A Deep Learning Framework for Visual Attention Prediction and Analysis of News Interfaces

Matthew Kenely
Dept. of AI
University of Malta
Msida, Malta
matthew.kenely@ieee.org

Dylan Seychell
Dept. of AI
University of Malta
Msida, Malta
dylan.seychell@um.edu.mt

Carl James Debono
Dept. of Comm. & Comp. Eng.
University of Malta
Msida, Malta
c.debono@ieee.org

Chris Porter
Dept. of Comp. Info. Systems
University of Malta
Msida, Malta
chris.porter@um.edu.mt

***Abstract*—News outlets' competition for attention in news interfaces has highlighted the need for demographically-aware saliency prediction models. Despite recent advancements in saliency detection applied to user interfaces (UI), existing datasets are limited in size and demographic representation. We present a deep learning framework that enhances the SaRa (Saliency Ranking) model with DeepGaze IIE, improving Salient Object Ranking (SOR) performance by 10.7%. Our framework optimizes three key components: saliency map generation, grid segment scoring, and map normalization. Through a two-fold experiment using eye-tracking (30 participants) and mouse-tracking (375 participants aged 13–70), we analyze attention patterns across demographic groups. Statistical analysis reveals significant age-based variations ($p < 0.05$, $\epsilon^2 = 0.042$), with older users (36–70) engaging more with textual content and younger users (13–35) interacting more with images. Mouse-tracking data closely approximates eye-tracking behavior (sAUC = 0.86) and identifies UI elements that immediately stand out, validating its use in large-scale studies. We conclude that saliency studies should prioritize gathering data from a larger, demographically representative sample and report exact demographic distributions.**

***Index Terms*—Computer vision, saliency prediction, eye-tracking, visual attention, AI news analysis, user interface design**

Citation: M. Kenely, D. Seychell, C. J. Debono and C. Porter, "A Deep Learning Framework for Visual Attention Prediction and Analysis of News Interfaces," *2025 IEEE Conference on Artificial Intelligence (CAI)*, Santa Clara, CA, USA, 2025, pp. 278-283, doi: 10.1109/CAI64502.2025.00050.

This work is part of the project 'AI-Driven Media Representation Analysis for Social Equity' (AIRAS). It is financed by Xjenza Malta and The Malta Digital Innovation Authority for and on behalf of the Foundation for Science and Technology through the FUSION: R&I Research Excellence Programme.

I. Introduction

The growing demand for a human visual system model to predict gaze behavior and ensure accountable, user-friendly user interfaces (UI) in news websites has highlighted the potential of saliency, a subfield of computer vision [1], [2].

Although saliency prediction has been applied to user interfaces [3]–[5], existing datasets are smaller than those used to train state-of-the-art models for traditional photographs [6]–[8]. Moreover, prior studies often feature narrow demographics [6], [9], [10], small sample sizes [6], [9]–[11], or fail to report participant demographics in detail [4], [6], [10]–[12]. Through this study, we highlight the importance of precise demographic reporting in data-driven saliency research.

This study makes three primary contributions:

1) The optimization of an existing saliency ranking framework (SaRa), which can generate the ranks of elements in an interface by using any saliency model as a backbone and passing element masks as input.
2) The curation of a demographically diverse dataset in a typical UI A/B-testing evaluation context, which captures attention shifts in news websites. Gaze data was gathered through an eye-tracking experiment (n=30) and a mouse-tracking experiment (n=375). The exact demographic distribution of the participants is reported.
3) Statistical analysis showing significant age-based differences in visual attention patterns, demonstrating how AI systems must account for demographic diversity to create inclusive interfaces that serve diverse needs.

II. Related Work

A. Demographic Representation in Saliency Applications

Recent applications of automatic saliency detection to UI remain limited. Gupta *et al.* [3] developed a deep learning model for saliency prediction on mobile UI elements, collecting gaze data from 111 participants (aged 19–46) without reporting gender or detailed age distribution. Similarly, Leiva *et al.* [4] analyzed gaze data from 30 participants (average age 25.9) on 193 mobile UIs, though specific age distribution was not provided. Shen *et al.* [9] and Jiang *et al.* [12] studied 11 and 62 participants respectively, with Jiang reporting gender

(23 males, 43 females) but lacking precise age data. These datasets, though valuable, are small and demographically narrow, potentially biasing gaze predictions.

This issue extends beyond UI studies. Widely used saliency datasets (SALICON [6], MSRA [11], MIT1003 [10]) also rely on small participant groups (< 16) without reporting age distributions, limiting model generalizability and potentially compromising safety.

B. Mouse-tracking as a Complement to Eye-tracking

While eye-tracking technology can be considered more sophisticated, mouse-tracking provides complementary insights. Jiang *et al.* [6] demonstrated that mouse-tracking closely approximates eye-tracking, achieving an sAUC of 0.86 compared to 0.89 for eye-tracking, outperforming saliency models (sAUC < 0.8). This highlights mouse-tracking's potential for large-scale studies with diverse participants. In this study, we leverage mouse-tracking to collect large-scale data and evaluate its applicability in a UI context.

III. Saliency Ranking Framework

A. Original Saliency Ranking Model

Seychell and Debono [13] introduced SaRa, a framework that segments images and ranks segment saliency using any saliency model as a backbone. SaRa divides the input image into a $k \times k$ grid G, generating a saliency map where each segment s is scored based on entropy (H), center bias (CB), and optional depth values (DS):

$$S_s = w_H \cdot H_s + w_{CB} \cdot CB_s + w_{DS} \cdot DS_s \quad (1)$$

where w_H, w_{CB} and w_{DS} are the weights assigned to entropy, center bias and depth, respectively.

B. Proposed Saliency Ranking Model

1) Optimization σ – Saliency Map Generator: We replace the saliency map generator in the original framework proposed by Itti *et al.* [14] with DeepGaze IIE [15]. The MIT/Tuebingen Saliency Benchmark [16] (Table I) shows that DeepGaze IIE outperforms Itti's method and all other benchmarked techniques across all metrics on the MIT300 dataset [17].

TABLE I
PERFORMANCE COMPARISON OF ITTI'S MODEL AND DEEPGAZE IIE ON MIT/TUEBINGEN SALIENCY BENCHMARK METRICS. METRICS WITH ↓ INDICATE THAT LOWER VALUES ARE BETTER.

Model	**IG**	**AUC**	**sAUC**	**NSS**	**CC**	**KLDiv ↓**	**SIM**
Itti *et al.*	N/A	0.54	0.54	0.41	0.13	1.50	0.34
DG IIE	1.07	**0.88**	**0.79**	**2.53**	**0.82**	**0.35**	**0.70**

2) Optimization ϵ – Grid Segment Saliency Score Equation: To better leverage DeepGaze IIE's capabilities, we assign greater weight to the direct pixel values (labeled SS_s) in the generated saliency map. A revised saliency score formula is proposed. For this study, score component weights are set to 1.

$$S_s = w_H \cdot H_s + w_{SS} \cdot SS_s + w_{CB} \cdot CB_s + w_{DS} \cdot DS_s \quad (2)$$

C. Optimization ν – Saliency Map Normalization

Saliency maps may contain noise in non-salient regions that can inflate entropy values, particularly with high bit depths. We include a post-processing step in the pipeline, which applies a 31×31 Gaussian filter, normalizes to $[0, 255]$, and reduces bit depth from 2^8 to 2^5 by dividing values by 8. This approach minimizes noise and results in 32 discrete saliency levels. Hyperparameters were optimized on a 482-image subset of MS-COCO (Figure 1).

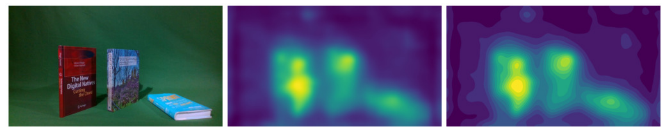

Fig. 1. From left to right: Original image, saliency map from DeepGaze IIE, and map after 31×31 Gaussian filter and bit depth normalization.

D. Evaluation

Spearman's rank correlation coefficient (SRCC) measures the strength of a monotonic relationship between two variables [18]. It is well-suited for saliency ranking tasks and frequently used in related work [19]–[21]. Unlike linear correlation, SRCC detects correlations in relative rank order.

SRCC returns $\rho \in [-1, 1]$, where 1 indicates perfect rank order, -1 indicates complete contrast and ρ is calculated as:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (3)$$

where $\sum d_i^2$ is the sum of squared rank differences, and n is the sample size.

The Salient Object Ranking (SOR) metric, introduced by Islam *et al.* [22], normalizes SRCC to $[0, 1]$ for clearer interpretation. The optimized SaRa framework will be quantitatively evaluated for saliency ranking on a dataset combining MS-COCO object masks and SALICON fixation sequences using SOR, following the approach in [23].

E. Discussion

TABLE II
RESULTS OF THE QUANTITATIVE EXPERIMENT IN [23] COMPARING STATE-OF-THE-ART SALIENCY MODELS (AVERAGE WEIGHTED BY IMAGES USED) AND SARA WITH COMBINATIONS OF OPTIMIZATIONS IN ORDER PERFORMANCE IMPROVEMENT MAGNITUDE.

Model	**SOR ↑**	**#Images used ↑**
RSDNet	0.728	2418
S4Net	0.891	1507
BASNet	0.707	2402
CPD-R	0.766	2417
SCRN	0.756	2418
Siris *et al.*	0.792	2365
Average	0.765	2278
Original SaRa [13]	0.654	2347
SaRa + ν	0.670	2347
SaRa + ϵ	0.685	2347
SaRa + σ	0.714	2347
SaRa + $\epsilon\sigma$	0.715	2347
SaRa + $\epsilon\sigma\nu$	0.718	2347
SaRa + $\epsilon\sigma\nu$, $k = 30$	0.724	2347

Table II demonstrates that each optimization incrementally improved SOR performance. Notably, the Grid Segment

Saliency Score Equation enhanced performance even when using Itti's model, while DeepGaze IIE provided the most substantial boost. A segment grid size of $k = 30$ balanced performance with computational efficiency ($\mathcal{O}(k^2)$ complexity). Applying all optimizations achieved a 10.7% SOR increase over the original technique, reaching performance comparable to state-of-the-art models.

IV. Methodology

Our approach is grounded in human-centered AI principles, recognizing that visual attention modeling must account for demographic diversity to create inclusive systems [24], [25]. We designed our experiments to capture and quantify demographic variations in visual attention explicitly [26]. Data and implementation available on GitHub[1].

A. Gaze Dataset

The dataset comprises 10 pairs of news website interfaces, selected to represent diverse content consumption patterns [27] while controlling for interface design variables. Seven major Maltese news outlets (Times of Malta, Lovin' Malta, Illum, The Malta Independent, Malta Today, The Shift and TVM) are included to represent a wide range of UI design. Each pair consists of a control version containing distracting elements (such as advertisements) and an experimental version with these elements removed. Differences between desktop and mobile versions are also observed. The elements which change between versions, termed Areas of Interest (AOIs), highlight the impact of distractions on gaze attention.

The dataset comprises 10 pairs of news website interfaces in desktop and mobile forms, selected to represent diverse content consumption patterns [27]. This design choice reflects the human-centered understanding that interface elements affect different demographic groups uniquely [28], requiring evaluation frameworks that can account for these variations.

B. Eye-tracking Experiment

This experiment established a baseline for comparing gaze and mouse-tracking behaviors. Thirty participants were split into control and experimental groups, viewing interfaces with either highly salient or neutral elements. Gaze data was recorded using a GazePoint eye-tracker at 60 Hz, with 9-point calibration for accuracy.

Participants viewed 10 interfaces for 10 seconds each, in random order to minimize exposure bias. Followed by a questionnaire on demographics (age, gender, or "rather not say") and awareness of distracting elements. While only 5 participants were female, statistical analysis showed minimal gender influence on gaze patterns. Ages ranged from 19 to 26, potentially biasing results and emphasizing the need for a more age-diverse dataset (see Table IV).

[1]https://github.com/matthewkenely/framework-attention-news

C. Mouse-tracking Experiment

This online experiment engaged 375 participants, offering insights into demographic influences on attention. Participants viewed 10 news interfaces, interacting by hovering or clicking on elements of interest. Mouse movements and clicks/taps were tracked, generating attention heatmaps.

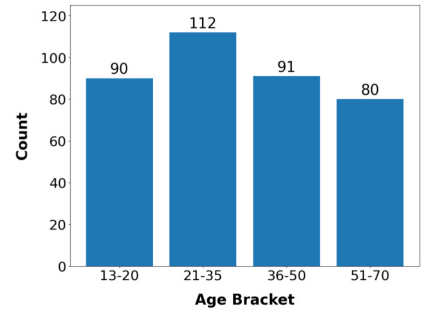

Fig. 2. Age distribution in the mouse-tracking experiment, binned into 4 groups.

Participants provided their age and gender (or "Rather not say") before assignment to control or experimental groups. Each group viewed a unique shuffled sequence to mitigate bias. Desktop users hovered over a central dot before each image for standardization. Mouse data was tracked using JavaScript's MouseEvent API and stored in JSON on Firebase Cloud.

Among participants, 64% were female (131 male, 240 female, 3 other, 1 rather not say). Mann-Whitney U tests showed that gender had minimal influence on attention patterns ($p > 0.05$ across 90% of interfaces).

The age distribution was structured into four balanced groups, enabling robust Kruskal-Wallis tests to examine age-attention correlations. This approach captures demographic variations essential for developing inclusive AI systems.

Table IV reports:

- p-Value (p): Indicates the probability of observing the test statistic under the null hypothesis (that there is no significant difference in the click/tap location based on the demographic variable). A strong likelihood of statistical significance is assumed at $p < 0.05$.
- Effect Size (ε^2): Measures the variance explained by the grouping variable, indicating practical significance.

V. Evaluation

A. Dataset

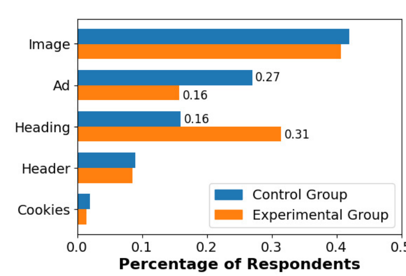

Fig. 3. Responses to "Which type of element do you feel stood out the most?" from the control group (blue) and the experimental group (orange).

TABLE III

RESULTS FROM THE EYE-TRACKING EXPERIMENT COMPARED TO RANK SHIFTS DETECTED BY SARA. EACH RESULT CONCERNS THE AOIS IN THE INTERFACES. BETTER PERFORMANCE, DENOTED IN BOLD, IMPLIES THAT THE INTERFACE WAS LESS DISTRACTING TO PARTICIPANTS. TMI REFERS TO THE MALTA INDEPENDENT.

Image	**Time Viewed % ↓**		**Avg. Fixations ↓**		**Revisitors% ↓**		**Avg. Revisits ↓**		**Avg. 1st View ↑**		**SaRa Rank ↑**	
	CTRL	EXPR	CTRL	EXPR	CTRL	EXPR	CTRL	EXPR	CTRL	EXPR	CTRL	EXPR
Custom (DESKTOP)	17.90	**14.03**	7.73	**6.00**	93.33	**86.67**	3.43	**2.69**	0.57	**1.13**	1.00	**5.00**
Custom (MOBILE)	7.60	**2.56**	3.13	**2.00**	73.33	**13.33**	**2.27**	3.50	2.27	**6.40**	1.00	**2.00**
Times of Malta (1)	**28.16**	32.60	**10.60**	10.73	86.67	86.67	2.77	**1.69**	0.37	**0.50**	1.00	1.00
Lovin' Malta	30.91	**5.62**	12.07	**2.44**	100.00	**13.33**	4.27	**1.25**	0.84	**1.59**	1.00	**4.00**
Illum	8.04	**7.78**	4.83	**3.64**	80.00	80.00	3.08	**2.58**	1.44	**2.18**	4.00	**6.00**
TMI	6.59	**5.65**	3.04	**2.96**	**16.67**	20.00	**2.50**	2.60	2.65	**4.17**	**6.00**	2.00
Malta Today	10.87	**5.67**	5.28	**2.50**	43.33	**40.00**	3.48	**2.33**	**4.61**	3.52	2.00	**4.00**
The Shift	11.28	4.76	5.17	**1.75**	20.00	**0.00**	3.20	**0.00**	5.24	**8.89**	3.00	**4.00**
Times of Malta (2)	**10.89**	13.13	5.05	**5.01**	33.33	**16.67**	3.14	**2.33**	**4.73**	4.12	2.00	**7.00**
TVM	12.06	**10.66**	5.08	4.07	73.33	**46.67**	2.00	**1.71**	2.61	**3.18**	5.00	**7.00**
Average	14.43	**10.25**	5.08	**4.12**	62.00	**40.33**	3.01	**2.07**	2.53	**3.57**	2.60	**4.20**

We observe an apparent reduction in the distraction factor of the AOIs in most interfaces shown to the experimental group. As shown in Table III, GazePoint data reveals that, on average, AOIs were viewed 4.2% less (0.42s), fixated on 0.96 fewer times, revisited by 21.7% fewer participants, revisited 0.94 times less, and first viewed 1.03 seconds later.

The eye-tracking questionnaire further supports this, with Figure 3 showing a significant shift in attention. Participants in the experimental group focused more on article headings (relevant content), while attention to images dropped by 0.3% and to advertisements by 10.3%.

B. Eye-tracking vs Mouse-tracking

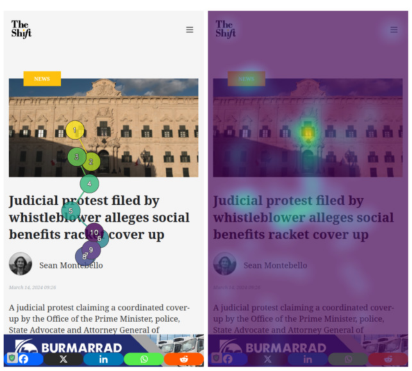

Fig. 4. Gaze location results for the interface "The Shift" shown to the experimental group. Left: average fixation location per second in the eye-tracking experiment, right: heatmap from the mouse-tracking experiment.

Heatmaps reveal that eye-tracking highlights elements that sustain attention over time (e.g., 10 seconds), while mouse-tracking captures what initially stands out on the interface. In this UI context, eye-tracking participants tended to read through screen captures, as shown in Figure 4.

Notably, early fixation locations from eye-tracking align with the top salient regions that are identified by mouse-tracking, highlighting their complementary roles. Eye-tracking suits models for sustained attention, while mouse-tracking better reflects initial visual saliency. Expanding on the work done by Jiang *et al.* in Subsection II-B, we suggest that mouse-tracking should augment, rather than replace, eye-tracking in UI studies.

TABLE IV

RESULTS OF THE AGE KRUSKAL-WALLIS TEST. NULL HYPOTHESIS REJECTIONS ARE UNDERLINED. X: HORIZONTAL GAZE MOVEMENT. Y: VERTICAL GAZE MOVEMENT.

Image	**Group**	***p* (X)**	**ε^2 (X)**	***p* (Y)**	**ε^2 (Y)**
Custom (DESKTOP)	CTRL	0.416	0.000	0.155	0.013
Custom (MOBILE)	**CTRL**	0.137	0.014	**0.038**	**0.031**
Times of Malta 1	CTRL	0.686	0.000	0.710	0.000
Lovin' Malta	CTRL	0.123	0.016	0.209	0.009
Illum	**CTRL**	**0.042**	**0.029**	0.792	0.000
TMI	CTRL	0.395	0.000	0.317	0.003
Malta Today	**CTRL**	0.549	0.000	**0.019**	**0.039**
The Shift	CTRL	0.844	0.000	0.684	0.000
Times of Malta 2	CTRL	0.704	0.000	0.953	0.000
TVM	CTRL	0.165	0.012	0.510	0.000
Custom (DESKTOP)	EXPR	0.480	0.000	0.404	0.000
Custom (MOBILE)	EXPR	0.128	0.000	0.762	0.000
Times of Malta 1	EXPR	0.477	0.015	0.206	0.009
Lovin' Malta	EXPR	0.683	0.000	0.350	0.000
Illum	**EXPR**	0.929	0.000	**0.015**	**0.042**
TMI	EXPR	0.985	0.009	0.213	0.008
Malta Today	EXPR	0.151	0.000	0.146	0.014
The Shift	**EXPR**	**0.019**	**0.040**	0.532	0.000
Times of Malta 2	EXPR	0.543	0.000	0.356	0.001
TVM	EXPR	0.620	0.000	0.619	0.000

C. Demographic Findings

The demographic statistical tests and subsequent qualitative analyses reveal notable differences in gaze tendencies across demographic groups.

Gender played a weaker role in influencing gaze patterns. The only significant result from the Mann-Whitney U Test occurred with the "The Malta Independent" interface shown to the control group. A bias toward the first article image was evident among female respondents, potentially due to their superior ability to recognize faces [29], [30].

Age, on the other hand, was shown to have a much stronger influence on where people were likely to look, with the difference between the gaze tendencies of the age groups being statistically significant in 5 out of the 20 examined interfaces (25%). We observed the following through heatmap analyses:

1) Within headings, the specific words which stood out to participants tended to shift based on their age group, e.g.

Fig. 5. Custom (MOBILE) interface heatmaps. The control group is on the left, and the experimental group is on the right. From left to right for each group: heatmaps from the eye-tracking experiment, heatmaps from the mouse-tracking experiment, saliency maps generated by DeepGaze IIE and the corresponding SaRa ranks.

Fig. 6. The Shift interface heatmaps. Within each pair, the control group is on the left and the experimental group is on the right. Top-left: heatmaps from the eye-tracking experiment; bottom-left: heatmaps from the mouse-tracking experiment; top-right: saliency maps generated by DeepGaze IIE; bottom-right: the corresponding SaRa ranks.

"Judicial" and "whistleblower" in the 36–50 age group and "protest" in the 51–70 age group.

2) Participants in the 21–35 age group were more likely to reject cookies, whereas the 36–70 age group were more likely to accept them;
3) Older demographics (36–70) were more likely to look at news article headings rather than the image;
4) Participants tended to look at images featuring people who are the same age as them;

D. Saliency Ranking Framework

This section discusses the findings from both experiments in comparison to the predictions of the AI framework. We present interfaces from two shift types – content and responsiveness – where discrepancies were found between demographic groups. The potential effects of the narrow demographic range in the eye-tracking experiment (mostly male, ages 19–26) are cross-checked with the demographic findings in Subsection V-C.

1) Custom (Mobile): The results and SaRa ranks for this interface are shown in Figure 5. This custom interface assessed content attention shifts by including and removing the primary cat image as the AOI. The shift was less significant, with the image still receiving considerable attention. DeepGaze IIE and the resulting SaRa ranks accurately captured this, with the ablated AOI receiving a rank difference of only -1, and the attention shift toward the second header, "A Game-Changer in Technology", was also well represented. Demographic heatmap analyses revealed that younger participants were more likely to direct attention to the "Accept" button in the cookies bar. This biased behavior is apparent in the eye-tracking experiment (ages 19–26).

2) The Shift: The results and SaRa ranks for this interface are shown in Figure 6. This interface aimed to assess responsiveness attention shifts by comparing desktop and mobile versions, specifically the ablation of the large ad on the right. In both experiments, attention toward the ad from the experimental group was negligible, with the focus shifting to the main image, heading, and content. DeepGaze IIE and the corresponding SaRa ranks captured this shift well, with the AOIs receiving a rank shift of -1. However, the top bar in the control group, which received no attention, was erroneously assigned rank 2 due to entropy. Demographic heatmap analyses revealed that younger participants were likelier to show attention to the article image. Again, this biased behavior is apparent in the eye-tracking experiment.

VI. Discussion

A. Demographic Findings

Past research has often focused on young adults (21–35), potentially overlooking key differences in attention patterns across younger and older demographics. Future studies should prioritize both participant quantity and diversity. A large sample size alone does not guarantee representativeness if it does not account for demographic variations such as age, gender, and digital literacy.

Training data must reflect user diversity for AI-based UI evaluation tools to be effective for a broad audience. Models trained mainly on younger, tech-savvy participants [31] may exhibit biases, neglecting the preferences and limitations of older or less tech-savvy users [32]. Participant recruitment should be methodical, using stratified sampling to ensure accurate representation of diverse groups based on regional and national demographic data. Depending on the application, this may involve narrowing or broadening the participant base. General-purpose interfaces require diverse representation to ensure inclusivity.

B. Experiments and the Ranking Framework

The AI framework composed of DeepGaze IIE and SaRa performed well in capturing attention across control and experimental groups in various interface designs. As per Table III, it excelled at predicting attention shifts, especially regarding AOIs. However, DeepGaze IIE struggled with interpreting semantic meaning in images and capturing saliency for distracting images, which was expected since it was trained on traditional photographs rather than user interfaces. This limitation was mitigated by the entropy component in Equation (2), which favors large elements.

VII. Conclusion

This study proposed a demographically representative dataset to capture attention shifts in responsive interfaces and evaluate the SaRa saliency ranking framework.

The experiment revealed that eye-tracking captures sustained attention while mouse-tracking reflects immediate attention—a distinction critical for training saliency models.

Demographic analysis emphasized that age significantly influences attention patterns, highlighting the necessity of representative participant pools when developing datasets. A transparent approach to dataset curation is essential for creating generalizable AI-based UI evaluation tools that encourage user-centric design.

Future research can expand on this study by incorporating a more nuanced understanding of cognitive variations and treating demographic factors as intersectional characteristics rather than discrete variables.

References

[1] D. Seychell, G. Hili, J. Attard, and K. Makantatis, "Ai as a tool for fair journalism: Case studies from malta," in *2024 IEEE Conference on Artificial Intelligence (CAI)*. IEEE, 2024, pp. 127–132.

[2] B. Oparaugo, "Relevance of photographs in newspaper journalism," *Central Asian Journal of Social Sciences and History*, vol. 2, no. 10, pp. 96–107, 2021.

[3] P. Gupta, S. Gupta, A. Jayagopal, S. Pal, and R. Sinha, "Saliency prediction for mobile user interfaces," in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2018, pp. 1529–1538.

[4] L. A. Leiva, Y. Xue, A. Bansal, H. R. Tavakoli, T. Köroðlu, J. Du, N. R. Dayama, and A. Oulasvirta, "Understanding visual saliency in mobile user interfaces," in *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '20)*, 2020, pp. 1–12.

[5] K. Krafka, A. Khosla, P. Kellnhofer, H. Kannan, S. Bhandarkar, W. Matusik, and A. Torralba, "Eye tracking for everyone," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2176–2184.

[6] M. Jiang, S. Huang, J. Duan, and Q. Zhao, "SALICON: Saliency in context," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1072–1080.

[7] Y. Li, X. Hou, C. Koch, J. M. Rehg, and A. L. Yuille, "The secrets of salient object segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 280–287.

[8] L. Wang, H. Lu, Y. Wang, M. Feng, D. Wang, B. Yin, and X. Ruan, "Learning to detect salient objects with image-level supervision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 136–145.

[9] C. Shen and Q. Zhao, "Webpage saliency," in *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part VII 13*. Springer, 2014, pp. 33–46.

[10] T. Judd, K. Ehinger, F. Durand, and A. Torralba, "Learning to predict where humans look," in *2009 IEEE 12th international conference on computer vision*. IEEE, 2009, pp. 2106–2113.

[11] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H.-Y. Shum, "Learning to detect a salient object," *IEEE Transactions on Pattern analysis and machine intelligence*, vol. 33, no. 2, pp. 353–367, 2010.

[12] Y. Jiang, L. A. Leiva, H. Rezazadegan Tavakoli, P. RB Houssel, J. Kylmälä, and A. Oulasvirta, "Ueyes: Understanding visual saliency across user interface types," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023, pp. 1–21.

[13] D. Seychell and C. J. Debono, "Ranking regions of visual saliency in rgb-d content," in *2018 International Conference on 3D Immersion (IC3D)*. IEEE, 2018, pp. 1–8.

[14] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.

[15] A. Linardos, M. Kümmerer, O. Press, and M. Bethge, "DeepGaze IIE: Calibrated prediction in and out-of-domain for state-of-the-art saliency modeling," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 12 919–12 928.

[16] M. Kümmerer, Z. Bylinskii, T. Judd, A. Borji, L. Itti, F. Durand, A. Oliva, and A. Torralba, "Mit/tübingen saliency benchmark." [Online]. Available: https://saliency.tuebingen.ai/

[17] T. Judd, F. Durand, and A. Torralba, "A benchmark of computational models of saliency to predict human fixations," in *MIT Technical Report*, 2012.

[18] P. Sedgwick, "Spearman's rank correlation coefficient," *Bmj*, vol. 349, 2014.

[19] W. Wang, P. Jones, and D. Partridge, "A comparative study of feature-salience ranking techniques," *Neural Computation*, vol. 13, no. 7, pp. 1603–1623, 2001.

[20] M. Kalash, M. A. Islam, and N. D. Bruce, "Relative saliency and ranking: Models, metrics, data, and benchmarks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 1, pp. 204–219, 2019.

[21] Y. Niu, J. Chen, and W. Guo, "Meta-metric for saliency detection evaluation metrics based on application preference," *Multimedia Tools and Applications*, vol. 77, pp. 26 351–26 369, 2018.

[22] M. A. Islam, M. Kalash, and N. D. Bruce, "Revisiting salient object detection: Simultaneous detection, ranking, and subitizing of multiple salient objects," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7142–7150.

[23] A. Siris, J. Jiao, G. K. Tam, X. Xie, and R. W. Lau, "Inferring attention shift ranks of objects for image saliency," in *Proceedings of*

the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 12 133–12 143.

[24] B. Shneiderman, "Human-centered artificial intelligence: Reliable, safe & trustworthy," *International Journal of Human–Computer Interaction*, vol. 36, no. 6, pp. 495–504, 2020.

[25] K. Holstein, J. Wortman Vaughan, H. Daumé III, M. Dudik, and H. Wallach, "Improving fairness in machine learning systems: What do industry practitioners need?" in *Proceedings of the 2019 CHI conference on human factors in computing systems*, 2019, pp. 1–16.

[26] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan, "A survey on bias and fairness in machine learning," *ACM computing surveys (CSUR)*, vol. 54, no. 6, pp. 1–35, 2021.

[27] Z. Wang, K. Qinami, I. C. Karakozis, K. Genova, P. Nair, K. Hata, and O. Russakovsky, "Towards fairness in visual recognition: Effective strategies for bias mitigation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 8919–8928.

[28] J. Buolamwini and T. Gebru, "Gender shades: Intersectional accuracy disparities in commercial gender classification," in *Conference on fairness, accountability and transparency*. PMLR, 2018, pp. 77–91.

[29] A. Herlitz and J. Lovén, "Sex differences and the own-gender bias in face recognition: A meta-analytic review," *Visual Cognition*, vol. 21, no. 9-10, pp. 1306–1336, 2013.

[30] A. M. Proverbio, "Sex differences in social cognition: The case of face processing," *Journal of neuroscience research*, vol. 95, no. 1-2, pp. 222–234, 2017.

[31] A. Dingli and D. Seychell, "The new digital natives," *JB Metzler: Stuttgart, Germany*, 2015.

[32] Y. E. Alkali and Y. Amichai-Hamburger, "Experiments in digital literacy," *CyberPsychology & Behavior*, vol. 7, no. 4, pp. 421–429, 2004.