https://doi.org/10.24867/GRID-2024-p47 Preliminary report

AI MODELS FOR PREDICTING VISUAL ATTENTION IN DIGITAL APPLICATIONS: A COMPARATIVE PILOT ANALYSIS WITH EYE TRACKING RESULTS

Neda Milić Keresteš, Gala Golubović, Sandra Dedijer, Živko Pavlović, Tamara Janjić
University of Novi Sad, Faculty of Technical Sciences,
Department of Graphic Engineering and Design, Novi Sad, Serbia

Abstract: *The paper presents a comparative pilot analysis of an AI-based visual attention prediction system with the traditional eye-tracking method in the context of UI/UX design. The study assesses the accuracy and reliability of AI-generated attention heatmaps compared to empirical eye-tracking data. The results indicate that AI systems show a strong ability to approximate human attention, achieving ROC-AUC values between 0.75 and 0.84. The study highlights AI's potential in early-stage design iterations, offering a suitable alternative to time-consuming eye-tracking testing sessions, but recommends combining AI tools with traditional methods for critical design decisions.*

Keywords: visual attention, eye-tracking, AI predictions, UI/UX design, ROC, saliency

1. INTRODUCTION

Human attention is an essential cognitive function due to the brain's limited capacity to simultaneously process all incoming visual information (Treue, 2003). Selective attention allows individuals to prioritize only the parts of a visual scene most relevant to their current task or situation, enabling them to respond efficiently in dynamic situations overloaded with visual stimuli (Posner & Petersen, 1990), such as interactions with digital interfaces.

Visual attention encompasses two key mechanisms: exogenous attention, which is reflexively triggered by external stimuli, such as bright objects or unexpected movements, and endogenous attention, which is voluntarily directed by the observer and guided by specific goals and tasks (Corbetta & Shulman, 2002). Understanding these mechanisms of visual attention provides the foundation for developing AI systems that simulate human attention and predict its distribution across visual scenes (Borji & Itti, 2013).

AI-based visual attention prediction relies on data collected through eye-tracking technologies and similar techniques that capture how users naturally distribute their attention (Tatler et al., 2011). The critical components of these systems are saliency and visual priorities. Saliency refers to the characteristics of objects or parts of a scene that make them stand out from their surroundings and attract our attention. These characteristics can be physical (such as contrast, color, and unusual shapes) or semantic (the meaning or relevance of an object in a given context), which together influence the automatic directing of the user's attention (Borji & Itti, 2013). Visual task-dependent priorities help AI models rank scene elements according to their importance.

1.1 Training and evaluating AI models for visual attention prediction

Saliency maps are a main tool for training AI models to simulate human attention mechanisms. They assign values to image points based on their likelihood to attract the user’s gaze. These maps are generated through the analysis of both low-level features (e.g., color, intensity, orientation, motion) and high-level features (e.g., semantic information like objects and faces) (Cornia et al., 2018).

That means that AI model training algorithms, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), rely heavily on large-scale datasets collected via eye-tracking technologies, which capture user gaze patterns on various visual scenes (Tatler et al., 2011).

The typical process in these models involves:

- Extraction of visual features: The model analyzes the input image to extract relevant features using computer vision techniques or deep learning methods, such as CNNs that automatically learn significant patterns from data.
- Generation of individual feature maps: For each extracted feature (e.g., color, contrast, orientation), a map is created showing its intensity across the entire image.

- Combination into a unified saliency map: The individual feature maps are merged through normalization and linear or nonlinear combination to produce a final saliency map that predicts areas of greatest interest.
- Prediction of fixation points: Based on the saliency map, the model predicts the probability that a user will direct their gaze to specific parts of the image, which can be validated against empirical eye-tracking data.

Different AI models apply saliency maps in various ways. Classical models focus on bottom-up processing based on physical scene properties, while deep learning models leverage complex representations of both low-level and high-level features (Kümmerer et al., 2017). Hybrid models incorporate both bottom-up and top-down factors like user intent.
The performance of these models is evaluated using standardized datasets such as MIT300 and CAT2000. The dataset of 300 imaged MIT300, part of the MIT/Tübingen Saliency Benchmark, is used specifically for objective performance assessment because its ground truth data are not publicly available, preventing overfitting (Kümmerer et al., 2018). The accuracy of the saliency maps is assessed through metrics such as AUC-Judd, sAUC, NSS, and CC, which measure different aspects of the model's predictive capabilities (Bylinskii et al., 2019), although Kümmerer et al. (2018) emphasize that a reliable AI model should demonstrate consistently high results across all metrics.

1.2 AI integration in the UI/UX design

AI has become integral to all stages of digital product development and usage, although its integration raises ethical issues (Jobin et al., 2019). One of the primary benefits of AI in UI/UX design is its ability to analyze large datasets. AI algorithms process extensive user data to uncover patterns in behavior and preferences, enabling designers to make data-driven improvements to design and functionality (Davenport & Ronanki, 2018). In addition to that, content personalization in digital applications becomes more achievable as AI predicts user needs based on previous interactions, allowing designers to create tailored experiences (Shneiderman, 2020). AI-driven tools also offer algorithmically generated design suggestions based on defined parameters and suggestions for enhancing accessibility by identifying issues such as poor color contrast (Lazar et al., 2015). Moreover, AI simplifies repetitive tasks by automating processes like resizing images or generating responsive designs, freeing designers to focus on more complex creative challenges. Additionally, AI enhances A/B testing by accurately predicting test outcomes based on previous data, making comparisons between different design elements more reliable (Kohavi et al., 2020). Finally, advancements in AI-based visual attention prediction can assess where users are likely to focus their attention, enable more informed design decisions, and accelerate design iterations (Borji & Itti, 2013).

1.3 Research aim and questions

The usability testing is a fundamental component of UI/UX design, involving the observation of users as they interact with a product to identify areas for design improvement (Nielsen, 1993). Traditional methods like eye-tracking provide valuable insights into user attention by recording eye movements and generating heatmaps that reveal patterns of focus (Duchowski, 2007). However, these methods require specialized equipment and significant time investment for recruiting, testing, and analyzing, making them resource-intensive and less practical during the early stages of design iteration (Moran, 2019).
The AI models offer promising alternatives by simplifying usability testing and predicting user attention without extensive eye-tracking studies (Xu et al., 2014). AI algorithms reduce manual effort while enabling quicker, automated testing cycles. This facilitates large-scale, continuous testing that can adapt rapidly to new data, design variations, and different user scenarios or focus groups.
This study aims to evaluate whether AI can reliably be used during the early stages and iterations of UI/UX design. Specifically, the research focuses on analyzing the performance of a widely used commercial AI system for visual attention prediction in a typical UI/UX design scenario. By conducting a pilot comparative analysis between AI-predicted visual attention maps and actual eye-tracking data obtained from real users, the study seeks to assess the reliability and accuracy of AI-based predictions in the context of UI/UX design.

The research questions of the study are the following:

1. How accurate are AI-based visual attention predictions compared to real eye-tracking data in the context of UI/UX design?
2. What are the key discrepancies, if any, between AI predictions and human visual behavior, and how can these gaps be minimized?
3. Can the AI system's predictions be reliably used to inform design decisions without the need for additional eye-tracking studies?

2. METHODS

2.1 Test samples

Four versions of a desktop application prototype for sharing travel experiences were created, with two featuring banner ads and two featuring native ads, differing only in the ad's position (top or bottom of the page). Banner ads, commonly used in applications, may be ignored by users due to so-called "banner blindness," where users recognize and avoid advertising spaces. In contrast, native ads are organically integrated into the application's content, making them appear more natural and informative to users.

2.2 The AI visual attention prediction

The analysis included six commercial AI tools that utilize attention prediction for evaluating user engagement with visual content. The selection criteria for these tools included the availability of a demo trial, suitability for testing web pages, mobile applications, or UI prototypes, and the availability of information regarding model evaluation. However, only two tools provided demo trials, with one proving unsuitable for analyzing web page content, as it yielded identical results across four test pages.
The remaining AI tool, Attention Insight, uses eye-tracking data from approximately 70,000 individual participants. to train its deep learning algorithms and generate precise visual attention heatmaps, which highlight areas where participants focus when viewing an image. The demography of 70,000 participants is 58% women and 42% men with ages ranging from 7 to over 60, although the majority of participants were between 21 and 30 years old and from the USA and Europe. In conducted eye-tracking studies, participants focused on visual stimuli for an average of 4 seconds.
Attention Insight's heatmap generation is powered by a Convolutional Neural Network (CNN), which mimics biological neural networks. Initially, the CNN's weights are randomized, resulting in inaccurate predictions, but through iterative training cycles, these weights are adjusted to reduce the error between the predicted and actual heatmaps obtained from eye-tracking studies. The model's accuracy continues to improve as new data is incorporated.
Attention Insight’s heatmaps have been tested and validated against the MIT/Tuebingen Saliency Benchmark. After testing with 300 images (MIT300 dataset), the model demonstrated 92.5% accuracy in general image prediction, with accuracy reaching up to 96%.

2.1 Traditional eye-tracking evaluation process

Seventeen participants, with an average age of 26, participated in the eye-tracking study. The Gazepoint device and its accompanying Control and Analysis software were used for calibration, data collection, and subsequent analysis. Prior to the experiment, the device was calibrated individually for each participant to ensure accuracy. Once calibration was complete, participants were instructed to observe the prepared design samples. After viewing each design, a black screen displaying a white cross was presented to reset the participants' gaze before proceeding to the next sample. After eye-tracking data collection, four cumulative heatmaps were generated, summarizing the visual attention of all 17 participants after 5 seconds of viewing each design sample.

2.3 The evaluation metric

When comparing an AI-predicted visual attention heatmap to a ground-truth heatmap obtained from eye-tracking data, selecting the appropriate evaluation metric is crucial. Several specialized metrics have been developed to assess how well a predicted saliency map matches human visual attention patterns. The fundamental tool is the Receiver Operating Characteristic (ROC) curve obtained from confusion matrix components.

A confusion matrix is presented in Table 1 and allows the performance visualization of a classification model by comparing actual values with predicted values. It includes the following components:

- True Positives (TP): Salient areas correctly predicted as salient.
- False Positives (FP): Non-salient areas incorrectly predicted as salient.
- True Negatives (TN): Non-salient areas correctly predicted as non-salient.
- False Negatives (FN): Salient areas incorrectly predicted as non-salient.

Table 1: A confusion matrix for ROC analysis

	Eye Tracking Map (P, positive)	Eye Tracking Map (N, negative)
AI prediction map (positive)	True Positive (TP)	False Positives (FP)
AI prediction map (negative)	False Negatives (FN)	True Negatives (TN)

From the confusion matrix, important metrics such as True Positive Rate (TPR) and False Positive Rate (FPR) can be derived. The ROC curve is a graphical representation that plots the TPR against the FPR at various threshold settings. It illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied.

The Area Under the ROC Curve (AUC-ROC) quantifies the overall ability of the model to discriminate between salient and non-salient areas. The AUC value ranges from 0 to 1, where:

- AUC = 0.5 suggests no discriminative ability (equivalent to random guessing),
- AUC < 0.6 indicates poor performance,
- AUC between 0.7 and 0.8 suggests acceptable performance,
- AUC > 0.8 is considered excellent.

3. RESULTS

Table 2 presents the original test pages alongside both the eye-tracking and AI-generated heatmaps for four different advertising setups: banner ads at the top and bottom of the page, and native ads at the top and bottom of the page. Table 3 provides metrics such as the percentage of attention (Area of Interest, AOI) and ROC-AUC values for each test design. Additionally, opacity maps are included in Table 3 to address the potential issue of different color palettes in heatmaps, ensuring consistency and eliminating outlier results from the cumulative heatmaps.

In ROC analysis, 0.5 is typically used as the standard or default threshold for binary classification problems. This means that if the predicted probability of an instance belonging to the positive class is greater than 0.5, it is classified as positive; otherwise, it is classified as negative. Also, AUC- Judd variant of AUC-ROC metric is used which includes all fixations.

Based on the data from Tables 2 and 3, the comparative analysis between traditional eye-tracking heatmaps and AI-generated heatmaps revealed notable insights. For the test page with the banner ad at the top, the area of interest (AOI) accounted for 19.4% of the attention time in the eye-tracking data, while the AI model predicted a lower percentage of 13.4%, with a corresponding ROC-AUC value of 0.79. For the banner ad at the bottom, the eye-tracking data indicated a smaller AOI of 6.8%, while the AI model predicted 8.1%, with a higher ROC-AUC value of 0.84. The native ad positioned at the top showed a significant focus, with 57.5% of attention time in the eye-tracking data, but the AI model estimated a lower 47.8%, with a corresponding ROC-AUC of 0.75. Similarly, the native ad at the bottom reflected 12.7% AOI in the eye-tracking data, with the AI model closely predicting 13.8% and a ROC-AUC value of 0.79. These results highlight the high degrees of alignment between human attention and AI predictions, with relatively strong ROC-AUC values indicating the model's ability to approximate human visual behavior in all four test cases.

Table 2: Heatmaps for test webpages

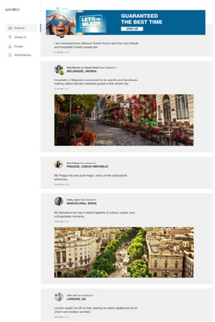
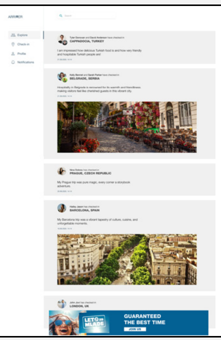
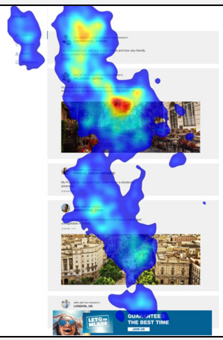
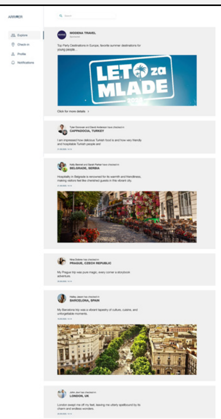
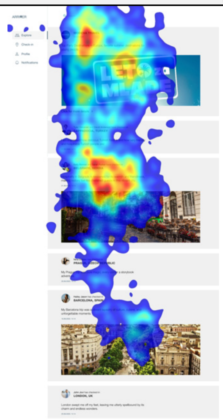
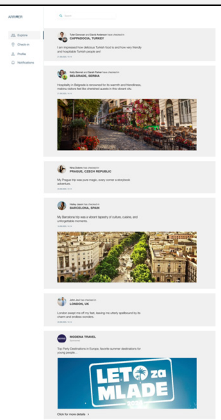
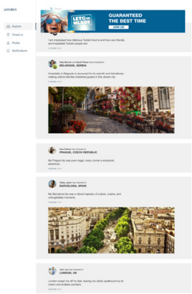
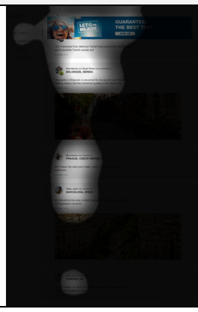
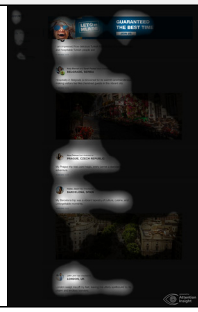
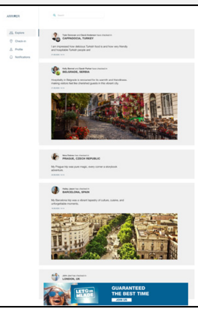
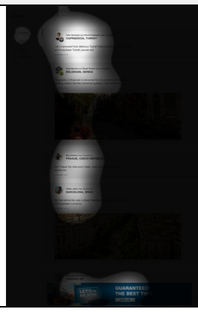
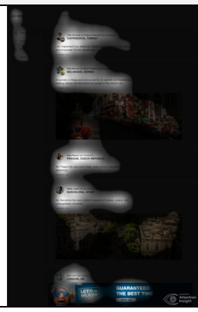
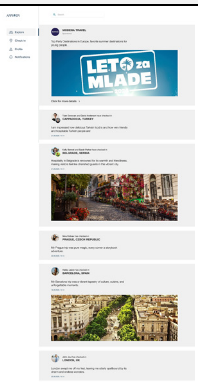
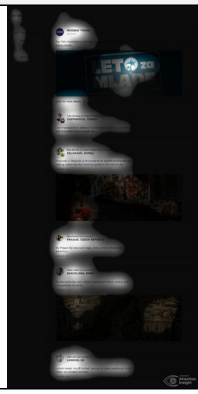
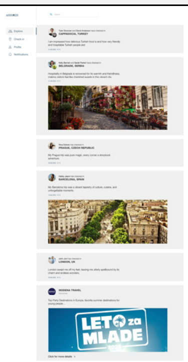
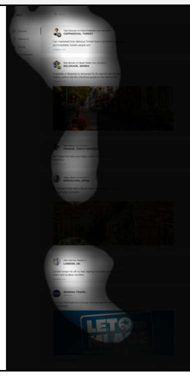
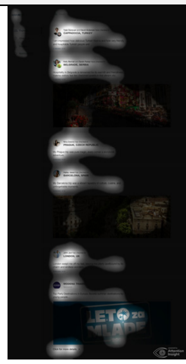
<table>
<tr><th>Original Test pages</th><th>Eye Tracking heatmap</th><th>Attention Insight heatmap</th></tr>
<tr><td colspan="3">Test page with the banner ad at the top of the page</td></tr>
<tr><td>
</td><td></td><td></td></tr>
<tr><td colspan="3">Test page with the banner ad at the bottom of the page</td></tr>
<tr><td>
</td><td></td><td></td></tr>
<tr><td colspan="3">Test page with the native ad at the top of the page</td></tr>
<tr><td>
</td><td></td><td></td></tr>
<tr><td colspan="3">Test page with the native ad at the bottom of the page</td></tr>
<tr><td>
</td><td></td><td></td></tr>
</table>

Table 3: Focus maps with metrics – AOI % of attention and ROC-AUC

Original Test pages	Eye Tracking heatmap	Attention Insight heatmap
Test page with the banner ad at the top of the page		
		
% time viewed AOI in the sum time	19.4%	13.4%
ROC-AUC value	0.79	
Test page with the banner ad at the bottom of the page		
		
% time viewed AOI in the sum time	6.8%	8.1%
ROC-AUC value	0.84	
Test page with the native ad at the top of the page		
		
% time viewed AOI in the sum time	57.5%	47.8%
ROC-AUC value	0.75	
Test page with the native ad at the bottom of the page		
		
% time viewed AOI in the sum time	12.7%	13.8%
ROC-AUC value	0.79	

4. DISCUSSION

The primary aim of this study was to evaluate the accuracy of AI-based visual attention predictions compared to real eye-tracking data in the UI/UX design context. By addressing the research questions, we sought to determine the viability of using AI tools for early-stage design iterations and to understand any discrepancies between AI predictions and human visual attention.

Accuracy of AI-Based Predictions (Research Question 1)

The comparative analysis between traditional eye-tracking and AI-generated heatmaps revealed a noteworthy alignment in the percentage values of time spent viewing the AOI (ad). In addition, ROC-AUC values, ranging from 0.75 to 0.84, indicate acceptable to excellent model performance, suggesting that the AI model can approximate human visual attention with a reasonable degree of accuracy.

Discrepancies Between AI Predictions and Human Visual Behavior (Research Question 2)

Despite the overall good alignment, some discrepancies were observed. For the native ad positioned at the top, the eye-tracking data indicated a significant focus with 57.5% attention time, whereas the AI model estimated a lower value of 47.8%. Also, the AUC-ROC values are lower compared to their evaluation with the MIT300 dataset. This variance may stem from the AI model's training predominantly on images that trigger exogenous attention meaning automatic responses to salient stimuli, whereas UX interfaces often require endogenous attention, which is goal-directed and task-oriented.

The discrepancies highlight that while the AI model is effective in predicting general attention patterns, it may not fully capture the nuances of user behavior in task-specific contexts. To minimize these gaps, incorporating UX-specific data into the AI model's training set could enhance its ability to predict endogenous attention patterns more accurately.

Reliability of AI Predictions for Design Decisions (Research Question 3)

The findings suggest that AI-based predictions can be a valuable asset in informing design decisions, especially during the early stages of the design process. The acceptable ROC-AUC values indicate that AI tools like Attention Insight can reliably predict areas of user focus, reducing the need for extensive eye-tracking studies, which are time-consuming. However, for crucial design decisions or interfaces requiring a precise understanding of user interactions, supplementing AI predictions with traditional eye-tracking may be beneficial to capture the full spectrum of user attention.

This study's limitations include a relatively small sample size and a focus on desktop applications with specific advertisement types. The AI model's training on images eliciting exogenous attention may limit its applicability in contexts requiring endogenous attention. Future research should involve more extensive eye-tracking data with more diverse test sets.

5. CONCLUSIONS

The pilot analysis demonstrates that AI attention prediction tools like Attention Insight can effectively approximate human visual attention in UX design contexts. Despite being primarily trained on exogenous stimuli, the AI model achieved acceptable ROC-AUC values, indicating its potential utility in predicting user attention without the need for extensive eye-tracking studies. The ability of AI tools to quickly and efficiently predict user attention areas presents a significant advantage in the iterative design process. While traditional eye-tracking provides precise data on user focus and gaze both order and duration, it requires specialized equipment and more time to conduct. AI predictions offer a practical alternative for rapid assessments, enabling designers to make timely and informed decisions to enhance usability and user satisfaction. A combined approach is recommended for comprehensive analyses and a deeper understanding of user behavior. Utilizing AI tools for initial predictions followed by validation through eye-tracking studies can achieve an optimal balance between speed and accuracy.

6. ACKNOWLEDGMENTS

This research has been supported by the Ministry of Science, Technological Development and Innovation (Contract No. 451-03-65/2024-03/200156) and the Faculty of Technical Sciences, University of Novi Sad through project “Scientific and Artistic Research Work of Researchers in Teaching and Associate Positions at the Faculty of Technical Sciences, University of Novi Sad” (No. 01-3394/1).

7. REFERENCES

Attention Insight. (2024) *AI-based visual attention prediction tool*. Available from: https://www.attentioninsight.com [Accessed 5th October 2024].

Borji, A. & Itti, L. (2013) State-of-the-art in visual attention modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 35(1), 185-207. Available from: doi: 10.1109/TPAMI.2012.89

Bylinskii, Z., Judd, T., Borji, A., Itti, L., Durand, F., Oliva, A. & Torralba, A. (2019) MIT saliency benchmark. Available from: https://saliency.tuebingen.ai/results.html [Accessed 5th October 2024].

Bylinskii, Z., Judd, T., Oliva, A., Torralba, A., & Durand, F. (2019) What do different evaluation metrics tell us about saliency models? *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 41(3), 740–757. https://doi.org/10.1109/TPAMI.2018.2815601

Corbetta, M. & Shulman, G. L. (2002) Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience.* 3(3), 201-215. Available from: doi: 10.1038/nrn755

Cornia, M., Baraldi, L., Serra, G. & Cucchiara, R. (2018) Predicting human eye fixations via an LSTM-based saliency attentive model. *IEEE Transactions on Image Processing*, 27(10), 5142-5154. Available from: doi: 10.1109/TIP.2018.2851672

Davenport, T. H., & Ronanki, R. (2018) Artificial intelligence for the real world. *Harvard Business Review*. 96(1), 108–116.

Duchowski, A. T. (2007) *Eye Tracking Methodology: Theory and Practice* (2nd ed.). Springer. Available from: doi: 10.1007/978-1-84628-609-4

Jobin, A., Ienca, M., & Vayena, E. (2019) The global landscape of AI ethics guidelines. *Nature Machine Intelligence.* 1(9), 389–399. Available from: doi: 10.1038/s42256-019-0088-2

Kohavi, R., Tang, D., & Xu, Y. (2020) *Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing*. Cambridge University Press.

Kümmerer, M., Wallis, T. S. A., & Bethge, M. (2017) Understanding low- and high-level contributions to fixation prediction. In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 4799–4808. IEEE. Available from: doi: 10.1109/ICCV.2017.513

Kümmerer, M., Wallis, T. S. A. & Bethge, M. (2018) Saliency benchmarking made easy: Separating models, maps and metrics. In: *Proceedings of the European Conference on Computer Vision (ECCV)*, 9-14 September 2018, Munich, Germany. Springer, pp. 630-645.

Lazar, J., Goldstein, D. F., & Taylor, A. (2015) *Ensuring Digital Accessibility Through Process and Policy*. Morgan Kaufmann.

Moran, K. (2019) Usability Testing 101: Introduction to Usability Testing. *Nielsen Norman Group*. Available from https://www.nngroup.com/articles/usability-testing-101/ [Accessed 5th October 2024].

Nielsen, J. (1993) *Usability Engineering*. Academic Press.

Posner, M. I. & Petersen, S. E. (1990) The attention system of the human brain. *Annual Review of Neuroscience*. 13(1), 25-42. Available from: doi: 10.1146/annurev.ne.13.030190.000325

Shneiderman, B. (2020) Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human-Computer Interaction.* 36(6), 495–504. Available from: doi: 10.1080/10447318.2020.1741118

Tatler, B. W., Hayhoe, M. M., Land, M. F. & Ballard, D. H. (2011) Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision*. 11(5), 5-5. Available from: doi: 10.1167/11.5.5

Treue, S. (2003) Visual attention: The where, what, how and why of saliency. *Current Opinion in Neurobiology*. 13(4), 428-432. Available from: doi: 10.1016/S0959-4388(03)00105-3

Xu, J., Jiang, M., Wang, S., Kankanhalli, M., Zhao, Q. (2014) Predicting human gaze beyond pixels. *Journal of Vision*. 14(1):28. Available from: doi: 10.1167/14.1.28.