What Color Scheme is More Effective in Assisting Readers to Locate Information in a Color-Coded Article?

Ho Yin Ng* Zeyu He† Ting-Hao 'Kenneth' Huang‡

Pennsylvania State University

ABSTRACT

Color coding, a technique assigning specific colors to cluster information types, has proven advantages in aiding human cognitive activities, especially reading and comprehension. The rise of Large Language Models (LLMs) has streamlined document coding, enabling simple automatic text labeling with various schemes. This has the potential to make color-coding more accessible and benefit more users. However, the impact of color choice on information seeking is understudied. We conducted a user study assessing various color schemes' effectiveness in LLM-coded text documents, standardizing contrast ratios to approximately 5.55:1 across schemes. Participants performed timed information-seeking tasks in color-coded scholarly abstracts. Results showed non-analogous and yellow-inclusive color schemes improved performance, with the latter also being more preferred by participants. These findings can inform better color scheme choices for text annotation. As LLMs advance document coding, we advocate for more research focusing on the "color" aspect of color-coding techniques.

Index Terms: Color, Color coding, Information seeking, Text visualization, Document.

1 INTRODUCTION

Color coding is a simple yet powerful technique that assigns specific colors to different types of information within an article, effectively performing a clustering task [18] in the domain of text visualization. Despite being simple, color coding is remarkably powerful, helping readers quickly identify key information in an essay and assisting writers in analytically evaluating their own compositions [25]. Studies also showed that color coding benefits human cognitive activities [4, 8, 23, 28]. Meanwhile, the recent emergence of LLMs has transformed the document annotation landscape. LLMs enabled effortless labeling of arbitrary text documents with arbitrary label schemes automatically [10, 34]. This development holds significant potential for enhancing the accessibility of color coding, allowing a broader audience to reap its benefits in reading and writing support. From a Human-Computer Interaction (HCI) perspective, **given LLM's advancements in the "coding" aspect of color coding, it is opportune to focus more on the "color" aspects of this technique.** We noted a lack of discourse on the significance of color choice, particularly in supporting information-seeking in text through different color schemes.

This paper presents a user study designed to evaluate the efficacy of color schemes in enhancing readers' information-seeking for color-coded documents. To replicate the scenario of LLM-coded documents, we used GPT-4 to annotate the documents with our specified color schemes. We designed information-seeking tasks using abstracts from scholarly papers. We developed the main user study with 10 color schemes generated by 4 base colors in 2 distinct color temperatures (warm: red and yellow; cool: blue and green) to investigate how color schemes generated with specific base colors could impact the information-seeking performance in the text document. The results (Section 4) indicated that *(i)* the **non-analogous (mixed color temperature)** color schemes significantly improved overall performance compared to **analogous** color schemes, *(ii)* **dichromatic** schemes resulted in shorter response time than **monochromatic** schemes, *(iii)* **yellow-inclusive** schemes resulted in shorter response times and were more preferred by most participants, *(iv)* **red-inclusive** schemes led to longer response times and were least favored by the participants. This paper shows that the choice of color schemes significantly affects readers' performance in seeking information in color-coded documents.

2 BACKGROUND

2.1 Color Coding: Enhancing Learning, Reading, and Information Handling

Color coding is an efficient technique that utilizes stimuli to enhance information handling [1]. Dating back to the 1950s, early research explored color's role in visual search tasks for target numbers [12]. Hitt's study compared coding methods across operator tasks, indicating that color and numeral coding outperformed others [14]. As mentioned in the Introduction (Section 1), various research supports color coding's advantages in aiding human cognitive activities [4, 8, 23, 28], especially reading and comprehension [19, 24, 31, 38]. Recognizing these benefits, color coding has gained popularity as a tool in classrooms, supporting student learning and reading. It enables text analysis by highlighting different parts of speech, classifies text genres using color patterns to help students understand sentence structure [36], teaches reading and composition through annotating different components [9], and identifies elements in analytical essays [25]. Notably, color-coding is particularly valuable in English as a Foreign Language education [2], as well as for students with learning disabilities [6, 21]. Most research on color-coding for better reading and comprehension comes from cognitive science or educational psychology, with little focus on quantitatively comparing color schemes in user tasks.

2.2 Color in Visualization

Previous studies explored color's role in visualization, particularly in areas like mapping quantitative data [33], categorical color maps [7], or 3D spatial representations [3]. The focus has primarily been on translating structural or numerical information into graphical forms. A notable example is the color scheme used in route tracing for the London metro map [20]. Discussions on color choice in visualization predominantly center around graphical representations, such as investigating suitable color scales for spatial-temporal data [32], color utilization in map-based information visualization [5], recommendations for infographics color palettes [39], and applying color palettes from historical artists' paintings into scientific image visualization [30]. In contrast, textual information has received less attention. Weber proposed a color-coding scheme for text highlighting based on English grammar parts of speech,

*e-mail: hzn5135@psu.edu
†e-mail: zmh5268@psu.edu
‡e-mail: txh710@psu.edu

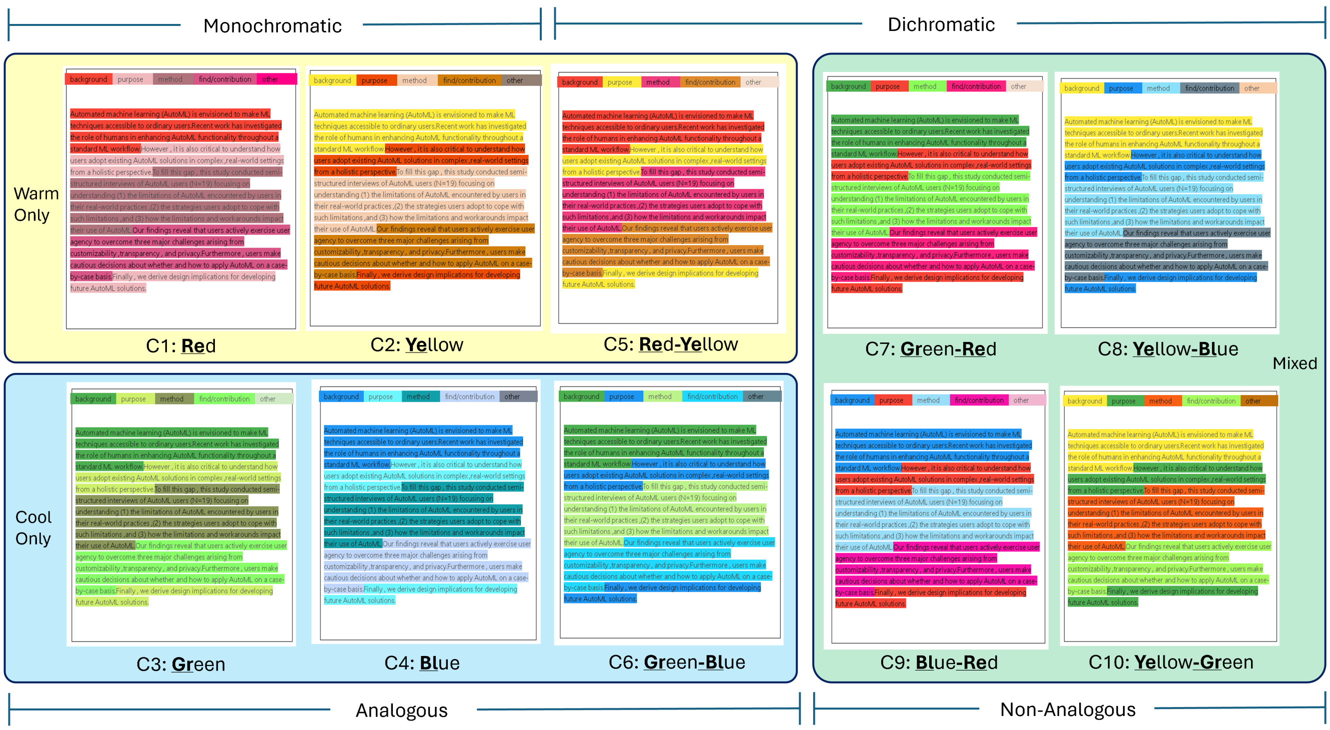

Fig. 1: The 10 color schemes used in our study, generated by combinations of 4 base colors—warm (Red, Yellow) and cool (Green, Blue). *Monochromatic* schemes consist of variations of a single base color; *Dichromatic* schemes are genereated by combining two different base colors; *Analogous* schemes feature base colors with the same or similar hues; *Non-analogous* schemes include contrasting base colors.

relying on the author's interpretation of color meanings in Western culture [36]. O'Connell and Fukao employed a reversed color-coded model, prompting students to match paragraph elements to colors based on context, with no explanation for the color choices provided in the research [27].

3 User Study

3.1 Task Design and Interface

We designed an information-seeking task that asks participants to identify the first text segment in the given abstract with a "target label" from our label scheme. The target label was one of the labels in the scheme, specified in the interface. The task contains 123 question items: 3 trial question items and 120 main question items. Each question item consists of an abstract and a multiple-choice question: "Which of the following is the FIRST segment belonging to the label [TARGET LABEL]?" All four options were chosen from the first segment of each label in the annotated text to distract participants if they were focusing on the first segment for the wrong label. An 8-second time limit was set for each question item to ensure a controlled environment for task difficulty across participants. Fig. 1 shows our interface. This task required participants to understand the matching between color and the label to seek information effectively. The task workflow aligns with regular reading habits for English text and the way humans seek information from passages (top to down, left to right): 1) Start by understanding what information to seek by identifying the target label (*e.g.*, "Method"), 2) Check the color of the corresponding label (*e.g.*, "Method" is indicated by "Blue"), 3) Locate the "color block", and 4) Identify the first segment (top of the color block).

The interface recorded the participant's answers and timestamps; we measured the response time on each question item by noting the start time at the commencement of the countdown timer. The end time was marked at the point of the participant's final selection made within the 8-second frame. In cases where the participant made no selection, no answer was recorded and the response time was automatically set to 8 seconds. The results were exported as a JSON file after the task was completed for each participant.

3.2 Experiment Material Preparation

3.2.1 Color Schemes Selection

Inspired by previous studies on the categorical perception of graphical stimulus using warm (red, yellow) and cool colors (green, blue) based on the CIELAB color space [15], we expanded this research to examine textual reading and a wider combination of colors in our study. We adopted the Material UI color system[1] and selected shade "500" for four identified hues: *Red 500 (#f44336)*, *Yellow 500 (#ffeb3b)*, *Blue 500 (#2196f3)*, and *Green 500 (#4caf50)* as the base color to generate the color schemes. We used Colorgorical [11] to generate 10 different color schemes using all possible combinations of 4 base colors (Fig. 1), with several groups of color schemes: **Monochromatic** schemes utilize a single base color, while **Dichromatic** schemes employ two. **Analogous** schemes comprise colors of the same or adjacent hues on the color wheel, and **non-analogous** schemes incorporate colors from disparate hue ranges. We used the consistent setting for the scheme generation: 1) High in Perceptual Distance and Pair Preference, 2) Low in Name Difference and Name Uniqueness, and 3) lightness range: 50 - 90. For the hue filters, we set "+/- 15°" for each base color (Red: 0°, Yellow: 60°, Green: 120°, Blue: 240°), except "+/- 35°" for blue to generate effective color schemes. We adopted the WCAG AA standard for normal texts that the contrast ratio is 4.5:1, using the contrast checker tools by WebAIM [35] to ensure sufficient contrast between the text color and each color from the generated color

[1] Material UI: https://mui.com/material-ui/customization/color/

Table 1. Contrast ratios for 10 color schemes. The color was used as the baseline with the lowest contrast ratio is denoted by ♠.

Color Scheme	Color 1	Color 2	Color 3	Color 4	Color 5
Re	5.56:1	5.55:1	5.55:1	5.56:1	5.56:1
Ye	5.63:1	5.57:1	5.55:1	5.59:1	5.57:1
Gr	5.57:1	5.6:1	5.56:1	5.6:1	5.59:1
Bl	5.61:1	5.59:1	5.6:1	5.63:1	5.56:1
Re-Ye	5.56:1	5.63:1	5.56:1	5.56:1	5.58:1
Gr-Bl	5.57:1	5.55:1	5.61:1	5.61:1	5.55:1 ♠
Gr-Re	5.55:1	5.56:1	5.6:1	5.55:1	5.6:1
Ye-Bl	5.63:1	5.55:1	5.63:1	5.55:1	5.61:1
Bl-Re	5.55:1	5.56:1	5.61:1	5.57:1	5.61:1
Ye-Gr	5.63:1	5.57:1	5.55:1	5.61:1	5.57:1

scheme. If any color fails the test, another set of color schemes would be generated until all colors from the scheme pass the contrast test.

Our study focuses on the performance difference among the color schemes for annotation only. We standardized contrast ratio to approximately 5.55:1 between highlight color and text color across all color schemes (Table 1). This ratio was derived from the lowest acceptable contrast between the highlight color(#658994 from Gr-Bl) and black text (#000000).

3.2.2 Data Processing and Label Scheme

The data used in the study contained abstracts extracted from recent publications in the arXiv HCI (`cs.HC`) field (from February to April 8th, 2024). We aimed to simulate the information-seeking scenario of college students reading paper abstracts during the literature review process. We used CODA-19's label scheme [17], categorizing sentence segments in paper abstracts into **Background, Purpose, Method, Finding, and Other**. We chose this scheme because it is closely relevant and useful for a graduate student's reading scenario. For instance, a student might use color-coding annotations to identify the "Finding" of a paper. To process the data, we first used Stanford CoreNLP [22] for tokenizing and segmenting sentences in all the abstracts. We then used commas (,) and periods (.) to divide sentences into smaller text segment parts. Each segment contained at least six tokens (excluding punctuation), ensuring the first segment in each label had sufficient content. We selected 706 abstracts with a token range from 150 to 250. The average abstract had 7.28 sentences (SD=1.47), which were further divided into 12.42 text segments (SD=2.50). Each abstract had 189.29 tokens (SD=26.32). We then used He's [13] zero-shot prompt on OpenAI's GPT-4 [26], which had a high accuracy (83.6%) evaluated by expert labels, to automatically label each segment in abstracts for this study. We excluded abstracts that lack one of four primary classes, namely Background, Purpose, Method, and Finding.

Data Selection. From the labeled abstracts, we randomly sampled 123 unique abstracts for the user study. The first 3 abstracts were used for trial question items. We divided the remaining 120 abstracts evenly among 10 color schemes for color annotation. Within each color scheme, 12 abstracts were further divided into 4 main labels, with each label corresponding to 3 abstracts. This division created a set of question items for each color scheme.

3.3 Study Procedure and Setups

Participants. We recruited 32 participants for the main study, with diverse educational backgrounds (12 Master's students, 15 Ph.D. students, and 5 Undergraduate students). All participants reported no visual impairment or color blindness.

Study Setups. The study took place in person in a university campus lab room, which featured a quiet environment and full white lighting to provide optimal conditions. Participants used a 24" BenQ EW2440L display monitor connected to a laptop (ROG

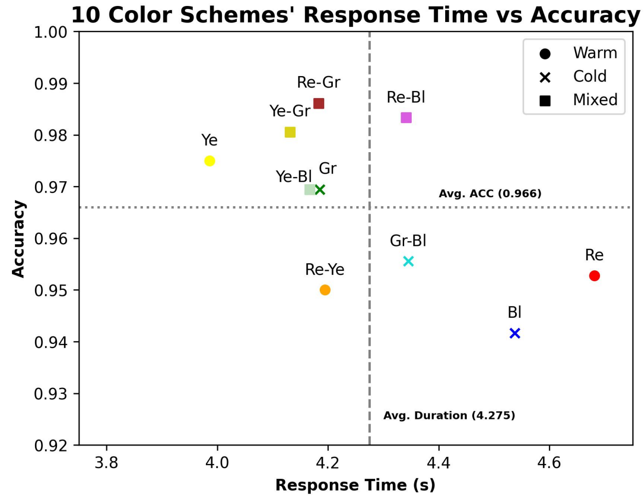

Fig. 2: Response Time and Accuracy Plot for 10 Color Scheme. Each color scheme was classified into 'Warm,' 'Cool,' and 'Mixed' categories and was denoted by distinct symbols.

Zephyrus G15) via HDMI cable. The monitor had a resolution of 1920 x 1080 with standard dynamic range (SDR) color space and specific settings: 1) Blue Light Setting: Low Blue Light - Multimedia (30%), 2) Brightness: 60, 3) Contrast: 50, and 4) Sharpness: 6. The monitor was centrally positioned on the desk, with a distance of 21" from the display edge to the desk (approximately the distance between participants' eyes and the monitor). This setup and setting remained constant for all participants throughout the study. During the task session, the researcher observed from a distance of approximately 40 inches to minimize disruption.

Study Procedure. Upon arrival, a researcher (one of the paper's co-authors) provided a briefing about the study's purpose, obtained oral consent, and guided the participants through a tutorial document to understand the web interface and tasks. The tutorial, displayed on the BenQ monitor in PDF format, ensured consistency between the tutorial and the actual task session. After addressing participant questions, the researcher launched the web interface and started screen recording. Participants entered their names and clicked "start" when ready. Participants first encountered the 3 trial question items to ensure they understood the task before proceeding. Then, they proceeded to complete the question set of 10 color schemes one by one, with 30 seconds in between each color scheme. To minimize the order effects, we randomized the order of question items within each color scheme and also the sequences of color schemes presented to each participant. After completing the test, participants completed an exit survey, rating their preference for each color scheme on a 5-point Likert scale, using the same text passage for all schemes (Fig. 1). The researcher then conducted follow-up interviews based on test observations and participants' preferences. Upon completion of the study, participants received a $5 cash compensation, confirmed by signing a compensation form. The user study took approximately 25 minutes for each participant.

4 Findings

Before analysis, we cleaned the data by removing two outliers who had low accuracies (0.25, 0.65) that fall below the average minus two standard deviations (0.934 - 2×0.140 = 0.654) [29]. The new accuracy after pruning is 0.966 (SD = 0.041). We documented the two outliers' behavior for detailed discussion in Section 5.

Participants performed significantly better when using non-analogous color schemes than analogous color schemes.

Table 2. User study results on different combinations of color palette. Accuracy pairs and Response Time pairs that passed the T-Test are denoted by † (p-value = 0.001), ♢ (p-value = 0.020), ♡ (p-value = 0.029), △ (p-value = 0.024), and ♣ (p-value < 0.001).

n=30	ACC		Response Time	
	Mean	95% CI	Mean	95% CI
Monochromatic	.960	[.938,.981]	4.348♢	[4.163,4.532]
Dichromatic	.971	[.956,.985]	4.227♢	[4.067,4.387]
Analogous	.957†	[.939,.976]	4.322♡	[4.160,4.483]
Non-Analogous	.980†	[.966,.993]	4.206♡	[4.022,4.389]
Cool Only	.956	[.933,.978]	4.356	[4.190,4.522]
Warm Only	.959	[.936,.983]	4.287	[4.113,4.461]
Red-Inclusive	.968	[.950,.986]	4.350△	[4.179,4.521]
Red-Exclusive	.965	[.946,.985]	4.225△	[4.054,4.396]
Yellow-Inclusive	.969	[.949,.989]	4.120♣	[3.950,4.289]
Yellow-Exclusive	.965	[.950,.980]	4.379♣	[4.212,4.545]
Green-Inclusive	.973	[.961,.985]	4.211	[4.037,4.385]
Green-Exclusive	.962	[.939,.985]	4.318	[4.128,4.508]
Blue-Inclusive	.963	[.936,.989]	4.348	[4.147,4.549]
Blue-Exclusive	.969	[.954,.984]	4.227	[4.061,4.392]

Fig. 2 shows the results for all individual color schemes. Interestingly, the accuracies of all 'mixed' color schemes were above average, with most also achieving faster-than-average response times. In contrast, the majority of 'cool' color schemes displayed both accuracies and response times below average. The results for 'warm' color schemes varied, with their accuracies and response times distributed throughout the figure.

Given that color schemes with mixed colors performed noticeably better than those with warm/cool colors, we sought to explore the factors influencing accuracy and response times. We set up several pairwise comparisons to clarify the factors affecting performance. These comparisons included monochromatic versus dichromatic schemes, analogous versus non-analogous schemes, warm colors versus cool colors, and color-inclusive versus color-exclusive schemes. Table 2 shows the results for various color scheme combinations. Non-analogous schemes significantly improved participants' information location performance, showing higher accuracy (p=0.001) and faster response times (p=0.029) compared to analogous schemes. Dichromatic schemes yielded significantly faster responses than monochromatic ones. Among color-specific schemes, red-inclusive schemes slowed responses, while yellow-inclusive schemes accelerated them, compared to their respective color-exclusive counterparts.

Participants preferred yellow-inclusive schemes and disliked red-inclusive schemes. In the exit survey, participants were asked to rate the color schemes with the following statement, "This color palette helps me to recognize the structure of the abstract at glance." for each color scheme using a 5-point Likert scale from Strongly Disagree to Strongly Agree. The three most preferred color schemes based on were Ye-Gr (4.581), Ye (4.226), and Ye-Bl (4.194). Conversely, the three least-liked color schemes were Re (2.419), Bl-Re (2.968), and Gr (3.387).

In addition to quantitative data, we gathered qualitative insights through observation and brief interviews with participants. For yellow-inclusive color schemes, which had comparably better performance in response time and better preference, many participants found it effective for quickly identifying information. One participant mentioned that "yellow and light colors are the best". For red-inclusive color schemes, which resulted in poor performance and preference, many participants mentioned that "the red colors are too bright and distracting." For cool color schemes, most participants did not give comments, except one participant showed a strong preference towards green: "I find Green as an easy-to-go color, other colors are creating an information overload in the mind."

5 Discussion

The impacts of red and yellow. It is surprising that red and yellow are both warm color groups but led to two different direction effects in response time: response time for color schemes generated by red as base color led to a significantly slower response, while yellow led to a significantly faster response. It might echo other psychological research about the "yellow priority" that yellow is more prominent compared to other colors [16]. More research is needed to understand the causes.

Suggestions for color coding of textual documents. A few suggestions emerged from our findings that can be used to inform the color-coding practices of textual documents: 1) Color schemes with **mixed color temperatures** give better performance and are recommended to be used for annotating textual documents to help readers accurately search for the information from specific labels. 2) **Yellow is recommended** to be included in the color schemes to help the reader locate information in a shorter time. 3) **Avoid the red** in the color schemes as it could delay the search time and cause discomfort to readers' eyes. These suggestions could inform the selection of color schemes for color-coded documents.

Outliers' Behavior. Two participants were identified as outliers in the study due to specific circumstances affecting their performance. One participant reported experiencing significant stress and frustration during the task, citing their slow reading speed and insufficient time to process the text options before selecting answers. This highlights the impact of cognitive load and the need for inclusive design considerations in tasks involving reading. The other participant expressed discomfort using the mouse, which impeded their ability to select answers quickly in several questions. These cases show a potential limitation in the study design, specifically the requirement for participants to use a mouse to complete the task, which may not accommodate all users equally.

Limitations. We acknowledge a few limitations in this work. Firstly, our study focused on examining the impact of color schemes on locating specific textual information, omitting a direct assessment of reader comprehension. While comprehension is crucial, it is more resource-intensive to evaluate than information location. We identify this as a potential avenue for future research. Secondly, perceived visibility or "pop-out" effect may still vary across different color combinations due to factors such as chromatic contrast and individual perceptual differences [37] despite our effort to standardize the contrast ratio between each highlight color and the respective text color. Future studies could benefit from incorporating more advanced color appearance models or conducting supplementary perceptual tests to address color visibility. Lastly, our focus on scholarly articles limits generalizability. Future studies should examine these effects across diverse text types.

6 Conclusion and Future Work

This paper conducts a user study to assess the impact of 10 color schemes, generated from the combinations of 4 base colors, on rapid and accurate information seeking in color-coded documents. The results indicate that the **non-analogous** color scheme leads to better information-seeking performance, **yellow-inclusive** schemes lead to shorter response time and are also more preferred by most participants. These could inform the better choice of color scheme for annotating text documents. As LLMs enhance our ability to code textual documents, we advocate for additional research specifically focusing on the "color" aspect of color-coding techniques.

References

[1] E. A. Alluisi and P. F. Muller Jr. Verbal and motor responses to seven symbolic visual codes: A study in sr compatibility. *Journal of Experimental Psychology*, 55(3):247, 1958. 1

[2] R. A. Azeez. Color-code strategy for improving writing academic paragraphs in efl classes. *Zanco Journal of Humanity Sciences*, 25(2), 2021. 1

[3] L. Besançon, A. Ynnerman, D. F. Keefe, L. Yu, and T. Isenberg. The state of the art of spatial interfaces for 3d visualization. In *Computer Graphics Forum*, vol. 40, pp. 293–326. Wiley Online Library, 2021. 1

[4] M. A. Dzulkifli and M. F. Mustafar. The influence of colour on memory performance: A review. *The Malaysian journal of medical sciences: MJMS*, 20(2):3, 2013. 1

[5] S. Einakian and T. S. Newman. An examination of color theories in map-based information visualization. *Journal of Computer Languages*, 51:143–153, 2019. 1

[6] K. B. Ewoldt and J. J. Morgan. Color-coded graphic organizers for teaching writing to students with learning disabilities. *Teaching Exceptional Children*, 49(3):175–184, 2017. 1

[7] H. Fang, S. Walton, E. Delahaye, J. Harris, D. Storchak, and M. Chen. Categorical colormap optimization with visualization case studies. *IEEE transactions on visualization and computer graphics*, 23(1):871–880, 2016. 1

[8] S. Folker, H. Ritter, and L. Sichelschmidt. Processing and integrating multimodal material—the influence of color-coding. In *Proceedings of the annual meeting of the Cognitive Science Society*, vol. 27, 2005. 1

[9] B. A. Geigle. How color coding formulaic writing enhances organization: A qualitative approach for measuring student affect. *Online Submission*, 2014. 1

[10] F. Gilardi, M. Alizadeh, and M. Kubli. Chatgpt outperforms crowd-workers for text-annotation tasks. *arXiv preprint arXiv:2303.15056*, 2023. 1

[11] C. C. Gramazio, D. H. Laidlaw, and K. B. Schloss. Colorgorical: Creating discriminable and preferable color palettes for information visualization. *IEEE transactions on visualization and computer graphics*, 23(1):521–530, 2016. 2

[12] B. F. Green and L. K. Anderson. Color coding in a visual search task. *Journal of experimental psychology*, 51(1):19, 1956. 1

[13] Z. He, C.-Y. Huang, C.-K. C. Ding, S. Rohatgi, and T.-H. K. Huang. If in a crowdsourced data annotation pipeline, a gpt-4. In *Proceedings of the CHI Conference on Human Factors in Computing Systems at CHI 2024*, CHI '24. Association for Computing Machinery, New York, NY, USA, 2024. doi: 10.1145/3613904.3642834 3

[14] W. D. Hitt. An evaluation of five different abstract coding methods—experiment iv. *Human Factors*, 3(2):120–130, 1961. 1

[15] K. J. Holmes and T. Regier. Categorical perception beyond the basic level: The case of warm and cool colors. *Cognitive science*, 41(4):1135–1147, 2017. 2

[16] K. Hu, E. De Rosa, and A. K. Anderson. Yellow is for safety: perceptual and affective perspectives. *Psychological research*, 84:1912–1919, 2020. 4

[17] T.-H. K. Huang, C.-Y. Huang, C.-K. C. Ding, Y.-C. Hsu, and C. L. Giles. CODA-19: Using a non-expert crowd to annotate research aspects on 10,000+ abstracts in the COVID-19 open research dataset. In *Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020*. Association for Computational Linguistics, Online, July 2020. 3

[18] K. Kucher and A. Kerren. Text visualization techniques: Taxonomy, visual survey, and community insights. In *2015 IEEE Pacific visualization symposium (pacificVis)*, pp. 117–121. IEEE, 2015. 1

[19] K.-R. Lee. Raising grammatical consciousness by color-coding to improve english reading performance. *English Teaching*, 63(3):159–178, 2008. 1

[20] P. B. Lloyd, P. Rodgers, and M. J. Roberts. Metro map colour-coding: Effect on usability in route tracing. In *Diagrammatic Representation and Inference: 10th International Conference, Diagrams 2018, Edinburgh, UK, June 18-22, 2018, Proceedings 10*, pp. 411–428. Springer, 2018. 1

[21] C. Maldonado-Otto and C. Ormsbee. Color-coding affect on writing instruction for students with learning difficulties. In *ICERI2019 Proceedings*, pp. 11421–11432. IATED, 2019. 1

[22] C. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. Bethard, and D. McClosky. The Stanford CoreNLP natural language processing toolkit. In K. Bontcheva and J. Zhu, eds., *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pp. 55–60. Association for Computational Linguistics, Baltimore, Maryland, June 2014. doi: 10.3115/v1/P14-5010 3

[23] A. Marcus. *Graphic design for electronic documents and user interfaces*. Association for Computing Machinery, New York, NY, USA, 1991. 1

[24] E. Mirzabeiki and L. A. Nurliyana. Visualization of subjective extracted text using the parse tree. 1

[25] C. B. Olson and R. Land. A cognitive strategies approach to reading and writing instruction for english language learners in secondary school. *Research in the Teaching of English*, pp. 269–303, 2007. 1

[26] OpenAI. Gpt-4 api. `https://openai.com/gpt-4/`, 2023. Accessed: January 21, 2024. 3

[27] G. O'Connell and A. Fukao. Collaborative colour-coding for academic paragraph writing. *ICU Language Research Bulletin*, (33):34–43, 2019. 2

[28] P. A. Pruisner. From color code to color cue: Remembering graphic information. 1993. 1

[29] R. Ratcliff. Methods for dealing with reaction time outliers. *Psychological bulletin*, 114(3):510, 1993. 3

[30] F. Samsel, L. Bartram, and A. Bares. Art, affect and color: Creating engaging expressive scientific visualization. In *2018 IEEE VIS Arts Program (VISAP)*, pp. 1–9. IEEE, 2018. 1

[31] F. S. Sandvold, T. Schuller, A. Rolfsvåg, K.-E. Sikkerbøl, F. O. Medola, and F. E. Sandnes. Revisiting redundant text color coding in user interfaces. In *International Conference on Human-Computer Interaction*, pp. 467–476. Springer, 2023. 1

[32] P. Schulze-Wollgast, C. Tominski, and H. Schumann. Enhancing visual exploration by appropriate color coding. 2005. 1

[33] C. Tominski, G. Fuchs, and H. Schumann. Task-driven color coding. In *2008 12th International Conference Information Visualisation*, pp. 373–380. IEEE, 2008. 1

[34] P. Törnberg. Chatgpt-4 outperforms experts and crowd workers in annotating political twitter messages with zero-shot learning. *arXiv preprint arXiv:2304.06588*, 2023. 1

[35] J. Wang and S. Zionts. Webaim: an online aspiration-level interactive method. *Journal of Multi-Criteria Decision Analysis*, 13(2-3):51–63, 2005. 2

[36] W. Weber. Text visualization-what colors tell about a text. In *2007 11th International conference information visualization (IV'07)*, pp. 354–362. IEEE, 2007. 1, 2

[37] L. E. Wool, S. J. Komban, J. Kremkow, M. Jansen, X. Li, J.-M. Alonso, and Q. Zaidi. Salience of unique hues and implications for color theory. *Journal of vision*, 15(2):10–10, 2015. 4

[38] J.-H. Wu and Y. Yuan. Improving searching and reading performance: the effect of highlighting and text color coding. *Information & Management*, 40(7):617–637, 2003. 1

[39] L.-P. Yuan, Z. Zhou, J. Zhao, Y. Guo, F. Du, and H. Qu. Infocolorizer: Interactive recommendation of color palettes for infographics. *IEEE Transactions on Visualization and Computer Graphics*, 28(12):4252–4266, 2021. 1