transparently highlights potentially misleading information, supports visual literacy in a way that goes beyond previous efforts that have primarily focused on the text-based domain [44].

The participatory design approach used to develop a user-centred interface resulted in a UI that users generally found intuitive and helpful. This was expressed both in qualitative statements and SUS scores. The feature of **enabling a direct comparison between the misleading and non-misleading chart side by side** (design implication (1), Table 4) in a clear manner was highly praised, thus mirroring previous findings from Wijnker et al. [116] on the importance of direct visual comparisons to facilitate understanding of misleading information. Correcting misinformation is generally challenging, as repeating misleading information can reinforce belief in it [33, 93]. It is, therefore, generally unclear to what extent the focus should be on the original, potentially misleading chart. However, in our think-aloud study, participants explicitly valued this comparison as it helped them identify specific misleading tactics.

Transparency is crucial for building user trust and combating misinformation. Participants valued the use of transparent indicators to highlight potential deception, both visually with a recommended chart without deception and through text describing the specific potentially deceptive features contained in the chart. Such **prioritisation of transparency and explainability** (design implication (2), Table 4), therefore, also emerged as an overarching design implication when supporting users in navigating misleading charts. Similar positive evaluations of transparent, indicator-based explanations have been observed in text-based [44], audio-based [45] and video-based domains [43]. This research extends these findings to the chart domain, further demonstrating the potential of transparency in helping users navigate complex, potentially misleading information. While some indicators, such as emotionality, are transferable between text, audio and video, most indicators in charts are specific and refer to typical deviations from design conventions [116].

While the comprehensibility of the misleading features was generally well-rated, some participants preferred more detailed explanations, while others found the text overwhelming. In addition, a minority of participants struggled with certain terms such as "linearity" when the explanations were intended to point out unusual axes. Individual differences thus emerged, confirming that different groups of users engage with interventions in different ways depending on their prior knowledge and level of expertise. Thus, this research confirms the potential of **personalisation** (design implication (3), Table 4) to improve the efficiency of interventions, as has been previously shown both in the broader area of online privacy and security [14, 58] and in misinformation interventions in mainly text-based domains [13, 51, 52]. Specifically, this could be implemented by allowing users to switch between simplified and detailed descriptions of misleading features based on their level of visualisation literacy. The overall aim of misinformation interventions is to, ideally, improve both short-term understanding and long-term literacy. In the think-aloud studies, some participants demonstrated a transfer of knowledge gained through ChartChecker's annotations to novel stimuli, suggesting a potential for improving visualisation literacy. However, as this transfer was not evident for all participants, explanations should ideally be tailored to individual needs in order to effectively improve visualisation literacy for all.

Finally, while most participants found ChartChecker useful, some raised a key concern. The tool can only detect design violations based on standard conventions, but under standard conventions may not always best convey data meaning. In such cases, departures from these conventions may enhance the clarity and communicative power of the data. For example, Correll et al. [30] discuss how truncating y-axes can introduce bias, but in some cases, avoiding truncation might be even more misleading. Cairo [21] argues that baselines should be *"logical and meaningful"* rather than fixed at zero in all graphs. ChartChecker does not currently account for these contextual factors, which may occasionally result in detections that confuse rather than assist users. This highlights an important limitation, as ChartChecker cannot take into account the nuances of chart contexts or the potential intent behind certain design choices, as this would represent its own set of complex challenges. For this reason, some users actually preferred the original charts over the recommended ones in our study. This issue raises an important design implication: while automated tools can assist in detecting design flaws, it remains essential to **balance automation with human judgement** (design implication (4), Table 4) and interventions should be designed to enable this. As the 'neutral' presentation of data is sometimes seen as a myth [39], it remains important to present interventions as guidance rather than definitive corrections to enable informed and nuanced engagement with information in charts. ChartChecker has attempted to do this by not indicating that the original chart is "wrong" or "misleading" per se, but that it presents a "recommended" chart without departing from typical design conventions.

5.3 Limitations and Future Work

First and foremost, the empirical results of this study provide qualitative insights based on a smaller group of participants from a developed Western European country. Representativeness is not the aim of such qualitative research, and therefore, it should be acknowledged that the conclusions drawn here may differ with a different subset of users. In particular, our work involved younger people with higher education backgrounds as part of the participatory design approach. In addition, the think-aloud study included participants with at least some technical affinity, as they were recruited via a digital platform such as Prolific. The results also showed a relatively high affinity for technical interaction, suggesting a gap for those with lower general skills and comfort in this area. This was further highlighted by the fact that while most participants found the explanations of misleading features useful and understandable, a subset struggled with terms such as 'non-linearity'. Therefore, future work is needed to assess the transferability of these findings to other user groups. In addition, the smaller sample sizes of the participatory design approach and the think-aloud evaluation limit the generalisability of the results. While the sample sizes are not uncommon in qualitative HCI research [20, 82, 111], larger and more diverse samples would enable a broader applicability of the tool. However, given the lack of user-centred evidence in the area of supporting understanding of potentially misleading information in charts, this study deliberately prioritised depth and a more narrowly focused sample at an earlier stage of research in this area.

Moreover, the qualitative approach and reliance on self-reporting was intended to provide rich insights into user perceptions and engagement with the various target variables, including user interaction, comprehension, functionality and perceived usefulness. These insights are essential for the designing effective tools. In the process, we also gained valuable qualitative insights into how ChartChecker can improve the understanding of charts and the transfer of knowledge to novel content. To focus the effectiveness of ChartChecker, future research could incorporate quantitative formal pre- and post-intervention assessments to evaluate changes in chart literacy before and after tool usage, or experimental comparisons between a control group assessing misleading charts without ChartChecker and an experimental group using it. This could be done using the Visual Literacy Assessment Test, a 53-item questionnaire that measures visualisation reading skills [64].

Furthermore, while the current approach has improved automated data extraction and extended coverage to bar charts challenges remain in achieving fully accurate data extraction. In particular, the wide range of design choices made when creating axis designs and scales makes it a challenge to reliably assign correct values to chart axes, as their placement, orientation, and formatting can vary from chart to chart. There is, therefore, potential to further improve the accuracy of extraction and to extend the detection of misleading features commonly found in pie charts, another commonly used chart type.

Finally, it is important to recognise that the spread of misinformation through charts goes beyond the design violations identified in this paper. Lisnic et al. [69] highlight that misleading charts can include a wide range of other deceptive tactics that manipulate the chart or data in ways that ChartChecker currently cannot detect. These include factors such as the reliability of the source, the tone or polarity of the accompanying text, and inferential errors such as cherry-picking data. Furthermore, there may be contexts in which departures from certain design standards are reasonable and do not represent an intent to deceive. Therefore, future work that combines the detection of design violations with deeper analysis of background and contextual information has significant potential to more comprehensively address misinformation in charts.

6 Conclusion

Our work highlights the potential of a user-centred approach that transparently supports users in dealing with misleading information in the visual domain. The contributions and results include (C1) developing and evaluating a user-centred misinformation intervention for misleading charts, where we (F1) found that users preferred a direct comparison between the original and recommended charts with a transparent list of misleading features. ChartChecker received positive feedback in our think-aloud study, with (F2) high user satisfaction and (F3) praise for its transparent explanations. Importantly, (F3) the tool demonstrated the potential to correct misunderstandings and support knowledge transfer to new, unseen charts, although (F5) it cannot claim to offer the absolute truth, as some users preferred the original, potentially misleading charts. To

Design Implication	Explanation
(1) Enable direct comparisons between misleading and non-misleading charts	Provide users with a side-by-side comparison of misleading and non-misleading charts to help them understand the differences and how misleading features distort information to enhance visual literacy.
(2) Prioritise transparency and explainability	Clearly explain detected misleading features and any corrections made to charts. Transparent explanations build user confidence, ensuring they understand the intervention and the reasoning behind any chart modifications.
(3) Allow for personalisation	Design interventions that cater for different user groups by providing different levels of information detail. This helps different users to engage effectively with an intervention, based on their individual level of expertise.
(4) Balance automation with human judgement	While automating chart analysis increases efficiency, users should be encouraged to critically evaluate an intervention. Present intervention suggestions as guidance rather than definitive truth to maintain user agency.

Table 4: Design implications for user-centred strategies in dealing with misleading information in charts.

not fully rely on simulations but rather test the feasibility of an underlying technical foundation, we (C2) advanced the automatic detection of misleading charts by (F6) extending the detection range of related work from three to seven features, (F7) implementing a fast OCR data extraction process (<1.4 seconds), and (F8) incorporating bar charts besides line charts to extend the scope. On a broader level, our findings highlight the balance between providing transparent information about potentially misleading visualisations of information in charts and encouraging users to reflect for themselves. While ChartChecker provides transparency, the challenge remains to accommodate users' prior knowledge and preferences for certain visualisations, even when they might be considered misleading. This highlights the importance of further research into how users engage with supporting interventions and related limitations when dealing with deception in the visual domain.

Acknowledgments

This work was supported by the German Federal Ministry for Education and Research (BMBF) in the project NEBULA (13N16361), and by the German Federal Ministry of Education and Research and the Hessian Ministry of Higher Education, Research, Science and the Arts within their joint support of the National Research Center for Applied Cybersecurity ATHENE.

References

[1] Troy L. Adams, Yuanxia Li, and Hao Liu. 2020. A Replication of Beyond the Turk: Alternative Platforms for Crowdsourcing Behavioral Research – Sometimes Preferable to Student Groups. *AIS Transactions on Replication Research* 6 (10 2020), 15. Issue 1. doi:10.17705/1atrr.00058

[2] Rabah A. Al-Zaidy and C. Lee Giles. 2015. Automatic Extraction of Data from Bar Charts. In *Proceedings of the 8th International Conference on Knowledge Capture*, Ken Barker and José Manuel Gómez-Pérez (Eds.). ACM, New York, NY, 1–4. doi:10.1145/2815833.2816956

[3] Obead Alhadreti and Pam Mayhew. 2016. "Thinking about Thinking Aloud": An Investigation of Think-Aloud Methods in Usability Testing. In *BCS Learning & Development*, Vol. 1. BCS Learning & Development, Bournemouth, 1–3. doi:10.14236/ewic/hci2016.101

[4] Malik Almaliki. 2019. Online Misinformation Spread: A Systematic Literature Map. In *Proceedings of the 2019 3rd International Conference on Information System and Data Mining (ICISDM 2019)*. Association for Computing Machinery, New York, NY, USA, 171–178. doi:10.1145/3325917.3325938

[5] Tiago Araújo, Paulo Chagas, João Alves, Carlos Santos, Beatriz Sousa Santos, and Bianchi Serique Meiguins. 2020. A Real-World Approach on the Problem of Chart Recognition Using Classification, Detection and Perspective Correction. *Sensors* 20, 16 (2020), 4370. doi:10.3390/s20164370

[6] Alberto Ardevol-Abreu, Patricia Delponti, and Carmen Rodriguez-Wanguemert. 2020. Intentional or Inadvertent Fake News Sharing? Fact-checking Warnings and Users' Interaction with Social Media Content. *Profesional de la Informacion* 29, 5 (2020), 1–13. doi:10.3145/epi.2020.sep.07

[7] Ranojoy Barua, Rajdeep Maity, Dipankar Minj, Tarang Barua, and Ashish Kumar Layek. 2019. F-NAD: An Application for Fake News Article Detection Using Machine Learning Techniques. In *2019 IEEE Bombay Section Signature Conference (IBSSC)*. Institute of Electrical and Electronics Engineers, Mumbai, 1–6. doi:10.1109/IBSSC47189.2019.8973059

[8] Leilani Battle, Peitong Duan, Zachery Miranda, Dana Mukusheva, Remco Chang, and Michael Stonebraker. 2018. Beagle: Automated Extraction and Interpretation of Visualizations from the Web. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–8. doi:10.1145/3173574.3174168

[9] Vivien Beattie and Michael John Jones. 2002. The Impact of Graph Slope on Rate of Change Judgments in Corporate Reports. *Abacus-a Journal of Accounting Finance and Business Studies* 38, 2 (2002), 177–199. doi:10.1111/1467-6281.00104

[10] David K. Berlo. 1960. *The Process of Communication; an Introduction to Theory and Practice.* New York, Holt, Rinehart and Winston, New York. http://archive.org/details/processofcommuni0000berl

[11] Puneet Bhargava, Katie MacDonald, Christie Newton, Hause Lin, and Gordon Pennycook. 2023. How Effective Are TikTok Misinformation Debunking Videos? *Harvard Kennedy School Misinformation Review* 4, 2 (2023), 1–17. doi:10.37016/mr-2020-114

[12] MD Momen Bhuiyan, Hayden Whitley, Michael Horning, Sang Won Lee, and Tanushree Mitra. 2021. Designing Transparency Cues in Online News Platforms to Promote Trust: Journalists' and Consumers' Perspectives. *Proceedings of the ACM: Human Computer Interaction (PACM): Computer-Supported Cooperative Work and Social Computing* 5, CSCW2 (2021), 395:1–395:31. doi:10.1145/3479539

[13] Tom Biselli, Katrin Hartwig, and Christian Reuter. 2025. Mitigating Misinformation Sharing on Social Media through Personalised Nudging. *Proceedings of the ACM: Human Computer Interaction (PACM): Computer-Supported Cooperative Work and Social Computing* 9 (2025).

[14] Tom Biselli, Laura Utz, and Christian Reuter. 2024. Supporting Informed Choices about Browser Cookies: The Impact of Personalised Cookie Banners. *Proceedings on Privacy Enhancing Technologies (PoPETs)* 2024, 1 (2024), 171–191. doi:10.56553/popets-2024-0011

[15] T. Boren and J. Ramey. 2000. Thinking Aloud: Reconciling Theory and Practice. *IEEE Transactions on Professional Communication* 43, 3 (2000), 261–278. doi:10.1109/47.867942

[16] Jeremy Boy, Ronald A. Rensink, Enrico Bertini, and Jean-Daniel Fekete. 2014. A Principled Way of Assessing Visualization Literacy. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (2014), 1963–1972. doi:10.1109/TVCG.2014.2346984

[17] Virginia Braun, Victoria Clarke, Nikki Hayfield, and Gareth Terry. 2019. Thematic Analysis. In *Handbook of Research Methods in Health Social Sciences*, Pranee Liamputtong (Ed.). Springer, Singapore, 843–860. doi:10.1007/978-981-10-5251-4_103

[18] John Brooke. 1995. SUS: A Quick and Dirty Usability Scale. *Usability Evaluation in Industry* 189, 194 (1995), 4–7.

[19] Grégoire Burel, Tracie Farrell, Martino Mensio, Prashant Khare, and Harith Alani. 2020. Co-Spread of Misinformation and Fact-Checking Content during the Covid-19 Pandemic. In *Social Informatics : 12th International Conference, SocInfo 2020*. Springer, Pisa, Italy, 28–42. doi:10.1007/978-3-030-60975-7{_}3

[20] Kelly Caine. 2016. Local Standards for Sample Size at CHI. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, San Jose California USA, 981–992. doi:10.1145/2858036.2858498

[21] Alberto Cairo. 2016. *The Truthful Art : Data, Charts, and Maps for Communication.* New Riders, Indianapolis, USA. https://elibrary.pearson.de/book/99.150005/9780133440553

[22] Jorge D. Camba, Pedro Company, and Vetria L. Byrd. 2022. Identifying Deception as a Critical Component of Visualization Literacy. *IEEE Computer Graphics and Applications* 42, 1 (2022), 116–122. doi:10.1109/MCG.2021.3132004

[23] Qing Chen, Shixiong Cao, Jiazhe Wang, and Nan Cao. 2023. How Does Automation Shape the Process of Narrative Visualization: A Survey of Tools. *IEEE Transactions on Visualization and Computer Graphics* 30, 8 (2023), 1–20. doi:10.1109/tvcg.2023.3261320

[24] Qing Chen, Fuling Sun, Xinyue Xu, Zui Chen, Jiazhe Wang, and Nan Cao. 2022. VizLinter: A Linter and Fixer Framework for Data Visualization. *IEEE Transactions on Visualization and Computer Graphics* 28, 1 (2022), 206–216. https://ieeexplore.ieee.org/document/9552878/?arnumber=9552878

[25] Sijing Chen, Lu Xiao, and Akit Kumar. 2023. Spread of Misinformation on Social Media: What Contributes to It and How to Combat It. *Computers in Human Behavior* 141 (2023), 107643. doi:10.1016/j.chb.2022.107643

[26] Yang Chen, Scott Barlowe, and Jing Yang. 2010. Click2Annotate: Automated Insight Externalization with Rich Semantics. In *2010 IEEE Symposium on Visual Analytics Science and Technology*. IEEE, Salt Lake City, 155–163. doi:10.1109/vast.2010.5652885

[27] Yang Chen, Jing Yang, Scott Barlowe, and Dong H. Jeong. 2010. Touch2Annotate. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems*. ACM, New York, NY, USA, 3703–3708. doi:10.1145/1753846.1754042

[28] Jinho Choi, Sanghun Jung, Deok Gun Park, Jaegul Choo, and Niklas Elmqvist. 2019. Visualizing for the Non–Visual: Enabling the Visually Impaired to Use Visualization. *Computer Graphics Forum* 38, 3 (2019), 249–260. doi:10.1111/cgf.13686

[29] Michael Correll. 2019. Ethical Dimensions of Visualization Research. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3290605.3300418

[30] Michael Correll, Enrico Bertini, and Steven Franconeri. 2020. Truncating the Y-Axis: Threat or Menace?. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3313831.3376222

[31] K. Anders Ericsson and Herbert A. Simon. 1998. How to Study Thinking in Everyday Life: Contrasting Think-Aloud Protocols with Descriptions and Explanations of Thinking. *Mind, Culture, and Activity* 5, 3 (1998), 178–186. doi:10.1207/s15327884mca0503{_}3

[32] Arlen Fan, Yuxin Ma, Michelle Mancenido, and Ross Maciejewski. 2022. Annotating Line Charts for Addressing Deception. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 80, 12 pages. doi:10.1145/3491102.3502138

[33] Lisa K. Fazio, Raunak M. Pillai, and Deep Patel. 2022. The Effects of Repetition on Belief in Naturalistic Settings. *Journal of Experimental Psychology: General* 151, 10 (2022), 2604–2613. doi:10.1037/xge0001211 Place: US Publisher: American Psychological Association.

[34] Renee Fleck. 2021. Dribbble - Discover the World's Top Designers & Creative Professionals. https://dribbble.com/resources/education/ui-design-principles

[35] Thomas Franke, Christiane Attig, and Daniel Wessel. 2019. A Personal Resource for Technology Interaction: Development and Validation of the Affinity for Technology Interaction (ATI) Scale. *International Journal of Human–Computer Interaction* 35, 6 (2019), 456–467. doi:10.1080/10447318.2018.1456150

[36] Isabelle Freiling. 2019. Detecting Misinformation in Online Social Networks: A Think-Aloud Study on User Strategies. *SCM Studies in Communication and Media* 8, 4 (2019), 471–496. doi:10.5771/2192-4007-2019-4-471

[37] Lily W. Ge, Yuan Cui, and Matthew Kay. 2023. CALVI: Critical Thinking Assessment for Literacy in Visualizations. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–18. doi:10.1145/3544548.3581406

[38] Amira Ghenai, Mark Smucker, and Charles Clarke. 2020. A Think-Aloud Study to Understand Factors Affecting Online Health Search. In *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*. Association for Computing Machinery, Vancouver, Canada, 273–282. doi:10.1145/3343413.3377961

[39] Lisa Gitelman. 2013. *'Raw Data' Is an Oxymoron.* MIT Press, Cambridge, Mass.

[40] Xinning Gui, Yubo Kou, Kathleen H. Pine, and Yunan Chen. 2017. Managing Uncertainty: Using Social Media for Risk Assessment during a Public Health Crisis. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, Denver Colorado USA, 4520–4533. doi:10.1145/3025453.3025891

[41] Lee Hadlington, Lydia J. Harkin, Daria Kuss, Kristina Newman, and Francesca C. Ryding. 2023. Perceptions of Fake News, Misinformation, and Disinformation amid the COVID-19 Pandemic: A Qualitative Exploration. *Psychology of Popular Media* 12, 1 (2023), 40–49. doi:10.1037/ppm0000387

[42] Michael Hameleers, Thomas E. Powell, Toni G.L.A. van der Meer, and Lieke Bos. 2020. A Picture Paints a Thousand Lies? The Effects and Mechanisms of Multimodal Disinformation and Rebuttals Disseminated via Social Media. *Political Communication* 37, 2 (2020), 281–301. doi:10.1080/10584609.2019.1674979

[43] Katrin Hartwig, Tom Biselli, Franziska Schneider, and Christian Reuter. 2024. From Adolescents' Eyes: Assessing an Indicator-Based Intervention to Combat Misinformation on TikTok. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–20. doi:10.1145/3613904.3642264

[44] Katrin Hartwig, Frederic Doell, and Christian Reuter. 2024. The Landscape of User-centered Misinformation Interventions – A Systematic Literature Review. *ACM Computing Surveys (CSUR)* 56, 11 (2024), 1–36. doi:10.1145/3674724

[45] Katrin Hartwig, Ruslan Sandler, and Christian Reuter. 2024. Navigating Misinformation in Voice Messages: Identification of User-Centered Features for Digital Interventions. *Risk, Hazards & Crisis in Public Policy* 15, 2 (2024), 203–235. doi:10.1002/rhc3.12296

[46] Aspen K. Hopkins, Michael Correll, and Arvind Satyanarayan. 2020. VisuaLint: Sketchy In Situ Annotations of Chart Construction Errors. *Computer Graphics Forum* 39, 3 (2020), 219–228. doi:10.1111/cgf.13975

[47] Darrell Huff. 1991. *How to Lie with Statistics.* Penguin, London.

[48] Jessica Hullman, Nicholas Diakopoulos, and Eytan Adar. 2013. Contextifier. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (ACM Digital Library)*, Wendy E. Mackay (Ed.). ACM, New York, NY, 2707–2716. doi:10.1145/2470654.2481374

[49] Andrew Iliadis and Federica Russo. 2016-12-01, 2016. Critical Data Studies: An Introduction. *Big Data & Society* 3, 2 (2016-12-01, 2016), 2053951716674238. doi:10.1177/2053951716674238

[50] Md Rafiqul Islam, Shaowu Liu, Xianzhi Wang, and Guandong Xu. 2020. Deep Learning for Misinformation Detection on Online Social Networks: A Survey and New Perspectives. *Social Network Analysis and Mining* 10, 1 (2020), 82.

[51] Farnaz Jahanbakhsh, Amy X. Zhang, Adam J. Berinsky, Gordon Pennycook, David G. Rand, and David R. Karger. 2021. Exploring Lightweight Interventions at Posting Time to Reduce the Sharing of Misinformation on Social Media. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (April 2021), 18:1–18:42. doi:10.1145/3449092

[52] Farnaz Jahanbakhsh, Amy X. Zhang, and David R. Karger. 2022. Leveraging Structured Trusted-Peer Assessments to Combat Misinformation. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (Nov. 2022), 524:1–524:40. doi:10.1145/3555637

[53] Anjali Jain, Avinash Shakya, Harsh Khatter, and Amit Kumar Gupta. 2019. A Smart System for Fake News Detection Using Machine Learning. In *2019 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT)*, Vol. 1. IEEE, Ghaziabad, 1–4. doi:10.1109/ICICT46931.2019.8977659

[54] Cristhy Jimenez, Pablo Lozada, and Pablo Rosas. 2016. Usability Heuristics: A Systematic Review. In *2016 IEEE 11th Colombian Computing Conference (CCC)*. Institute of Electrical and Electronics Engineers Inc., United States, 1–8. https://ieeexplore.ieee.org/document/7750805/?arnumber=7750805

[55] Shankar Kantharaj, Rixie Tiffany Ko Leong, Xiang Lin, Ahmed Masry, Megh Thakkar, Enamul Hoque, and Shafiq Joty. 2022. Chart-to-Text: A Large-Scale Benchmark for Chart Summarization. https://arxiv.org/abs/2203.06486

[56] Jan Kirchner and Christian Reuter. 2020. Countering Fake News: A Comparison of Possible Solutions Regarding User Acceptance and Effectiveness. *Proceedings of the ACM: Human Computer Interaction (PACM): Computer-Supported Cooperative Work and Social Computing* 4, CSCW2 (2020), 1–27. doi:10.1145/3415211

[57] Rob Kitchin, Martin Dodge, and Chris Perkins. 2011. Introductory Essay: Power and Politics of Mapping. *The map reader: Theories of mapping practice and cartographic representation* 1 (2011), 387–394. doi:10.1002/9780470979587.ch50

[58] Bart P. Knijnenburg. 2017. Privacy? I Can't Even! Making a Case for User-Tailored Privacy. *IEEE Security & Privacy* 15, 4 (2017), 62–67. doi:10.1109/MSP.2017.3151331

[59] Ha-Kyung Kong, Zhicheng Liu, and Karrie Karahalios. 2019. Trust and Recall of Information across Varying Degrees of Title-Visualization Misalignment. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland Uk, 1–13. doi:10.1145/3290605.3300576

[60] Krahmer. 2004. Thinking about Thinking Aloud: A Comparison of TwoVerbal Protocols for Usability Testing. *IEEE Transactions on Professional Communication* 47, 2 (2004), 105–117. doi:10.1109/TPC.2004.828205

[61] Claire Lauer and Shaun O'Brien. 2020. How People Are Influenced by Deceptive Tactics in Everyday Charts and Graphs. *IEEE Transactions on Professional Communication* 63, 4 (2020), 327–340. doi:10.1109/tpc.2020.3032053

[62] Hei Law and Jia Deng. 2018. CornerNet: Detecting Objects as Paired Keypoints. In *Proceedings of the European Conference on Computer Vision (ECCV) (Lecture Notes in Computer Science)*, Vittorio Ferrari (Ed.). Springer, Cham, 765–781. doi:10.48550/arXiv.1808.01244

[63] Crystal Lee, Tanya Yang, Gabrielle D Inchoco, Graham M. Jones, and Arvind Satyanarayan. 2021. Viral Visualizations: How Coronavirus Skeptics Use Orthodox Data Practices to Promote Unorthodox Science Online. In *Proceedings of the*

2021 CHI Conference on Human Factors in Computing Systems. ACM, Yokohama Japan, 1–18. doi:10.1145/3411764.3445211

[64] Sukwon Lee, Sung-Hee Kim, and Bum Chul Kwon. 2017. VLAT: Development of a Visualization Literacy Assessment Test. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2017), 551–560. doi:10.1109/TVCG.2016.2598920

[65] Clayton Lewis. 1982. *Using he Thinking-aloud Method in Cognitive Interface Design*. IBM TJ Watson Research Center, NY.

[66] Shufan Li, Congxi Lu, Linkai Li, and Haoshuai Zhou. 2022. Chart-RCNN: Efficient Line Chart Data Extraction from Camera Images. https://arxiv.org/pdf/2211.14362

[67] Yang-Jun Li, Jens Joachim Marga, Christy M.K. Cheung, Xiao-Liang Shen, and Matthew Lee. 2022. Health Misinformation on Social Media: A Systematic Literature Review and Future Research Directions. *AIS Transactions on Human-Computer Interaction* 14, 2 (2022), 116–149. doi:10.17705/1thci.00164

[68] Gionnieve Lim and Simon Tangi Perrault. 2021. Local Perceptions and Practices of News Sharing and Fake News. In *Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing*. ACM, Virtual Event USA, 117–120. doi:10.1145/3462204.3481767

[69] Maxim Lisnic, Cole Polychronis, Alexander Lex, and Marina Kogan. 2023. Misleading beyond Visual Tricks: How People Actually Lie with Charts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, Albrecht Schmidt, Kaisa Väänänen, Tesh Goyal, Per Ola Kristensson, Anicia Peters, Stefanie Mueller, Julie R. Williamson, and Max L. Wilson (Eds.). ACM, New York, NY, USA, 1–21. doi:10.1145/3544548.3580910

[70] Can Liu, Liwenhan Xie, Yun Han, Datong Wei, and Xiaoru Yuan. 2020. AutoCaption: An Approach to Generate Natural Language Description from Visualization Automatically. In *2020 IEEE Pacific Visualization Symposium (PacificVis)*. IEEE, Tianjin, 191–195. doi:10.1109/pacificvis48177.2020.1043

[71] Xiaoyi Liu, Diego Klabjan, and Patrick NBless. 2019. Data Extraction from Charts via Single Deep Neural Network. https://arxiv.org/abs/1906.11906

[72] Leo Yu-Ho Lo, Yifan Cao, Leni Yang, and Huamin Qu. 2024. Why Change My Design: Explaining Poorly Constructed Visualization Designs with Explorable Explanations. *IEEE Transactions on Visualization and Computer Graphics* 30, 1 (2024), 955–964. doi:10.1109/tvcg.2023.3327155

[73] Leo Yu-Ho Lo, Ayush Gupta, Kento Shigyo, Aoyu Wu, Enrico Bertini, and Huamin Qu. 2022. Misinformed by Visualization: What Do We Learn from Misinformative Visualizations? *Computer Graphics Forum* 41, 3 (2022), 515–525. doi:10.1111/cgf.14559

[74] Junyu Luo, Zekun Li, Jinpeng Wang, and Chin-Yew Lin. 2021. ChartOCR: Data Extraction from Charts Images via a Deep Hybrid Framework. In *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, Waikoloa, 1916–1924. doi:10.1109/wacv48630.2021.00196

[75] F. Marin, A. Rohatgi, and S. Charlot. 2017. WebPlotDigitizer, a Polyvalent and Free Software to Extract Spectra from Old Astronomical Publications: Application to Ultraviolet Spectropolarimetry. http://arxiv.org/pdf/1708.02025v1

[76] Alexandre Maros, Anastasia Giachanou, Viktoria Spaiser, Francesca Spezzano, Anna George, Alexandra Pavliuc, Alexandre Bright, Jonathan, Jussara M. Almeida, and Marisa Vasconcelos. 2021. A Study of Misinformation in Audio Messages Shared in WhatsApp Groups. In *Disinformation in Open Online Media*, Vol. 12887. Springer International Publishing, Cham, 85–100. doi:10.1007/978-3-030-87031-7_6

[77] Giovanni Da San Martino, Shaden Shaar, Yifan Zhang, Seunghak Yu, Alberto Barrón-Cedeño, and Preslav Nakov. 2020. Prta: A System to Support the Analysis of Propaganda Techniques in the News. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. Association for Computational Linguistics, Online, 287–293.

[78] Andrew McNutt, Gordon Kindlmann, and Michael Correll. 2020. Surfacing Visualization Mirages. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–16. doi:10.1145/3313831.3376420

[79] Shubha Mishra, Piyush Shukla, and Ratish Agarwal. 2022. Analyzing Machine Learning Enabled Fake News Detection Techniques for Diversified Datasets. *Wireless Communications and Mobile Computing* 2022 (2022), 1–18. doi:10.1155/2022/1575365

[80] Sina Mohseni, Fan Yang, Shiva Pentyala, Mengnan Du, Yi Liu, Nic Lupfer, Xia Hu, Shuiwang Ji, and Eric Ragan. 2021. Machine Learning Explanations to Prevent Overtrust in Fake News Detection. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 15. AAAI Press, virtually, 421–431. doi:10.1609/icwsm.v15i1.18072

[81] Veronica Nanni, Enrico Caprio, Giulia Bombieri, Stefano Schiaparelli, Carlo Chiorri, Stefano Mammola, Paolo Pedrini, and Vincenzo Penteriani. 2020. Social Media and Large Carnivores: Sharing Biased News on Attacks on Humans. *Frontiers in Ecology and Evolution* 8 (2020), 486886. doi:10.3389/fevo.2020.00071

[82] Jakob Nielsen. 1994. Estimating the number of subjects needed for a thinking aloud test. *International Journal of Human-Computer Studies* 41, 3 (1994), 385–397. doi:10.1006/ijhc.1994.1065

[83] Shuo Niu, Zhicong Lu, Amy X. Zhang, Jie Cai, Carla F. Griggio, and Hendrik Heuer. 2023. Building Credibility, Trust, and Safety on Video-Sharing Platforms. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, 1–7. doi:10.1145/3544549.3573809

[84] Mie Nørgaard and Kasper Hornbæk. 2006. What Do Usability Evaluators Do in Practice?. In *Proceedings of the 6th Conference on Designing Interactive Systems*. ACM, New York, NY, USA, 209–218. doi:10.1145/1142405.1142439

[85] Duarte N. J. Nunes. 2001. *Object Modeling for User-Centered Development and User Interface Design: The Wisdom Approach*. Ph. D. Dissertation. Universidade da Madeira (Portugal), Madeira, Portugal. https://www.proquest.com/dissertations-theses/object-modeling-user-centered-development/docview/1894059681/se-2

[86] Brendan Nyhan and Jason Reifler. 2010. When Corrections Fail: The Persistence of Political Misperceptions. *Political Behavior* 32, 2 (2010), 303–330. doi:10.1007/s11109-010-9112-2

[87] Shaun O'Brien and Claire Lauer. 2018. Testing the Susceptibility of Users to Deceptive Data Visualizations When Paired with Explanatory Text. In *Proceedings of the 36th ACM International Conference on the Design of Communication*. ACM, Milwaukee WI USA, 1–8. doi:10.1145/3233756.3233961

[88] Katherine O'Toole. 2023. Collaborative Creativity in TikTok Music Duets. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–16. doi:10.1145/3544548.3581380

[89] Anshul Vikram Pandey, Katharina Rall, Margaret L. Satterthwaite, Oded Nov, and Enrico Bertini. 2015. How Deceptive Are Deceptive Visualizations?: An Empirical Analysis of Common Distortion Techniques. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, Seoul Republic of Korea, 1469–1478. doi:10.1145/2702123.2702608

[90] Irene V. Pasquetto, Eaman Jahani, Shubham Atreja, and Matthew Baum. 2022. Social Debunking of Misinformation on WhatsApp: The Case for Strong and In-group Ties. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW1 (2022), 1–35. doi:10.1145/3512964

[91] Eyal Peer, Laura Brandimarte, Sonam Samat, and Alessandro Acquisti. 2017. Beyond the Turk: Alternative Platforms for Crowdsourcing Behavioral Research. *Journal of Experimental Social Psychology* 70 (2017), 153–163. doi:10.1016/j.jesp.2017.01.006

[92] Gordon Pennycook, Jonathon McPhetres, Yunhao Zhang, Jackson G. Lu, and David G. Rand. 2020. Fighting COVID-19 Misinformation on Social Media: Experimental Evidence for a Scalable Accuracy-Nudge Intervention. *Psychological science* 31, 7 (2020), 770–780. doi:10.1177/0956797620939054

[93] Raunak M. Pillai and Lisa K. Fazio. 2021. The Effects of Repeating False and Misleading Information on Belief. *WIREs Cognitive Science* 12, 6 (2021), e1573. doi:10.1002/wcs.1573 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wcs.1573.

[94] Jorge Poco and Jeffrey Heer. 2017. Reverse-Engineering Visualizations: Recovering Visual Encodings from Chart Images. *Computer Graphics Forum* 36, 3 (2017), 353–363. doi:10.1111/cgf.13193

[95] Jacob Ritchie, Daniel Wigdor, and Fanny Chevalier. 2019. A Lie Reveals the Truth: Quasimodes for Task-Aligned Data Presentation. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland Uk, 1–13. doi:10.1145/3290605.3300423

[96] Maryam Salahshour Rad, Mehrbakhsh Nilashi, and Halina Mohamed Dahlan. 2018. Information Technology Adoption: A Review of the Literature and Classification. *Universal Access in the Information Society* 17, 2 (2018), 361–390. doi:10.1007/s10209-017-0534-z

[97] Joni Salminen, Soon-Gyo Jung, Shammur Chowdhury, Sercan Sengün, and Bernard J. Jansen. 2020. Personas and Analytics: A Comparative User Study of Efficiency and Effectiveness for a User Identification Task. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–13. doi:10.1145/3313831.3376770

[98] Manolis Savva, Nicholas Kong, Arti Chhajta, Li Fei-Fei, Maneesh Agrawala, and Jeffrey Heer. 2011. ReVision. In *Proceedings of the 24th Annual ACM Symposium in User Interface Software and Technology*. ACM, New York, NY, USA, 393–402. doi:10.1145/2047196.2047247

[99] Anastasia Schaadhardt, Yue Fu, Cory Gennari Pratt, and Wanda Pratt. 2023. "Laughing so I Don't Cry": How TikTok Users Employ Humor and Compassion to Connect around Psychiatric Hospitalization. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3544548.3581559

[100] Joseph S. Schafer, Kate Starbird, and Daniela K. Rosner. 2023. Participatory Design and Power in Misinformation, Disinformation, and Online Hate Research. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*. ACM, Pittsburgh PA USA, 1724–1739. https://dl.acm.org/doi/10.1145/3563657.3596119

[101] Filipo Sharevski and Donald Gover. 2021. Two Truths and a Lie: Exploring Soft Moderation of COVID-19 Misinformation with Amazon Alexa. In *The 16th International Conference on Availability, Reliability and Security*. ACM, Vienna Austria, 1–9. doi:10.1145/3465481.3470017

[102] Imani N. Sherman, Jack W. Stokes, and Elissa M. Redmiles. 2021. Designing Media Provenance Indicators to Combat Fake Media. In *24th International Symposium on Research in Attacks, Intrusions and Defenses*. ACM, San Sebastian Spain, 324–339. doi:10.1145/3471621.3471860

[103] Erica Shusas. 2024. Designing Better Credibility Indicators: Understanding How Emerging Adults Assess Source Credibility of Misinformation Identification and Labeling. In *Designing Interactive Systems Conference*. ACM, IT University of Copenhagen Denmark, 41–44. https://dl.acm.org/doi/10.1145/3656156.3665126

[104] Yunya Song, Sai Wang, and Qian Xu. 2022. Fighting Misinformation on Social Media: Effects of Evidence Type and Presentation Mode. *Health Education Research* 37, 3 (2022), 185–198. doi:10.1093/her/cyac011

[105] Kate Starbird, Ahmer Arif, and Tom Wilson. 2019. Disinformation as Collaborative Work. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–26. doi:10.1145/3359229

[106] Chase Stokes, Cindy Xiong Bearfield, and Marti A. Hearst. 2023. The Role of Text in Visualizations: How Annotations Shape Perceptions of Bias and Influence Predictions. *IEEE Transactions on Visualization and Computer Graphics* 30, 10 (2023), 1–12. doi:10.1109/TVCG.2023.3338451

[107] Huiyun Tang, Gabriele Lenzini, Samuel Greiff, Björn Rohles, and Anastasia Sergeeva. 2024. “Who Knows? Maybe It Really Works”: Analysing Users' Perceptions of Health Misinformation on Social Media. In *Designing Interactive Systems Conference*. ACM, IT University of Copenhagen Denmark, 1499–1517. https://dl.acm.org/doi/10.1145/3643834.3661510

[108] Edward R. Tufte. 2001. *The Visual Display of Quantitative Information* (2. edition ed.). Graphics Press, Cheshire, USA.

[109] Uttishta Sreerama Varanasi, Teemu Leinonen, Nitin Sawhney, Minttu Tikka, and Rahim Ahsanullah. 2023. Collaborative Sensemaking in Crisis: Designing Practices and Platforms for Resilience. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*. ACM, Pittsburgh PA USA, 2537–2550. https://dl.acm.org/doi/10.1145/3563657.3596093

[110] Sukrit Venkatagiri, Anirban Mukhopadhyay, David Hicks, Aaron Brantly, and Kurt Luther. 2023. CoSINT: Designing a Collaborative Capture the Flag Competition to Investigate Misinformation. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*. ACM, Pittsburgh PA USA, 2551–2572. https://dl.acm.org/doi/10.1145/3563657.3595997

[111] Robert A. Virzi. 1992. Refining the Test Phase of Usability Evaluation: How Many Subjects Is Enough? *Human Factors* 34, 4 (Aug. 1992), 457–468. doi:10.1177/001872089203400407 Publisher: SAGE Publications Inc.

[112] Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The Spread of True and False News Online. *Science* 359, 6380 (2018), 1146–1151. doi:10.1126/science.aap9559

[113] Emily K. Vraga, Sojung Claire Kim, John Cook, and Leticia Bode. 2020. Testing the Effectiveness of Correction Placement and Type on Instagram. *The International Journal of Press/Politics* 25, 4 (2020), 632–652. doi:10.1177/1940161220919082

[114] Yuxi Wang, Martin McKee, Aleksandra Torbica, and David Stuckler. 2019. Systematic Literature Review on the Spread of Health-Related Misinformation on Social Media. *Social science & medicine (1982)* 240 (2019), 112552. doi:10.1016/j.socscimed.2019.112552

[115] Daniel Wessel, Christiane Attig, and Thomas Franke. 2019. ATI-S - An Ultra-Short Scale for Assessing Affinity for Technology Interaction in User Studies. In *Proceedings of Mensch Und Computer 2019* (Hamburg, Germany) *(MuC '19)*. Association for Computing Machinery, New York, NY, USA, 147–154. doi:10.1145/3340764.3340766

[116] Winnifred Wijnker, Ionica Smeets, Peter Burger, and Sanne Willems. 2022. Debunking Strategies for Misleading Bar Charts. *Journal of Science Communication* 21, 07 (2022), A07. doi:10.22323/2.21070207

[117] Liang Wu, Fred Morstatter, Kathleen M. Carley, and Huan Liu. 2019. Misinformation in Social Media. *ACM SIGKDD Explorations Newsletter* 21, 2 (2019), 80–90. doi:10.1145/3373464.3373475

7 Appendix

A Technical Implementation

Chart Checker - All detectable misleading features

Name	Description	How it can be misleading	Example
Truncated Y-Axis	A truncated y-axis hides part of the vertical scale. Instead, a portion of the axis is cut off or removed.	This can be misleading as it makes the displayed data look more extreme or exaggerated than it actually is which can lead to misinterpretation and false impressions	Show example
Inverted Y-Axis	An inverted y-axis in a chart means that the values on the vertical axis are shown in reverse order, with the higher values at the bottom and the lower values at the top.	This can be misleading because it flips the expected progression of the data, making upwards trends look like downwards trends and vice versa. This can cause confusion and misinterpretation of the data.	Show example
Misleading Aspect Ratio	A chart with a misleading aspect ratio distorts the proportions and relationships between data points.	This can be misleading as it makes trends across the data appear more or less steep than they actually are. This misrepresentation can lead to incorrect interpretations and conclusions.	Show example
Missing labels	Labels in a chart provide context and clarity about the data being presented. If they are missing, this information is not available. Such labels include the chart title, titles for the x-axis and y-axis or a legend.	This can be misleading, as without labels, we may lack critical context about the presented data, which makes it difficult to interpret values and make accurate comparisons.	Show example
Multiple Axes	A chart can have multiple x-axis or y-axis, which represent different scales or measurements.	This can be misleading, as it can become unclear which axis belongs to which data series. Furthermore, different scales or measurements on each axis can distort comparisons and show misleading patterns. This can lead to incorrect conclusions about relations between different data series.	Show example
Non-linear axis scale	Inconsistent scales can make certain data points appear disproportionately large or small compared to others. This can exaggerate or downplay differences between data points.	This can be misleading as it distorts the representation of data, making some sections of the data appear disproportionately large or small compared to others. This misrepresentation can lead to incorrect interpretations and conclusions.	Show example
Inconsistent Tick intervals	The markings along the axes are placed at uneven intervals. This can make it harder to assess the values in the diagram.	This can be misleading as it makes it difficult to assess the true distribution of the data, leading to confusion and misinterpretation of the trends or values presented.	Show example

Figure 5: Shows the view of ChartChecker, where all misleading features that can currently be detected (independent from a specific chart with specific deceptions) are described. The view is accessible through clicking on the 'Show All Detectable Features' button in the main view).

B Participatory Design Approach

Metric	N	%
Age		
18-24	5	50
25-29	2	20
40-44	1	10
50-54	1	10
60-64	1	10
Gender		
Male	6	60
Female	4	40
Academic level		
Academic degree	3	30
University student	2	20
Vocational training	3	30
High school diploma	1	10
Secondary education	1	10

Table 5: Demographic information of participants in the participatory design approach.

Prototype	User Feedback Summary
Prototype 1 (P1)	**Pros:** Simplicity and intuitiveness **Cons:** Visual appearance (not specific to P1)
Prototype 2 (P2)	**Pros:** Control chart overall helpful **Cons:** Control chart requires further explanation and is too complex **Suggestions:** Show as additional feature but not directly
Prototype 3 (P3)	**Pros:** Comprehensibility, Story-telling aspects of layout, Control chart **Cons:** Usability, Difficulty to compare charts
Prototype 4 (P4)	**Cons:** Large empty space draws too much attention
Prototype 5 (P5)	**Pros:** Usability, Engagingness and interactivity **Suggestions:** Implement option to show and hide features separately, Implement a control chart
Prototype 6 (P6)	**Cons:** Overall structure

Table 6: Summary of the user feedback on different UI prototypes from the participatory design approach.

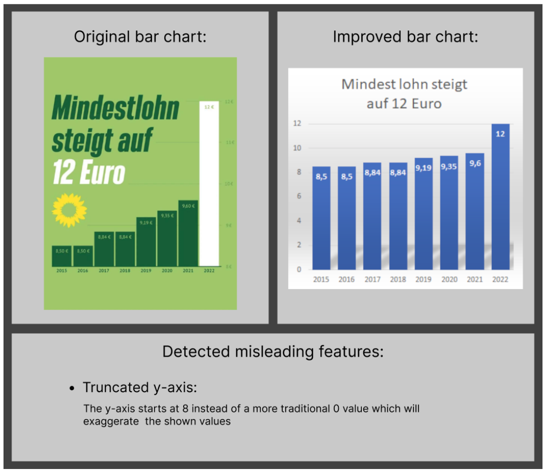
Rank	Feature	Score
1	Control chart	2
2	Table of all detectable misleading features	2,6
3	Toggle misleading features individually	2,7
4	Sharing function	2,8
5	Highlight bounding boxes	2,9
6	Inaccuracy warning label	3,5
7	Clear all data	4,5
8	Change chart type	5,6
9	Control bounding boxes	6,5

Table 7: Displays the average importance score for all feature ideas that were presented to participants. Explanations for each feature can be found in Appendix 10.

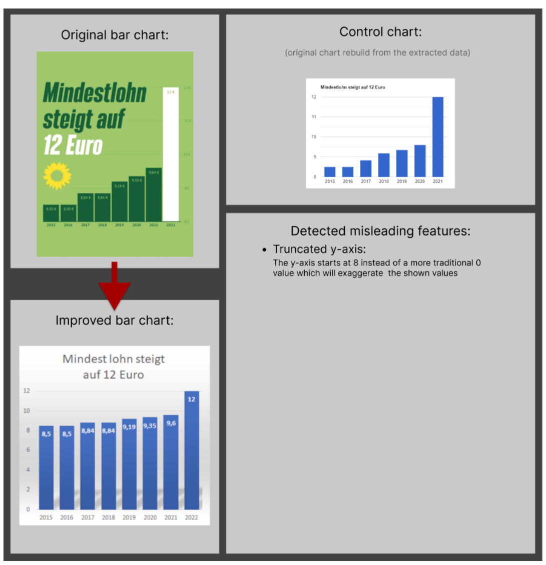
Design Principle	Description
Simplicity	Presenting only three main elements on the main page: the original chart, a modified version with deceptive features removed, and detailed textual descriptions of these removed features aids users in both phases of analysis.
Avoid unnecessary complexity	Implementing the widely known 3-click-rule and minimizing feature grouping and different views to keep user interaction simple and efficient.
Minimize cognitive load	Striving for an effortless and intuitive user experience by reducing the number of elements, displaying only necessary text, and incorporating user feedback on labeling and presentation.
Keep the users' needs in mind	Ensuring information presentation is simple and understandable, with detailed user study to determine real user needs.
Prioritise usability over design	Opting for neutral colors and familiar UI elements to prevent distraction from informative content.
Make the user feel in control	Incorporating interactive functionalities in future iterations to enhance user engagement and control.
Clear language	Constructing text elements to be clear and understandable, with feedback from user studies guiding language refinement.
Use familiar patterns	Employing UI elements that align with user expectations to reduce confusion and cognitive load.
Consistency	Maintaining a consistent design language across prototypes while remaining open to adjustments based on user feedback.
Give relevant feedback	Ensuring interactable elements clearly communicate their function and effects to the user.
Responsiveness	Designing the tool to adjust to different screen sizes for versatility in chart analysis.
Animations and transitions	Minimizing distracting animations and transitions unless serving a specific purpose.
Prioritise Functions	Ensuring essential elements remain visible in the main view, with less essential functionalities accessible through secondary interfaces like buttons or menus.

Table 8: Design principles guiding the creation of the user interfaces inspired by common guidelines and heuristics for interaction design [34, 54].

B.1 Prototype UI examples



(a) Prototype 1 (P1)



(b) Prototype 6 (P6)

Figure 6: Exemplary prototypes which have been developed for the participatory design approach.

Prototypes / Category	Mean	Median	All scores
P1 (average score)	2,02	2	
Rank	3.5	4	4-5-3-1-1-6-4-4-4-3
Comprehensibility	1,6	2	2-2-1-2-2-1-2-1-1-2
Intuitiveness	1,6	1	1-1-1-1-2-3-3-2-1-1
Visual Appeal	3	3	3-3-3-1-2-2-4-4-4-4
Usability	1,7	2	2-3-2-2-2-2-1-1-1-1
Overall	2,2	2	2-3-2-2-2-2-3-2-2-2
P2 (average score)	2,4	2	
Rank	2,9	2,5	2-2-4-3-2-5-5-3-1-2
Comprehensibility	2,1	2	2-1-2-2-4-1-4-1-2-2
Intuitiveness	2,4	3	3-1-2-3-3-2-3-3-1-3
Visual Appeal	3	3	3-3-3-2-4-2-2-4-4-3
Usability	2	2	2-2-2-3-4-2-2-1-1-1
Overall	2,5	2	3-2-2-3-4-2-3-2-2-2
P3 (average score)	2,44	2	
Rank	2,3	2,5	1-3-2-2-3-4-3-1-3-1
Comprehensibility	1,7	1,5	1-2-2-2-4-1-2-1-1-1
Intuitiveness	2,3	2	1-4-2-4-4-3-1-2-1-1
Visual Appeal	3,3	3	2-4-3-4-4-3-4-3-3-3
Usability	2,4	2	2-3-2-3-4-2-3-1-2-2
Overall	2,5	2,5	1-3-2-4-4-3-3-2-2-1
P4 (average score)	2,44	2	
Rank	4,6	5	5-6-5-5-4-3-2-5-6-5
Comprehensibility	1,6	1,5	1-2-1-2-3-1-2-1-1-2
Intuitiveness	2,2	2	1-2-2-4-3-2-1-2-3-2
Visual Appeal	3,5	3,5	3-3-4-4-3-2-3-4-4-5
Usability	2,2	2	2-2-2-3-4-1-1-2-3-2
Overall	2,6	3	2-2-3-3-3-2-2-3-3-3
P5 (average score)	2,5	2	
Rank	2,4	2	3-1-1-4-5-1-1-2-2-4
Comprehensibility	2,4	2,5	1-3-1-4-5-1-3-1-2-3
Intuitiveness	2,4	2	2-2-2-4-4-2-1-2-2-3
Visual Appeal	3,1	3	3-2-3-3-4-2-3-3-4-4
Usability	2,1	1,5	1-2-1-4-5-1-3-1-1-2
Overall	2,5	2	2-2-2-4-4-2-2-2-2-3
P6 (average score)	3,46	4	
Rank	5,3	6	6-4-6-6-6-2-6-6-5-6
Comprehensibility	3,3	3,5	3-4-3-5-4-1-4-2-3-4
Intuitiveness	3,7	4	3-5-3-5-4-2-4-3-4-4
Visual Appeal	3,8	4	4-4-3-4-4-2-4-5-4-4
Usability	3	3	2-4-2-3-4-2-4-2-3-4
Overall	3,5	4	3-4-3-4-4-2-4-3-4-4

Table 9: This table shows all scores the prototypes received in the respective categories. The average score is calculated without the rank as it followed a different scale.

B.2 Features from the Participatory Design Approach

Feature	Description
Control chart	A chart displaying the data extracted from the input image. It can be used to quickly identify if the chart data was correctly extracted from the original chart.
Table of all detectable MFs	A table view of all misleading features the tool can detect. It includes descriptions of each MF, as well as explanations how the are misleading, as well as the option to load an example for each into our tool.
Toggle MFs individually	The ability to individually show and hide all detected MFs. This can be utilised to better visualise the effects of each misleading feature on the recommended chart.
Sharing function	The functionality to quickly save or copy the results of out tool to the clipboard. Users can choose between sharing the main view of the original image, recommended chart and explanations of the detected misleading features, or just the recommended chart.
Highlight bounding boxes	When hovering over a detected misleading feature in the list, the tool would visually highlight the affected areas in the original chart. E.g. for a truncated y-axis, the y-axis would be highlighted by surrounding it with a red ellipsis.
Inaccuracy warning label	In some situations, a chart containing misleading features might be more applicable to display data than the alternative chart our tool presents. E.g. for a line charts that varies between values of 900 and 950, a chart containing a truncated y-axis might be appropriate. Our tool could contain a warning label in its UI highlighting this issue.
Clear all data	As the manual mode has a lot of text fields that would need to be individually cleared, our tool could include a button to clear all of them at once. It has to be noted, that the manual mode can be closed an re-opened to achieve the same effect.
Change chart type	We could include the option to change the chart type of the alternative chart to visualise the extracted data differently.
Control bounding boxes	Similar to how the control chart gives users the option to confirm that the chart data was correctly extracted, displaying them the bounding boxes together with the assigned type of the data would give them the opportunity to control if they were given the correct types.

Table 10: Features tested in the participatory design approach

C Think-Aloud Study

C.1 Study Structure

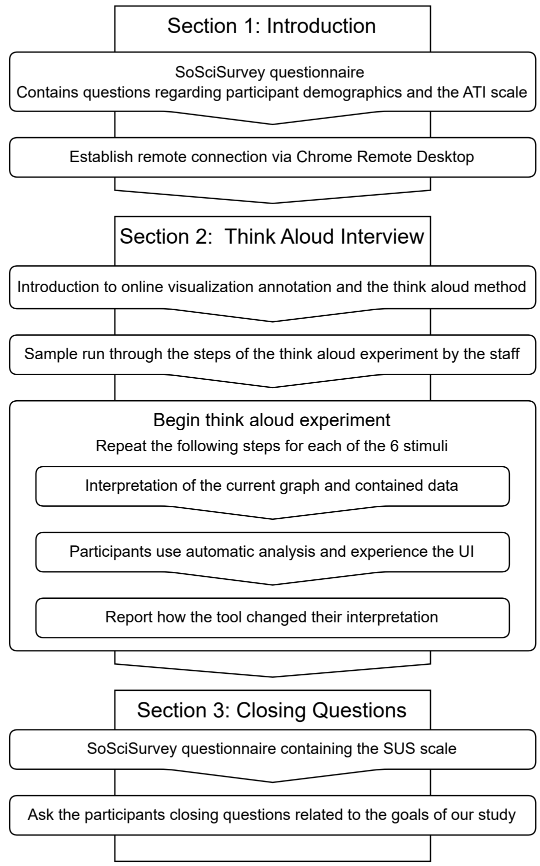
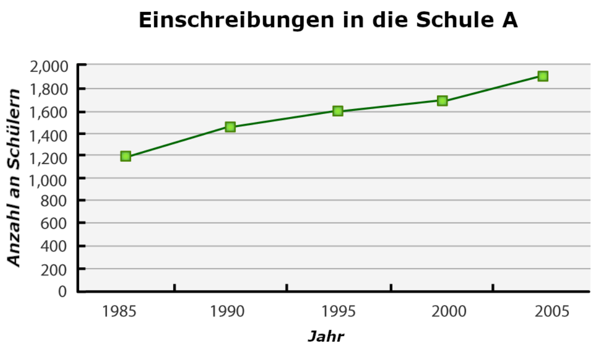
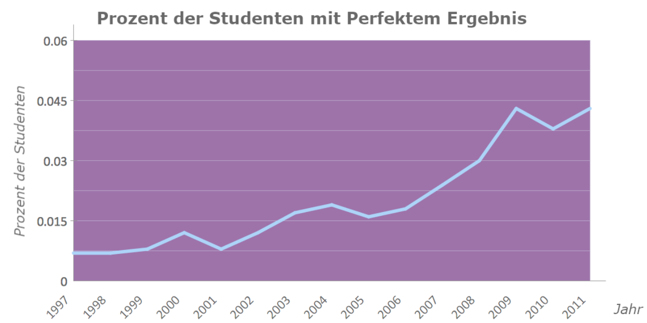


Figure 7: The Structure of our study. The goals we refer to in the diagram are defined at the beginning of subsection 4.1.

C.2 Stimuli

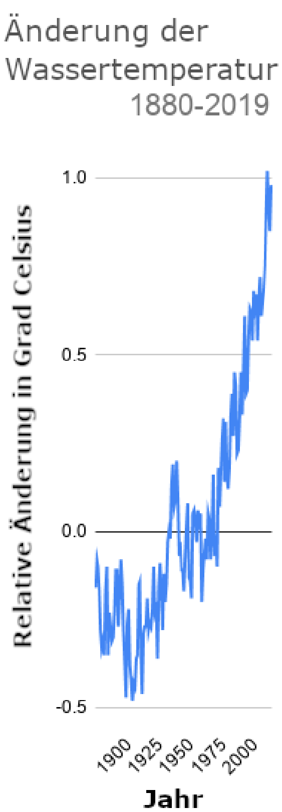


(a)

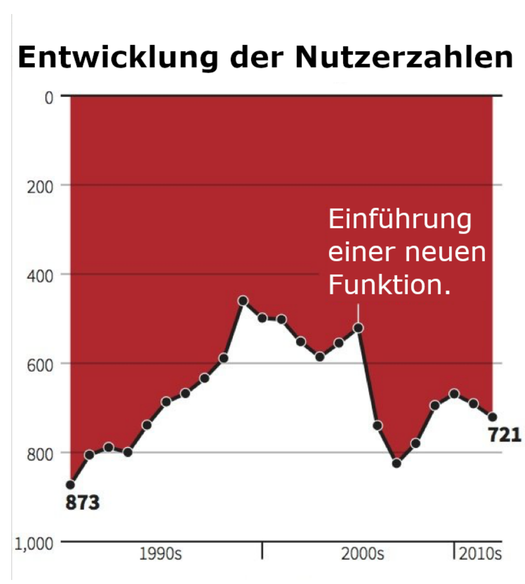


(b)

Figure 8: Stimuli (I) for the think-aloud experiment. Visualisations are translated real world examples that have been stripped of political connotations for the purpose of our study. Figure 8a and Figure 8b are positive examples that do not contain any misleading features.

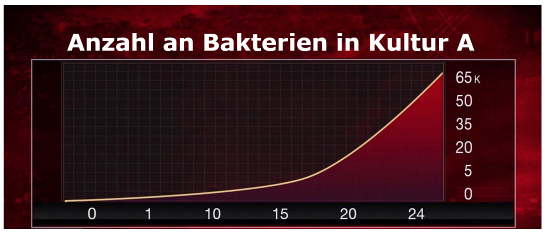


(a)

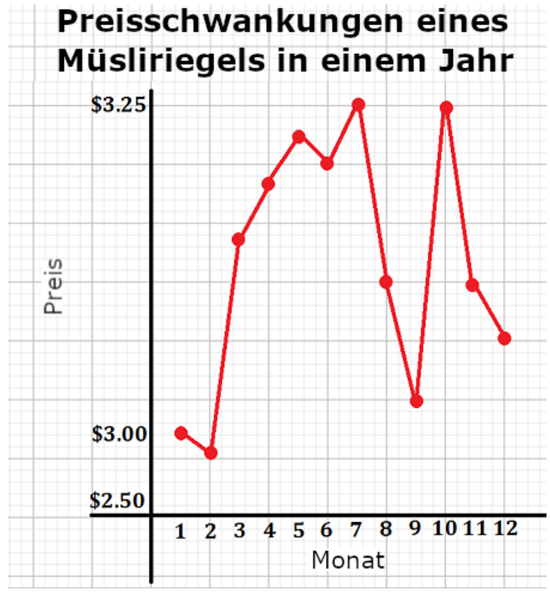


(b)

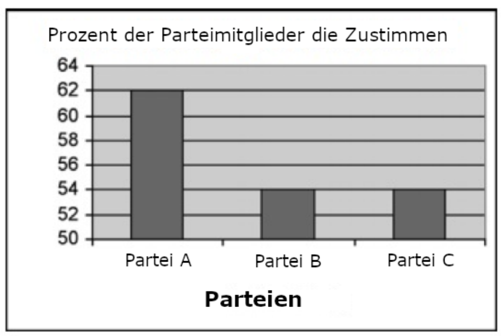
Figure 9: Stimuli (II) for the think-aloud experiment. Visualisations are translated real world examples that have been stripped of political connotations for the purpose of our study.



(a)



(b)



(c)

Figure 10: Stimuli (III) for the think-aloud experiment. Visualisations are translated real world examples that have been stripped of political connotations for the purpose of our study.

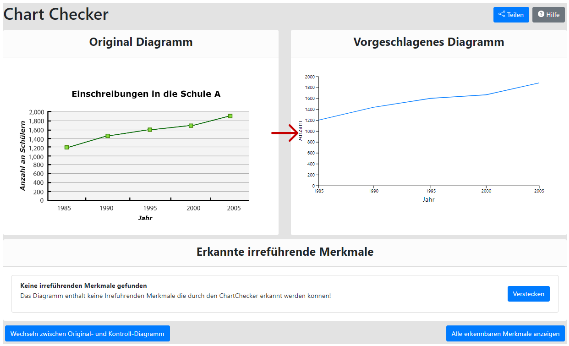


Figure 11: ChartChecker screenshot with stimulus 1

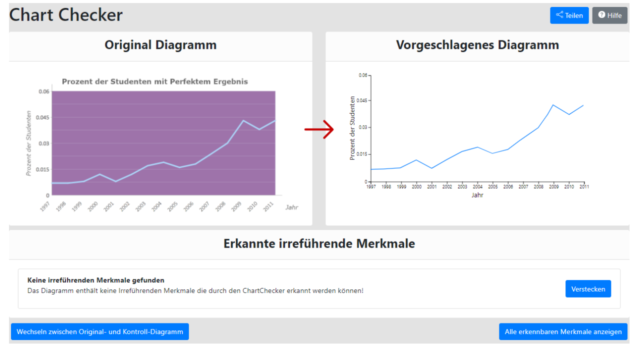


Figure 12: ChartChecker screenshot with stimulus 2

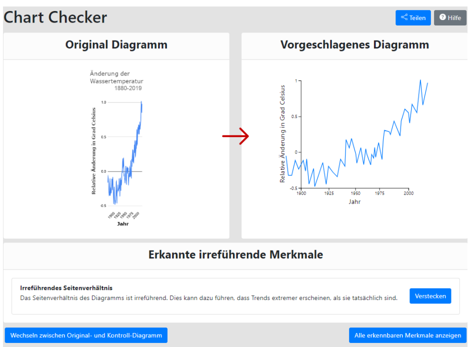


Figure 13: ChartChecker screenshot with stimulus 3

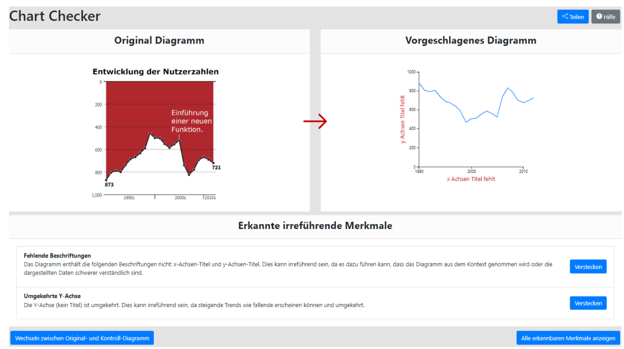


Figure 14: ChartChecker screenshot with stimulus 4

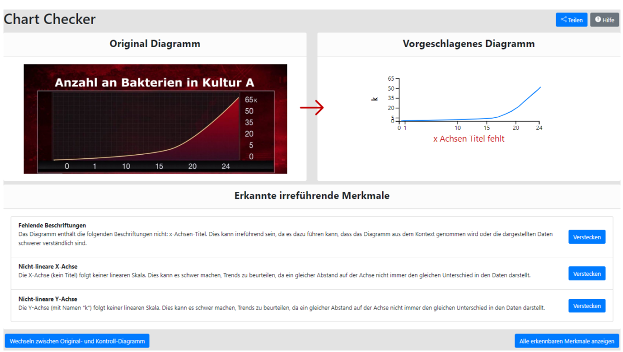


Figure 15: ChartChecker screenshot with stimulus 5

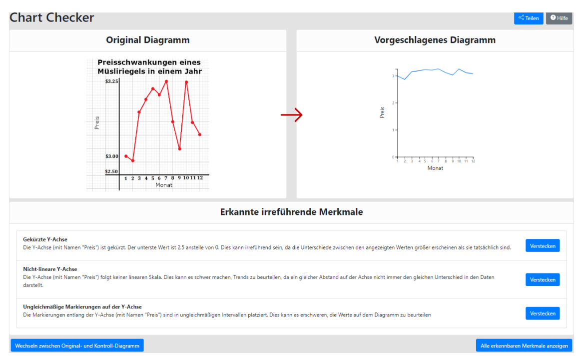


Figure 16: ChartChecker screenshot with stimulus 6

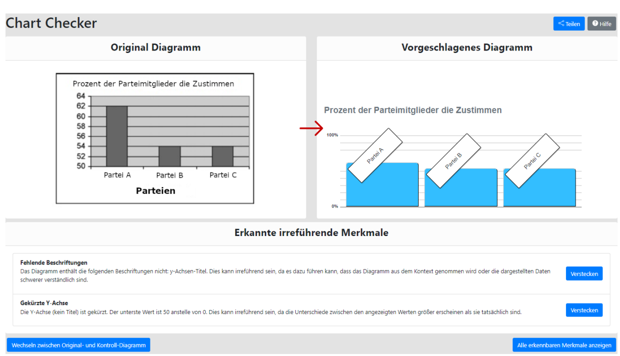


Figure 17: ChartChecker screenshot with stimulus 7

Feature	Description
Figure 9a: *Misleading aspect ratio*	This makes the increase in temperature seem even more dramatic than it is.
Figure 9b: *Inverted y-axis* and *missing y-axis label*	The inverted y-axis makes it seem as if the value displayed drops significantly after the highlighted point when it really goes up.
Figure 10a: *Non-linear x-axis*, *non-linear y-axis*, and *missing axis labels*	These features let the growth of the value depicted in the chart seem less steep than it really is.
Figure 10b: *Truncated y-axis*, *inconsistent tick placement*, and *non-linear y-axis*	The combination of these three features makes the value seem to fluctuate a lot when it really only fluctuates by about 10% of the initial value.
Figure 10c: *Truncated y-axis*	Exaggerates the difference between the bars

Table 11: Description of all misleading features within the stimuli.

C.3 Interview Questions

(1) User Interaction
   (a) How do you evaluate the usability/ user interface of the tool?
(2) Functionality
   (a) How do you evalute the functionality of the tool?
   (b) Do you think the time the tool takes to analyze diagrams is reasonable?
   (c) What was good in terms of features, if you liked anything in particular?
   (d) What functions do you think are missing?
(3) Comprehensibility
   (a) How would you evaluate the comprehensibility of the new diagrams produced by the tool?
   (b) How would you rate the comprehensibility of the misleading features described by the tool?
(4) Usefulness
   (a) Would you use the tool? And if so, in which cases? Why (not)?
   (b) Overall, how useful did you find the tool for interpreting the information? Do you think it can help you or others (e.g. younger siblings or an older neighbor) to better assess information in diagrams?

C.4 Codebook

- User interaction
  - Correct use of user interface
    - Run the tool independently
  - Wrong expectation of user interface
  - Positive evaluation
    - Confirming feeling when recognizing features already found by the user
    - Presentation clearer than original diagram
    - Comparison of original vs. improved diagram positive
    - Uncomplicated, intuitive operation
  - Negative evaluation
    - Presentation is less clear than original diagram
    - Thinks it wouldn't have been usable without an introduction
    - Would have expected more features
    - UI should be designed better
    - Dissatisfaction with improved chart
    - Dissatisfaction with hide function
    - Too much irrelevant area in the chart
    - Dissatisfaction with share function
  - Suggestion for improvement
- Comprehensibility
  - Correct understanding of a misleading feature description
  - Incorrect understanding of a misleading feature description
    - Non-linearity too complex
    - Inverted Y-axis too complex
    - Inconsistent tick placement too complex
  - Improved understanding of the chart overall
  - Unchanged understanding of the chart overall
    - Because everything was recognised beforehand
    - Because there is no deception taking place
  - Worsened understanding of the chart overall
    - Because of information loss
  - Positive evaluation
    - Generated chart comprehensible
    - Misleading feature texts comprehensible
  - Negative evaluation
    - Diagram comprehensibility poor/limited
    - Feature texts not easy to understand
    - Too much text
- Functionality
  - Positive evaluation
    - Adequate analysis time
    - Explanation of all features good
    - Generated diagrams look good/understandable
    - Help function good
    - Only one mouse click is needed for analysis
    - Encourages you to think about the diagram again
    - Share function good
    - Tool detects relevant misleading features
    - Hide function good
    - Before and after comparison good
  - Negative evaluation
    - Analysis time should be shorter
    - The chart title should appear
    - Loss of information
    - Disagree with misleading feature
    - Dissatisfaction with generated diagram
  - Suggestions for improvement
    - Batch analyse
    - Better color coding of functions
    - Button right next to diagram
    - Dark mode
    - The chart heading should appear
    - Advanced tutorial with examples

    * Feedback function to tell developers things
    * Feedback for users when using the tool
    * Smartphone version
    * Highlighting of misleading feature list desired
    * Hide irrelevant axis sections
    * Improve control chart function
    * Insert loading animation during analysis time
    * Add lines to the background of the chart
    * More context through automatic background search for diagram content
    * Support more languages
    * Show more intermediate steps on axis
    * Parallel view of the ChartChecker and all recognisable features
    * Parallel view of the tool and the original chart
    * Change position of misleading features
    * Make specific values visible when mouse over diagram point
    * Different chart design options
- Usefulness
  - Successful correction
  - No successful correction
  - Identifying previous features in new charts
  - Positive evaluation
    * Trust in chart is growing
    * Enable learning effect
    * Would use tool
    * Would recommend tool
  - Negative evaluation
    * Mistrust in tool
    * Extra effort to analyse a diagram is too high
    * Would not use tool
    * Would not recommend tool
  - Application scenario