Could AI eliminate the need for human eye-tracking testing in advert evaluation?

Katariina Pajuranta[a]*, Jarmo J. Ahonen[b], Ulla Hakala[c]

aEngineering and Business, Turku University of Applied Sciences, Turku, Finland; bEngineering and Business, Turku University of Applied Sciences, Turku, Finland; cTurku School of Economics, University of Turku, Turku, Finland

*Katariina Pajuranta, katariina.pajuranta@turkuamk.fi, Turku University of Applied Sciences, Lemminkäisenkatu 30, 20520, Turku Finland

Could AI eliminate the need for human eye-tracking testing in advert evaluation?

In today's increasingly digital and information-saturated environment, understanding attention is fundamental for advertising success. In this study we compared maps generated from eye tracking measurements between human participants and AI-generated maps. The comparison between the AI-generated maps and the map generated from the human participants' visual data implied that the AI maps of Blur Integrated Graphics and Smoothed Blur Integrated Graphics most closely resembled the authentic human visual attention. In their current form, AI-generated saliency maps provide a rough estimate of the features of an ad that people are likely to pay attention to. Human actions are predominantly motivation-driven and human visual attention is oftentimes influenced by top-down processing. Prediction of such goal-oriented consumers' visual behavior makes attention allocation for the AI troublesome. According to our findings, in top-down processing tasks, AI based saliency maps cannot eliminate the need for large-scale empirical eye tracking studies including numerous human participants.

Keywords: Print Advertising; AI Saliency maps; Attention Based Marketing (ABM); Human attention; Eye tracking

Introduction

Capturing human attention in advertising design is crucial for effective communication. No attention means no ad processing and hence, no brand communication effects. After exposure, attention is the next response necessary in ad processing (Rossiter et al., 2018). But what elements of an ad gain and hold attention? Attention-based marketing (ABM) is a marketing sub-discipline focused on understanding consumers' attention and cognitive processes. ABM operates on the assumptions that attention is a crucial aspect of consumer behavior, necessitating eye-tracking as a research method, and it plays a pivotal role in exploration, search, and choice (Orquin & Wedel, 2020). Eye-tracking measurements give precise and objective information about human attention. Eye behavior, encompassing gaze direction,

fixation, saccades, blinks, and pupil dilation, serve as a window into individuals' higher cognitive functions. (Holmqvist & Nyström, 2011; Wedel & Pieters, 2008). Fixations—periods of relative gaze stability—are indicative of attention (Barbierato et al., 2023; Du, 2016; Pieters & Warlop, 1999; Puurtinen et al., 2021; Wolf & Lappe, 2021).

However, eye tracking studies are expensive due to specific equipment and facilities, participant-related costs, as well as required expertise. Seeking cost-effective alternatives, researchers may be tempted to consider substitutes to laborious human eye-tracking testing. Saliency maps, generated by AI-based systems, can predict image parts likely to attract human attention and thereby, offer an avenue for exploring visual attention without traditional constraints (Treue, 2003). Could AI eliminate the need for human eye-tracking testing in advert evaluation? In this paper, we aim to answer this question, the purpose being to compare AI-generated saliency maps predicting human attention to eye-tracking visualizations of human participants.

AI has been acknowledged as a beneficial technological tool in the context of eye tracking, and recent studies have proposed neural networks as a solution for several purposes. These include easier eye-tracker calibration without active participation by human test subjects (Chang et al., 2019). Proposed benefits also include human scan-path prediction – visual mappings of participant viewing behaviors (Assens et al., 2018), saliency-map based tools for webpage design (Corradini et al., 2022) and prediction of the subsequent fixation of humans (Kadner et al., 2023; Mizuno et al., 2021). However, there is still a lack of knowledge of systematic exploration regarding the applicability of salience maps in advertising.

Methodology and Material

Stimulus preparation for human eye tracking

For this study, the researchers crafted a mock-up ad promoting an energy drink NOCCO,

shown in Figure 1. NOCCO is a functional drink, developed and first launched in Sweden. Today it is available in over 30 markets. The company graciously granted permission for the creation of the ad. The advertisement was created corresponding to the brand's key ad elements, including a textual brand logo, the advertised product, and a human actor.
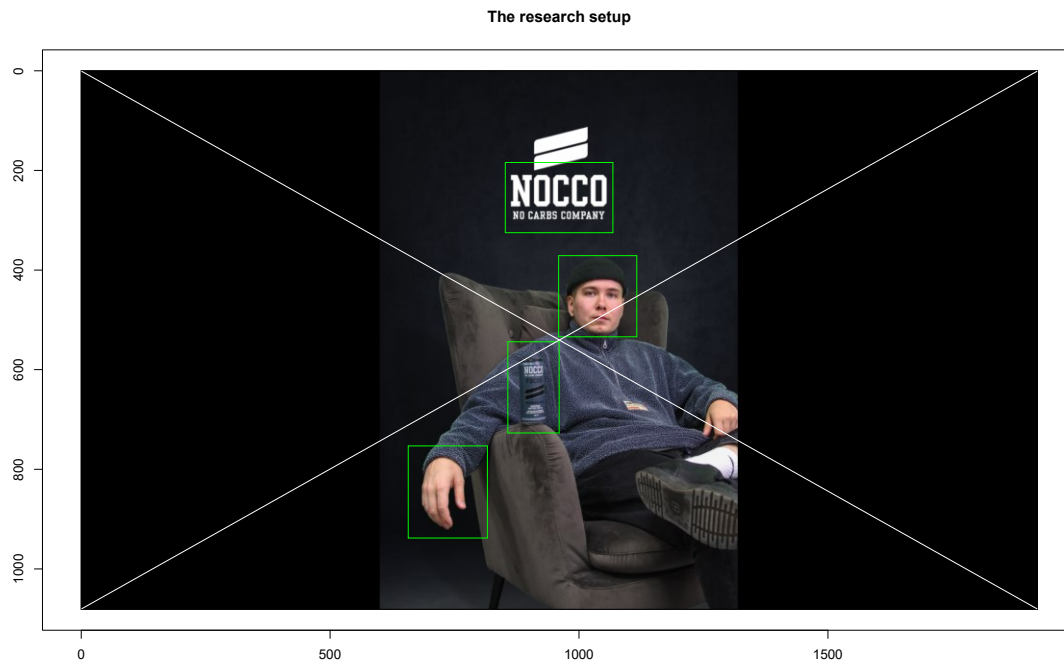

Figure 1: The stimulus used in the study. The white lines signify the central point of the ad and the AOIs are indicated with green rectangles.

Prior to commencing the eye tracking experiment with human participants, the ad was partitioned into four areas of interest (AOIs) marked with green squares: Product, Brand, Face and Hand. This was made to facilitate the aggregation of fixations and their analysis. The white dividing lines indicate the central point of the image, considered in the design and positioning of elements in the advertisement. Prior literature shows that in screen-based eye tracking experiments participants preferentially first fixate on locations in the middle, an observation known as central fixation bias (Tatler, 2007; van Heusden et al., 2023; Wolf & Lappe, 2021). The participants were asked to search for the advertised product within the ad,

and the advertised product was intentionally positioned at the center to aid in its identification during this task.

Hypothesis formulation

The hypotheses for the testing were:

> H1. Central fixation bias leads test subjects to initially fixate on the product or the face.
>
> H2. The test subjects pay the most attention to the advertised product due to top-down processing.
>
> H3. AI saliency maps cannot predict human top-down guided visual attention.

Hypothesis H1 is grounded on previous findings (Tatler, 2007; van Heusden et al., 2023; Venkatraman et al., 2014; Wolf & Lappe, 2021) suggesting that the center is an optimal spot for early information processing. Prior eye tracking studies have shown participants' tendency to direct their gaze preferentially toward the centre of the visual field, as opposed to peripheral areas (Fehd & Seiffert, 2010; Tatler, 2007). Based on this, participants are likely to fixate most on the face of the human character or product in the proximity to the image center and not on the features further away. In addition, human faces are particularly prospective visual cues capturing attention. This phenomenon has a potential of diverting attention from other visual elements which is attributable to the role faces play in the perception and interpretation of others' characteristics, personalities, intentions, and emotions (Adil et al., 2018; Bindemann et al., 2005; Emery, 2000; Ju & Johnson, 2010; Menon et al., 2016).

Hypotheses H2 and H3 were formulated based on top-down processing theory referring to situations in which attention is guided by the objectives and preferences which originate from the goals of the person influencing where they look (Corbetta & Shulman,

2002; Itti et al., 2005; Palermo & Rhodes, 2007; Xie et al., 2023). This very human characteristic impacts everyday life when consumers' motivation guides their attention and navigation in consumption environments. In order to engage the participants in such top-down processing, they were motivated to search for the advertised product in the advertisement.

Experiment procedure

The research protocol consisted of three steps: Human eye tracking testing, AI testing and comparison between human-testing and AI testing.

Human eye tracking

Prior to commencing the eye tracking experiment, each participant went through a 9-point calibration process to ensure precise measurements. The participants received on-screen and verbal instructions for a search task:

"You will be shown an advertisement. There is a product in the advertisement. Try to find the product and after finding the product, click the mouse. Concentrate only on the product."

The verbatim search task instruction endeavored to prompt participants with task-induced motivation and engage the participants in a top-down strategic rationale during the ad viewing. The goal of the search task was to render all other visual stimuli task-irrelevant and eliminate their attention allocation (Ambinder & Simons, 2005). After the instructions, participants were individually exposed to the advertisement stimulus, which was presented on a 23-inch monitor. The participants included 50 students between 20 and 25 years from the Turku University of Applied Sciences. The experimental procedures strictly adhered to the ethical standards outlined by the Research Ethics Committee of Turku University of Applied Sciences for human subject research (Ethical review number 08032022-1). Participation was

entirely voluntary and contingent upon informed consent, with no compensation offered in return.

Ocular movements were recorded using a Tobii Pro X3-120 eye tracker with a sampling rate of 120 Hz. The analysis utilized Tobii Pro Lab version 1.181.37603, configured with the Tobii I-VT standard fixation detection algorithm. Visual stimulus was displayed on a Dell P2319H monitor connected to Dell Precision 7560 laptop. The measurements are reported according to (Dunn et al., 2023) for research involving eye tracking and human participants in Appendix 1.

AI saliency map generation

The mock-up advertising image was analyzed by computer vision. Methods and frameworks utilized in the generation of saliency maps included Guided Integrated Gradients (Kapishnikov et al., 2021), XRAI in its normal and fast variant (Kapishnikov et al., 2019), SmoothGrad (Smilkov et al., 2017), Gradients (Erhan et al., 2009; Simonyan et al., 2013), Guided Backpropagation (Springenberg et al., 2014), Integrated Gradients (Sundararajan et al., 2017), Grad-CAM (Selvaraju et al., 2017), Blur IG (Xu et al., 2020), and combinations of the methods. The code used for generating the maps was from Google PAIR (Google PAIR, 2023) and the term "Vanilla" means no alternations to the original algorithm. Only small changes were made to the default settings of Google PAIR code. The software-based saliency maps were generated with the library provided by (Google PAIR, 2023).

Comparison between human eye tracking and AI saliency maps

The generated saliency maps were carefully assessed by a panel of experts, comprising two marketing scholars and fourteen junior researchers. Following the evaluation, Blur Integrated Gradients and Smoothgrad Blur Integrated Graphics saliency maps were selected for the test

phase, as they deemed most likely to correspond to human visual attention.

Results

Human testing

Altogether 30 participants fixated first on the face (61%, 95% CI [47%, 74%] and 36 participants fixated first on the Face or the Product AOIs (73%, 95% CI [60%, 84%]) which both are the closest to the central point of the image. Confidence intervals were calculated with the Wilson-method, as recommended by (Brown et al., 2001). The Kendall's coefficient of concordance W-measure was 0.57 and the related p-value below 0.001 (2.04e-18). The finding of the locations of the first fixations supports the Central Fixation Bias (p = 0.78, 95% CI [0.61, 0.89] for confidence level 0.95 with multinomial analysis using the Goodman method and confirms the validity of Hypothesis 1.

The locations of the human fixations are visualized as a heatmap in Figure 2. The predominance of red hues signals regions where gaze was most persistent, while yellow shades represent areas of secondary levels of attention and white squares mean a low level of attention. An area with no squares received no or insignificant attention.
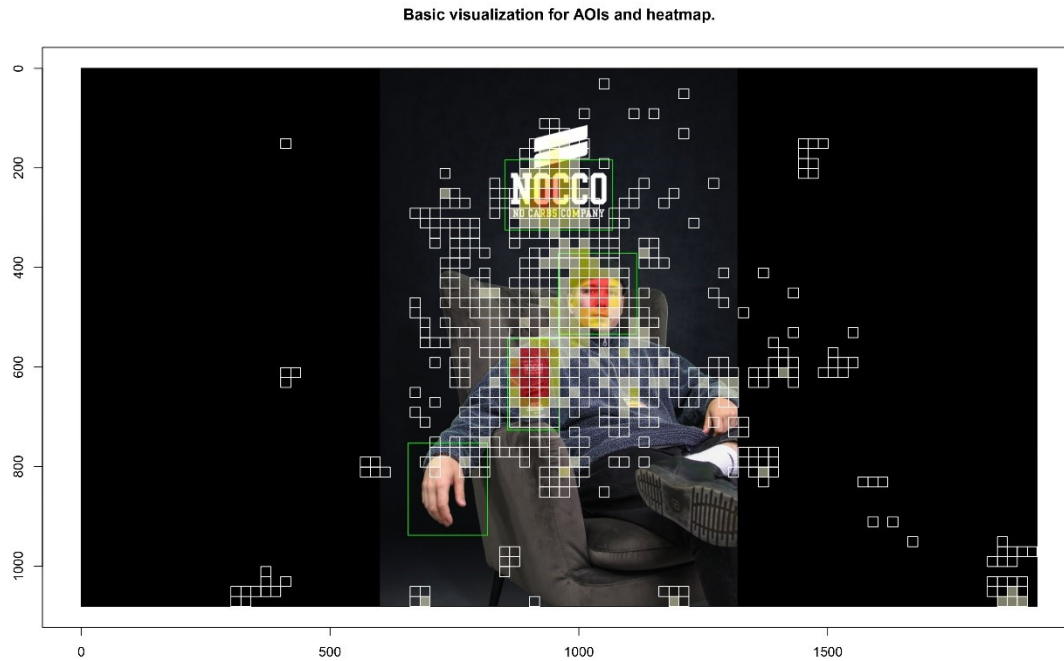
Figure 2. The heatmap visualization generated from human gaze.

The Area of Interest designated for the advertised product displays the most concentrated red center in Figure 2, indicating the highest sum of fixations. All participants fixated on the advertised product during the viewing, clearly influenced by the given search task. The results align with Hypothesis 2, confirming that top-down processing leads the test subjects to pay the most attention to the advertised product.

Comparison between human eye tracking and AI saliency maps

The AI-generated Blur Integrated Gradients and Smoothgrad Blur Integrated Graphics saliency maps are depicted in Figure 3.
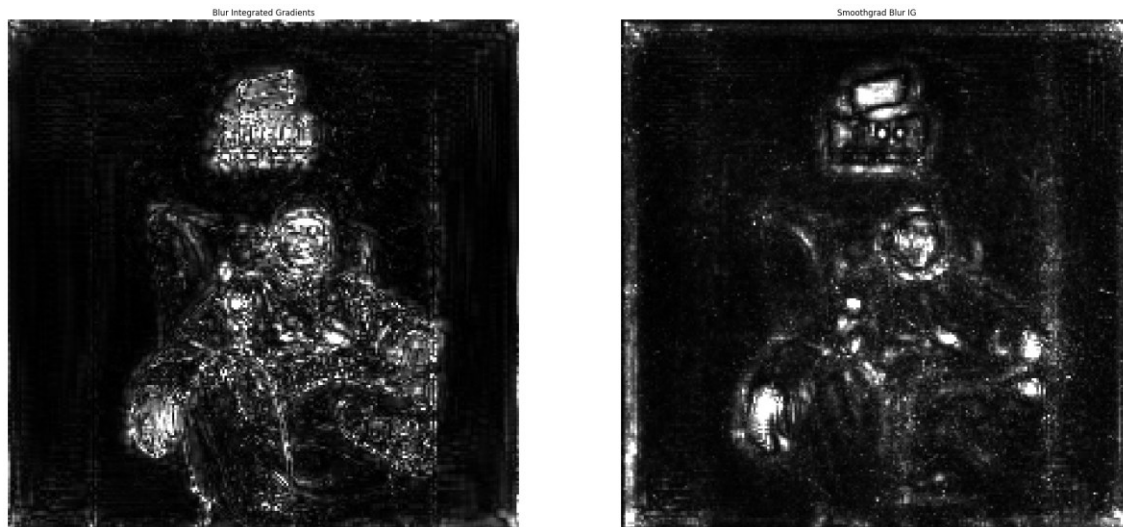
Figure 3. Blur Integrated Gradients on the left and Smoothgrad Blur Integrated Graphics saliency map on the right

In the saliency maps, the regions anticipated to attract the most attention are highlighted in white. Compared to the human gaze (see Figure 2), both AI-generated saliency maps underscored visual attention directed towards the human face, the actor's right hand, and the advertised product. However, the attention towards the advertised product is underestimated whereas the graphical part of the logo is unduly overestimated (Figure 3). Smoothgrad Blur Integrated Graphics exhibited additionally two significant errors: it neglected attention towards the textual component of the logo, and it inaccurately predicted attention allocation to the left hand of the actor. The map generated by Blur Integrated Gradients demonstrated greater accuracy but exhibited shortcomings in underestimating the attention on the advertised product and emphasizing the actor's foot. These limitations provide support for our Hypothesis 3, indicating that AI saliency maps are unable to accurately predict top-down processing of human attention.

Discussion

The findings suggest that further refinement is necessary before software-generated saliency

maps could substitute eye tracking measurements of human attention. While the best AI saliency maps correctly identified some areas that capture human attention, they still fall short in accurately predicting real human attention to ad elements. This limitation was evident in a test situation predicting task-oriented visual behavior requiring top-down processing, characteristics of everyday human actions. At least without explicit education, AI is incapable of considering the human characteristics and as for now, researchers cannot account for prediction by AI to replace human eye movement studies.

The visual human visual system and AI vision differ remarkably, and the shortcomings of AI prediction may be due to the human foveated nature of the visual system and the central fixation bias. Human visual attention operates by fixating on small areas at a time, allowing for high scrutiny in the area, and saccadic movements relocate the vision quickly. In addition, human vision is characterized by central bias, exhibiting a tendency to initiate viewing of the images from the middle of the screen. In contrast, AI relies on algorithm image scanning, and part of its functionality involves recognizing colour contrasts and shapes (Xu & Zhang, 2015).

This study was limited to 50 participants, a number consistent with studies on direct and indirect print advertisements, such as the one conducted by Simola et al. (2020) with a sample size of 45 participants. Although most of the statistical tests can be used and statistical models built with a reasonably small number of participants (Kwak & Kim, 2017), the statistical power would benefit from a larger number of participants. While expanding the participant pool could enhance the robustness of the findings, it would simultaneously escalate study costs.

Further research is needed to explore the applicability of automated tools generating saliency maps, as they could become highly valuable for advertising practitioners in the future. The rapid development of AI suggests that these models may become sufficiently

advanced for evaluation of the attentional value of advertisements in the coming years.
Additionally, it would be beneficial to investigate whether AI can accurately predict human
attention allocation in a free-viewing scenario of advertisements.

Acknowledgement

Appendices

Appendix 1

The measurement device used in this study was Tobii Pro X3-120 eye-tracker, the
measurement software used was Tobii Pro Lab 1.181.37603. The statistical analysis software
used was R version 4.2.1. The display used was Dell P2319H. The algorithm used for
detecting fixations and other types of eye movement was Tobii I-VT. The raw data generated
by Tobii Pro Lab was used for the analysis.

References

Adil, S., Lacoste-Badie, S., & Droulers, O. (2018). Face presence and gaze direction in print
advertisements: How they influence consumer responses— An eye-tracking study.
Journal of Advertising Research, *58*(4), 443–455. https://doi.org/10.2501/JAR-2018-004

Ambinder, M., & Simons, D. (2005). Attention Capture: The Interplay of Expectations,
Attention, and Awareness. In L. Itti, G. Rees, & J. K. Tsotsos (Eds.), *Neurobiology of
Attention* (pp. 69–75).

Assens, M., Giro-i-Nieto, X., McGuinness, K., & O'Connor, N. E. (2018, September). PathGAN: Visual Scanpath Prediction with Generative Adversarial Networks. *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*.

Barbierato, E., Berti, D., Ranfagni, S., Hernández-Álvarez, L., & Bernetti, I. (2023). Wine label design proposals: an eye-tracking study to analyze consumers' visual attention and preferences. *International Journal of Wine Business Research*. https://doi.org/10.1108/IJWBR-06-2022-0021

Bindemann, M., Burton, A. M., Hooge, I. T. C., Jenkins, R., & De Haan, E. H. F. (2005). Faces retain attention. In *Psychonomic Bulletin and Review* (Vol. 12, Issue 6, pp. 1048–1053). Psychonomic Society Inc. https://doi.org/10.3758/BF03206442

Brown, L. D., Cai, T. T., & DasGupta, A. (2001). Interval Estimation for a Binomial Proportion. *Statistical Science*, *16*(2), 101–133. https://doi.org/10.1214/ss/1009213286

Chang, Z., Matias Di Martino, J., Qiu, Q., Espinosa, S., & Sapiro, G. (2019, October). SalGaze: Personalizing Gaze Estimation using Visual Saliency. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*.

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

Corradini, E., Porcino, G., Scopelliti, A., Ursino, D., & Virgili, L. (2022). Fine-tuning SalGAN and PathGAN for extending saliency map and gaze path prediction from natural images to websites. *Expert Systems with Applications*, *191*, 116282. https://doi.org/https://doi.org/10.1016/j.eswa.2021.116282

Du, P. (2016). *Investigating effects of product visual designs on consumer judgments with the aid of eye-tracking*. Iowa State University.

Dunn, M. J., Alexander, R. G., Amiebenomo, O. M., Arblaster, G., Atan, D., Erichsen, J. T., Ettinger, U., Giardini, M. E., Gilchrist, I. D., Hamilton, R., Hessels, R. S., Hodgins, S., Hooge, I. T. C., Jackson, B. S., Lee, H., Macknik, S. L., Martinez-Conde, S., Mcilreavy, L., Muratori, L. M., … Sprenger, A. (2023). Minimal reporting guideline for research involving eye tracking (2023 edition). *Behavior Research Methods*, 1–7. https://doi.org/10.3758/S13428-023-02187-1/TABLES/1

Emery, N. J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience and Biobehavioral Reviews*, *24*, 581–604. www.elsevier.com/locate/neubiorev

Erhan, D., Bengio, Y., Courville, A., & Vincent, P. (2009). Visualizing higher-layer features of a deep network. *University of Montreal*, *1341*(3), 1.

Fehd, H. M., & Seiffert, A. E. (2010). Looking at the center of the targets helps multiple object tracking. *Journal of Vision*, *10*(4), 1–13. https://doi.org/10.1167/10.4.19

Google PAIR. (2023, February 19). *Saliency Library*. https://github.com/PAIR-code/saliency

Holmqvist, K., & Nyström, M. (2011). *Eye Tracking A Comprehensive Guide To Methods And Measures*. Oxford University Press.

Itti, L., Rees, G., & Tsotsos, J. K. (2005). Neurobiology of Attention. *Neurobiology of Attention*, 1–696. https://doi.org/10.1016/B978-0-12-375731-9.X5000-9

Ju, H. W., & Johnson, K. K. P. (2010). Fashion Advertisements and Young Women: Determining Visual Attention Using Eye Tracking. *Clothing and Textiles Research Journal*, *28*(3), 159–173.

Kadner, F., Thomas, T., Hoppe, D., & Rothkopf, C. A. (2023). Improving Saliency Models' Predictions of the Next Fixation With Humans' Intrinsic Cost of Gaze Shifts. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2104–2114.

Kapishnikov, A., Bolukbasi, T., Viegas, F., & Terry, M. (2019, October). XRAI: Better

    Attributions Through Regions. *Proceedings of the IEEE/CVF International Conference*

    on Computer Vision (ICCV).

Kapishnikov, A., Venugopalan, S., Avci, B., Wedin, B., Terry, M., & Bolukbasi, T. (2021).

    Guided Integrated Gradients: An Adaptive Path Method for Removing Noise.

    Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition

    (CVPR), 5050–5058.

Kwak, S. G., & Kim, J. H. (2017). Central limit theorem: the cornerstone of modern statistics.

    Kja, *70*(2), 144–156. https://doi.org/10.4097/kjae.2017.70.2.144

Menon, R. G. V., Sigurdsson, V., Larsen, N. M., Fagerstrøm, A., & Foxall, G. R. (2016).

    Consumer attention to price in social commerce: Eye tracking patterns in retail clothing.

    Journal of Business Research, *69*(11), 5008–5013.

    https://doi.org/10.1016/j.jbusres.2016.04.072

Mizuno, R., Nakazawa, S., & Nakada, Y. (2021). Simulator of Vehicle-Mounted Camera

    Video to Verify Dynamic Saliency Maps for Predicting Drivers' Gaze Points. *2021 IEEE*

    Symposium Series on Computational Intelligence (SSCI), 1–6.

    https://doi.org/10.1109/SSCI50451.2021.9660098

Orquin, J. L., & Wedel, M. (2020). Contributions to attention based marketing: Foundations,

    insights, and challenges. *Journal of Business Research*, *111*, 85–90.

    https://doi.org/https://doi.org/10.1016/j.jbusres.2020.02.012

Palermo, R., & Rhodes, G. (2007). Are you always on my mind? A review of how face

    perception and attention interact. *Neuropsychologia*, *45*(1), 75–92.

    https://doi.org/10.1016/j.neuropsychologia.2006.04.025

Pieters, R., & Warlop, L. (1999). Visual attention during brand choice: The impact of time pressure and task motivation. *Intern. J. of Research in Marketing*, *16*, 1–16. https://doi.org/https://doi.org/10.1016/S0167-8116(98)00022-6

Puurtinen, M., Hoppu, U., Puputti, S., Mattila, S., & Sandell, M. (2021). Investigating visual attention toward foods in a salad buffet with mobile eye tracking. *Food Quality and Preference*, 1–12. https://doi.org/10.1016/j.foodqual.2021.104290

Rossiter, J. R., Percy, L., & Bergkvist, L. (2018). *Marketing Communications: Objectives, Strategy, Tactics*. SAGE.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual Explanations From Deep Networks via Gradient-Based Localization. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 618–626.

Simonyan, K., Vedaldi, A., & Zisserman, A. (2013). *Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps*. arXiv. https://doi.org/10.48550/ARXIV.1312.6034

Smilkov, D., Thorat, N., Kim, B., Viégas, F., & Wattenberg, M. (2017). *SmoothGrad: removing noise by adding noise*. arXiv. https://doi.org/10.48550/ARXIV.1706.03825

Springenberg, J. T., Dosovitskiy, A., Brox, T., & Riedmiller, M. (2014). *Striving for Simplicity: The All Convolutional Net*. arXiv. https://doi.org/10.48550/ARXIV.1412.6806

Sundararajan, M., Taly, A., & Yan, Q. (2017). Axiomatic Attribution for Deep Networks. In D. Precup & Y. W. Teh (Eds.), *Proceedings of the 34th International Conference on Machine Learning* (Vol. 70, pp. 3319–3328). PMLR. https://proceedings.mlr.press/v70/sundararajan17a.html

Tatler, B. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision*, *7*(14). https://doi.org/10.1167/7.14.4

Treue, S. (2003). Visual attention: The where, what, how and why of saliency. In *Current Opinion in Neurobiology* (Vol. 13, Issue 4, pp. 428–432). Elsevier Ltd. https://doi.org/10.1016/S0959-4388(03)00105-3

van Heusden, E., Olivers, C. N. L., & Donk, M. (2023). The eyes prefer targets nearby fixation: Quantifying eccentricity-dependent attentional biases in oculomotor selection. *Vision Research*, *205*, 108177. https://doi.org/10.1016/J.VISRES.2023.108177

Venkatraman, V., Payne, J. W., & Huettel, S. A. (2014). An overall probability of winning heuristic for complex risky decisions: Choice and eye fixation evidence. *Organizational Behavior and Human Decision Processes*, *125*(2), 73–87. https://doi.org/10.1016/J.OBHDP.2014.06.003

Wedel, M., & Pieters, R. (2008). A Review of Eye-Tracking Research in Marketing. In N. K. Malhotra (Ed.), *Review of Marketing Research* (Vol. 4, pp. 123–147). Emerald Group Publishing Limited. https://doi.org/10.1108/S1548-6435(2008)0000004009

Wolf, C., & Lappe, M. (2021). Salient objects dominate the central fixation bias when orienting toward images. *Journal of Vision*, *21*(8), 23–23. https://doi.org/10.1167/JOV.21.8.23

Xie, W., Lee, M. H., Chen, M., & Han, Z. (2023). Understanding Consumers' Visual Attention in Mobile Advertisements: An Ambulatory Eye-Tracking Study with Machine Learning Techniques. *Journal of Advertising*, 1–19. https://doi.org/10.1080/00913367.2023.2258388

Xu, M., & Zhang, H. (2015). Saliency detection with color contrast based on boundary information and neighbors. *The Visual Computer*, *31*, 355–264.

Xu, S., Venugopalan, S., & Sundararajan, M. (2020, June). Attribution in Scale and Space. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.