Visual Attention Differences Between AI-Generated and Real Images Based on Decision Certainty and Gender

Rui Zhe Ooi
Computer Science
(Artificial Intelligence)
Faculty of Computer Science
University Malaya
W. Persekutuan Kuala Lumpur
23004947@siswa.um.edu.my

Rui Zhe Khor
Computer Science
(Artificial Intelligence)
Faculty of Computer Science
University Malaya
W. Persekutuan Kuala Lumpur
23004890@siswa.um.edu.my

Min Zi Teoh
Computer Science
(Artificial Intelligence)
Faculty of Computer Science
University Malaya
W. Persekutuan Kuala Lumpur
23004985@siswa.um.edu.my

Shin Yen Lee
Computer Science
(Artificial Intelligence)
Faculty of Computer Science
University Malaya
W. Persekutuan Kuala Lumpur
23004922@siswa.um.edu.my

Iman Soffea binti Jamal
Computer Science
(Artificial Intelligence)
Faculty of Computer Science
University Malaya
W. Persekutuan Kuala Lumpur
22002203@siswa.um.edu.my

Jia Xin Low
Computer Science
(Artificial Intelligence)
Faculty of Computer Science
University Malaya
W. Persekutuan Kuala Lumpur
23004985@siswa.um.edu.my

Abstract

Eye-tracking enables a deeper understanding of how individuals visually process complex visual stimuli, such as AI-generated images and real images. This study examines how gaze behavior varies across image authenticity, decision confidence, and gender. Eighty-seven participants viewed six pairs of AI and real images while their eye movements were tracked using a Tobii Pro Fusion eye tracker. Visual attention metrics, including total fixation duration, average fixation duration, fixation count, time to first fixation, and visit duration, were analyzed across three key dimensions: image type (AI vs. Real), decision certainty (certain vs. uncertain), and gender (male vs. female). Results show that participants exhibited significantly longer fixations and more focused gaze patterns on real images compared to AI-generated images, suggesting deeper cognitive processing. Uncertain decisions tended to show higher fixation counts and more exploratory gaze behavior, reflecting higher cognitive load. Gender-based differences were also observed, with both males and females demonstrating longer visit durations across different tasks. These findings provide insights into how individuals engage with visual content of varying authenticity and highlight the importance of considering gender and decision confidence in gaze analysis.

Keywords: eye-tracking, AI-generated images, gaze behavior, decision-making, gender differences, visual attention

1. Introduction

1.1 Gap/Problem/Challenge

As artificial intelligence advances, AI-generated images have flooded increasingly across the world, especially on social media. These synthetic visuals are like double-edged swords, offering creative opportunities, but also pose a risk to safety as users often struggle to distinguish them from real images. Prior studies have utilized eye-tracking to explore various fields, including psychology, user experience, interface design, and marketing and advertising. However, there is limited research that studies human gaze behaviour toward AI-generated and real images, not to mention how uncertainty in judgement or individual differences, such as gender, influence gaze behaviour during authenticity classification tasks.

1.2 Research Questions

1. How does visual attention differ between AI-generated and real images?
2. Does uncertainty in decision-making manifest in visual attention patterns when identifying AI vs real images?
3. Are there gender-based differences in gaze behavior when distinguishing between AI and real images?

1.3 Objectives

1. To identify if there's a difference in gaze behavior between AI and real images
2. To test whether uncertain decisions (e.g., incorrect or low-confidence answers) correlate with different gaze behavior.
3. To compare gaze behavior by gender while distinguishing AI vs. real images.

1.4 Hypothesis for the objectives/analysis

1. Null Hypothesis (H_0) :
   There is no significant difference in gaze patterns when participants view AI-generated images compared to real images.

   Alternative Hypothesis (H_1) :
   There is a significant difference in gaze patterns when participants view AI-generated images compared to real images.

2. Null Hypothesis (H_0) :
   There is no significant difference in total viewing time between participants who answered correctly and those who answered incorrectly.

   Alternative Hypothesis (H_1) :
   There is a significant difference in total viewing time between participants who answered correctly and those who answered incorrectly.

3. Null Hypothesis (H_0) :
   There are no significant differences in at least one measure of gaze behavior (as measured by Total Fixation Duration, Fixation Count, Time to First Fixation, and Total Visit Duration) between male and female participants when distinguishing between AI-generated and real images, potentially varying by image content or specific Areas of Interest (AOIs).

   Alternative Hypothesis (H_1) :
   There are significant differences in at least one measure of gaze behavior (as measured by Total Fixation Duration, Fixation Count, Time to First Fixation, and Total Visit Duration) between male and female participants when distinguishing between AI-generated and real images, potentially varying by image content or specific Areas of Interest (AOIs).

1.5 Contribution

This study contributes to a deeper understanding of visual cognition in the context of AI-generated images. By doing this research using an eye-tracker, an analysis of attention patterns in AI vs real Image classification could be done. It provides results for how different people perceive and interact with synthetic content.

2. Literature Review

2.1 Statistical Analysis of Eye Tracking Data (Krzysztof Krejtz)

The workshop "Statistical Analysis of Eye Tracking Data" by Krzysztof Krejtz, presented at the LEAD ME Summer Training School, has outlined a robust and automated eye-tracking data analysis pipeline with a focus on facial emotion recognition experiments. The workshop introduces an end-to-end methodology using Python and R scripts to streamline tasks such as data preprocessing, metric extraction, and statistical testing. This automation enhances reproducibility and efficiency in complex experimental designs. Key aspects include methods for ensuring tracking accuracy, visual data validation, and overcoming challenges like center-bias, gender balance, and emotion control in stimuli. He also demonstrates how to define Areas of Interest (AOIs) and extract gaze metrics such as fixation counts, durations, and pupil-based cognitive load indicators like the Index of Pupillary Activity (IPA) and K Coefficient, which reveal deeper layers of emotional engagement and attention patterns. These advanced metrics and data handling practices offer a strong methodological foundation for the current study's goals, which are to examine visual attention differences, decision-making uncertainty, and gender-based gaze behavior when viewing AI-generated versus real images.

2.2 "Introduction to Statistical Analysis" YouTube Video

The video "Introduction to Statistical Analysis" (YouTube) provides a foundational overview of key statistical principles, highlighting the role of both descriptive and inferential statistics in analyzing and interpreting data. It emphasizes the use of measures such as mean, median, standard deviation, and variance to summarize data distributions, while also introducing hypothesis testing, confidence intervals, and p-values to support decision-making based on sample data. Importantly, the video warns against common misinterpretations such as confusing correlation with causation or misusing averages in skewed distributions, which are crucial to avoid when interpreting visual attention metrics like fixation durations or gaze heatmaps. These concepts provide a solid grounding for analyzing eye-tracking data in the current study, where distinguishing between AI-generated and real images requires clear, statistically sound comparisons. Moreover, the emphasis on variability and data visualization aligns with best practices for representing gaze behavior across different user groups, such as by gender or uncertainty level in decision-making.

2.3 Analysis of Human Perception in Distinguishing Real and AI-Generated Faces: An Eye-Tracking Based Study (Huang et al.)

Huang et al. conducted an eye-tracking study to investigate how humans differentiate between real and AI-generated faces, using images created by StyleGAN3 and real photos from the FFHQ dataset. Involving over 7,000 images, their study revealed that participants achieved a classification accuracy of 76.8%, with significantly different gaze behaviors depending on whether the faces were real or synthetic. Specifically, suspected fake images triggered longer fixation durations, wider gaze dispersion, and higher gaze entropy, indicating a more cautious and exploratory viewing strategy. These behaviors reflect underlying uncertainty in decision-making, as participants took more time and visual effort when unsure. The study also used advanced gaze metrics such as the gaze convex hull and gaze entropy to quantify attention patterns. While the study did not analyze gender-based differences, the techniques it used provide a valuable methodological framework for exploring that dimension. This research directly supports the current study's focus on identifying visual attention differences between AI and real images, examining how uncertainty manifests in gaze patterns, and offers a strong foundation for extending analysis to gender-based gaze behavior.

2.4 Eye-tracking analysis of face observing and face recognition (Andrej Iskra, Helena Gabrijelčič Tomc)

Iskra & Tomc present an eye-tracking study focused on face observation and recognition, analyzing fixation patterns across key facial features, which are the eyes, nose, and mouth, during memory tasks. Conducted with participants memorizing facial images, the study found that these central features received the majority of fixations, reflecting their importance in face encoding and recognition. These results highlight foundational gaze behaviors essential for interpreting more complex interactions with facial stimuli. In the context of our research, their findings validate the use of *Areas of Interest* (AOIs) around facial landmarks and justify analyzing fixation metrics like duration and count. Moreover, the study underscores that such gaze patterns serve as reliable indicators of face processing strategies. This groundwork supports our first research question by confirming baseline gaze tendencies when viewing real faces and provides a benchmark against which AI-generated face gaze patterns can be compared.

3. Methodology

3.1 Participants

A total of 87 Year 2 undergraduate students enrolled in the Bachelor of Computer Science program, majoring in Artificial Intelligence, participated in the study. Gender information was available for 84 participants (53 male, 31 female) due to missing IDs in participant list data. All participants had normal or corrected-to-normal vision and provided informed consent to take part in the study.

3.2 Apparatus

All eye movements were recorded using the Tobii Pro Fusion eye tracker with sampling data rate of 50 Hz. Stimuli were displayed on a 24-inch monitor, with high video quality of 1920 px × 1080 px. A 9-point calibration procedure was conducted before each recording session. The experiment was managed through Tobii Pro Lab software (version 25.7.1400).

3.3 Stimuli and Conditions

Each participant completed a single assessment condition consisting of six questions. For each question, participants were required to choose one of two response options: A or B.

3.4 Experimental Design and Procedure

Six sets of pictures, each consisting of a real image and an AI-generated image, were shown to the participants. Each participant was instructed to choose the AI-generated picture as accurately as possible without time constraints. The eye-tracking process began upon presentation of each item and continued until responses were submitted.

3.5 Measures and Data Collected

The following data were recorded:

- Eye-tracking metrics: Total Fixation Duration, Fixation Count, Time to First Fixation, Total Visit Duration, Time to First Click, Scanpath
- Behavioral: Response Accuracy

4. Data Analysis

4.1 Preprocessing

We make an analysis of the stimuli provided and predict the possible Area of Interests (AOIs) of each of the pictures. Then, the AOIs are drawn using the AOI tools inside the Tobii Pro Lab. AOIs for each question are drawn manually for each recording. All the AOIs for 87 recordings are drawn. The metrics data for each question includes Total Fixation Duration, Average Fixation Duration, Fixation Count, Time to First Fixation, First Fixation Duration, Total Visit Duration, Average Visit Duration, and Visit Count.

After data is extracted from Tobii Pro Lab, the preprocessing involves selecting key gaze metrics by removing redundant sheets and data cleaning by merging the AOI attributes with repeating columns. The cleaned data was then merged with participant metadata and response accuracy for further analysis.

4.2 Features extracted

The features extracted from Tobii Pro Lab are as follows:

Features Extracted	Explanation
Tot Fixation duration	Sum of all fixation times within the area of interest (AOI)
Average Fixation duration	Mean time spent per fixation
Fixation count	Total number of fixations on the AOI
Time to First Fixation	Time taken to first fixate on the AOI
First Fixation duration	Duration of the initial fixation on the AOI
Total Visit duration	Sum of all visit times to the AOI
Average Visit duration	Mean time per visit to the AOI
Visit Count	Number of separate times the AOI was entered
Total Recording Duration	Total time of the recorded session
Total Time of Interest Duration	Total time during which any AOI was active or visible
Total Time of Interest Fixation Count	Total number of fixations across all AOIs during the interest time

4.3 Statistical analysis methods

- Paired-sample t-tests
  - Used to compare participants' visual attention between AI-generated and real images on each eye-tracking metric.
  - Statistical significance was set at $p < 0.05$, and separate t-tests were conducted for each metric to compare attention on AI vs. real AOIs.

- Independent Samples t-test
  - Used to compare the average value of eye-tracking metrics between 'Correct' and 'Incorrect' answer groups. A p-value < 0.05 was considered statistically significant.
  - Used to compare the average values of eye-tracking metrics between two independent groups ('Correct' vs. 'Incorrect' or 'Male' vs. 'Female'), when the data followed a normal distribution.

- Random Forest Classifier
  - A machine learning model was trained to predict answer correctness using eye-tracking data. Its performance was primarily evaluated using:
    - AUC (Area Under the Curve): To measure the model's overall ability to distinguish between correct and incorrect answers.
    - Feature Importance: To identify which specific eye-tracking metrics were the most powerful predictors.

- Shapiro–Wilk Test
  - Used to assess the normality of the distribution of gaze metrics within male and female participant groups for each Area of Interest (AOI).
  - This test determined whether parametric or non-parametric methods would be applied.
  - A p-value < 0.05 indicated non-normality, then the Mann–Whitney U test was applied, whereas if the data was normally distributed ($p \geq 0.05$), the Independent Samples t-Test was used.

- SMOTE (Synthetic Minority Over-sampling Technique)
  - Used to address class imbalance between Male and Female groups by synthetically oversampling the minority group before analysis.

- Mann–Whitney U Test
    - Used as a non-parametric alternative to the t-test when one or both gender groups did not meet the normality assumption.
    - It compared the distributions of gaze metrics between the two groups without assuming normality.
- Pearson Correlation Analysis
    - Used to assess linear relationships between selected eye-tracking metrics within each gender group.
    - Correlation matrices were then visualized using heatmaps to explore inter-metric dependencies between the gaze metrics

4.4 Software used

Tobii Pro Lab was used as the primary software to extract gaze data and metrics from the eye-tracking experiments. We used Tobii Pro Lab's built-in tools to define the Area of Interest (AOIs) for each image. Eye-tracking metrics like Total Fixation Duration, Fixation Count, Time to First Fixation, Total Visit Duration, and others were extracted per AOI.

Google Colab is used for preprocessing, statistical analysis, and visualization of the eye-tracking data of the three research questions. We use essential libraries like pandas for data manipulation, numpy for numerical computation, matplotlib and seaborn for data visualization, and scipy for statistical testing. We also used scikit-learn for building Random Forest classifiers, imbalanced-learn for balancing gender-related data using SMOTE, plotly for interactive visualizations, and tabulate for generating clear summary tables.

5. Results

5.1 Research Question 1

This study investigates visual attention differences between AI-generated and real images using eye-tracking data. A total of six image-question sets were analyzed, each containing one AI-generated image and one real image. Participants have to choose the AI-generated image in each pair.

To quantitatively assess visual attention, we used 5 key eye-tracking metrics for the analysis:

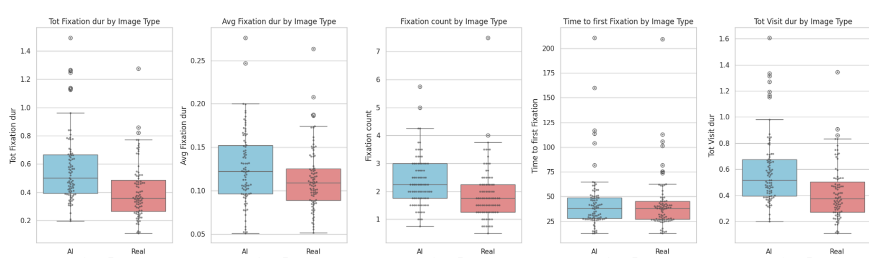
1. Total Fixation Duration: Total time the viewer's gaze remained fixed on the image.
2. Average Fixation Duration: Mean duration of individual fixations.
3. Fixation Count: Number of fixations made on the image.
4. Time to First Fixation: Time before the participant first looked at the image.
5. Total Visit Duration: Sum of all viewing sessions on the image (including revisits).

Each metric was compared between AI and real images using **paired-sample t-tests**, with a significance threshold of $p < 0.05$. A total of **87 participants** were included in the study.

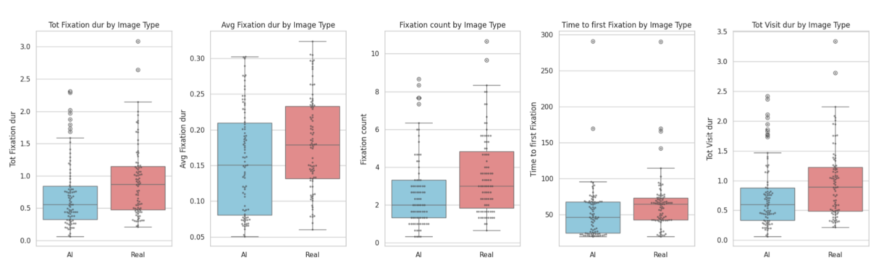
Visualization of Results:

Please refer to Appendix (5) for the full data visualisation of research question **1**

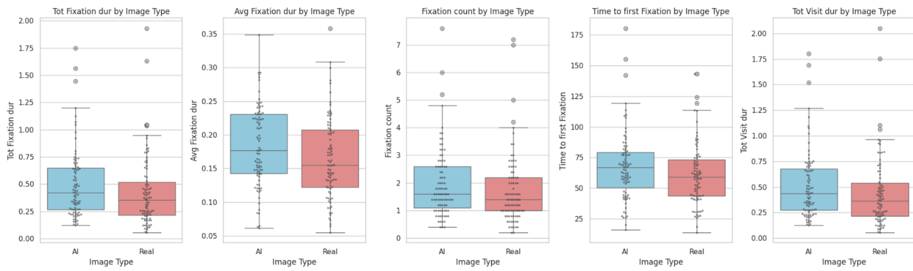
Question 1:



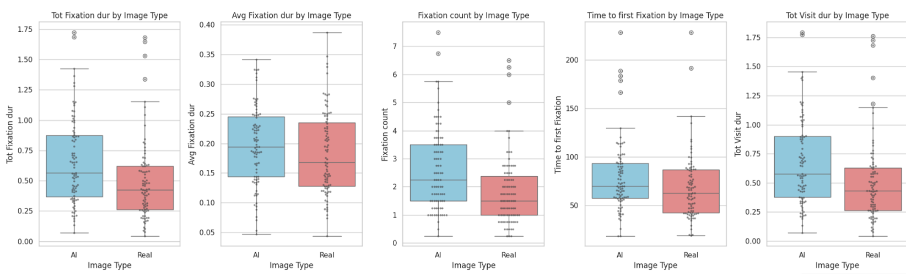
Question 2:



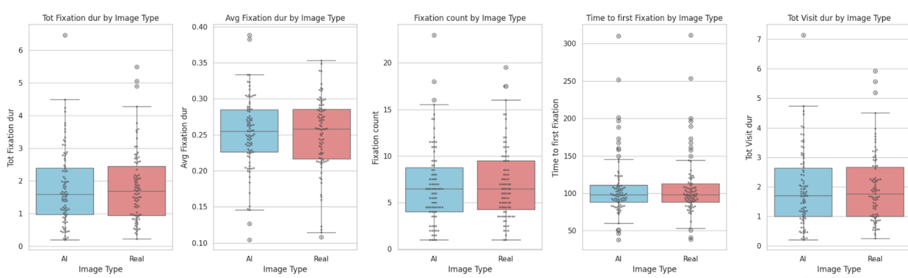
Question 3:



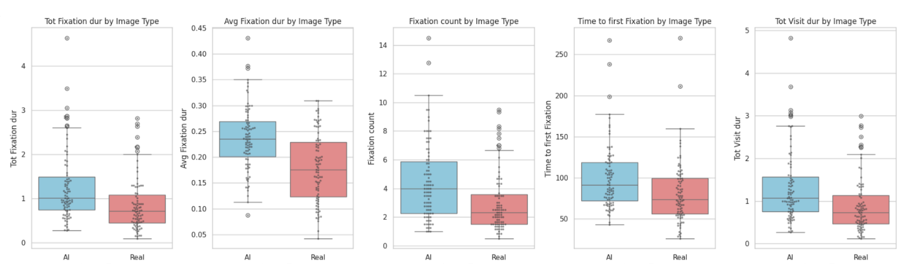
Question 4:



Question 5:



Question 6:



Overall Findings:

Metrics	Q1	Q2	Q3	Q4	Q5	Q6
Total Fixation Duration	Yes	Yes (Real > AI)	Yes	Yes	No	Yes
Average Fixation Duration	Yes	Yes (Real > AI)	Yes	No	No	Yes
Fixation Count	Yes	Yes (Real > AI)	Yes	Yes	No	Yes
Time to First Fixation	No	AI faster	Real faster	Real faster	No	Real faster
Total Visit Duration	Yes	Yes (Real > AI)	Yes	Yes	No	Yes

Note: Direction of difference indicated where AI did not perform better.

Hypothesis:
There is no significant difference in gaze patterns when participants view AI-generated images compared to real images.

Based on the results shown in the table, **the hypothesis is rejected**. There is a significant difference in gaze patterns when participants view AI-generated images compared to real images.

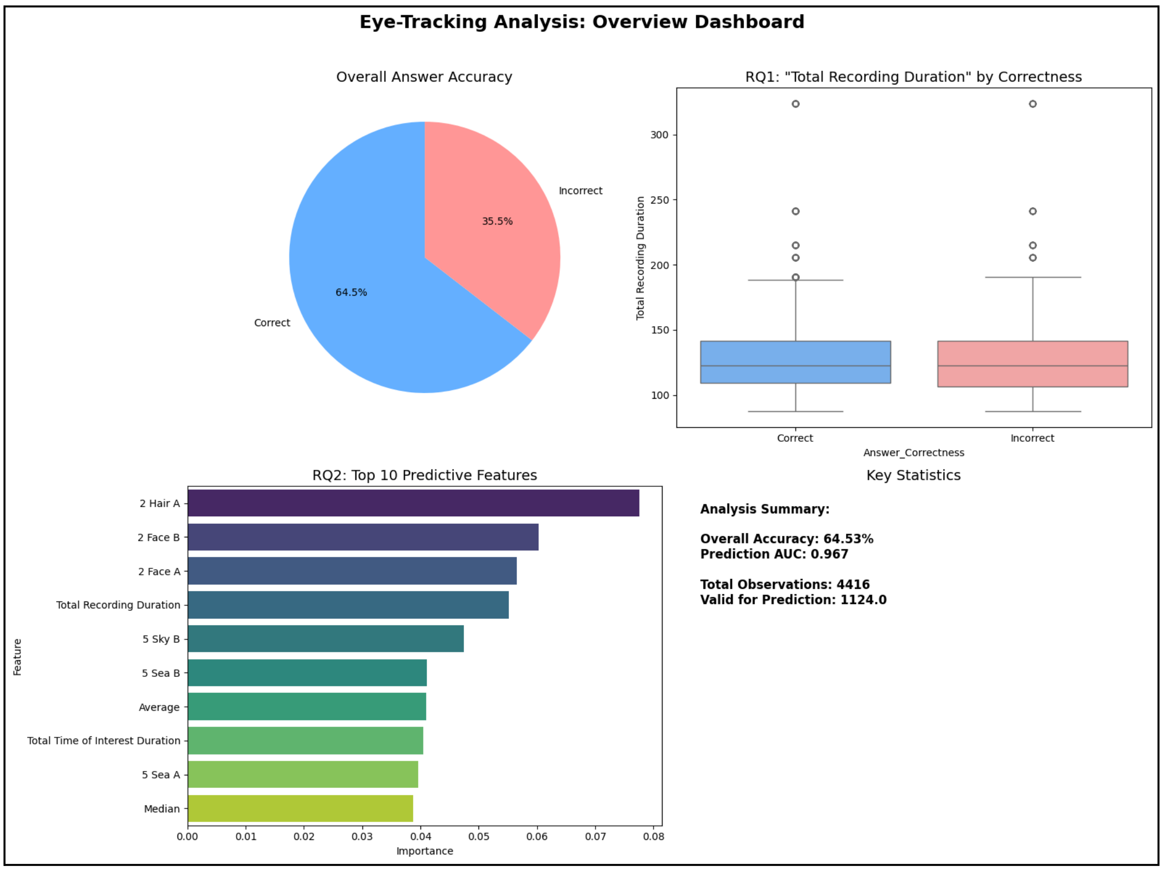
Justification:
Across the six questions (Q1–Q6), multiple eye-tracking metrics consistently show significant differences between real and AI-generated images:

- **Total Fixation Duration**: Significant in 5 out of 6 questions (Q1, Q2, Q3, Q4, Q6), with Q2 indicating participants fixated longer on real images.
- **Average Fixation Duration**: Significant in 4 out of 6 questions (Q1, Q2, Q3, Q6), with longer durations on real images in Q2.
- **Fixation Count**: Significant in 5 out of 6 questions (Q1, Q2, Q3, Q4, Q6), again with higher counts for real images in Q2.
- **Time to First Fixation**: Shows a consistent pattern (Q2, Q3, Q4, Q6) where real images were fixated on faster, suggesting a difference in initial attention, except for Q2, with the AI-generated image being fixated on faster.
- **Total Visit Duration**: Significant in 5 out of 6 questions (Q1, Q2, Q3, Q4, Q6), with Q2 showing longer visits for real images.

5.2 Research Question 2

This study investigates the relationship between uncertainty in decision-making and visual attention. Specifically, it examines whether incorrect or low-confidence answers are associated with different gaze behavior.

Visualization of Results:



The most critical insight from this dashboard is that while simply looking at images for a longer or shorter time does not determine success, where a participant looks is an extremely powerful predictor of whether they will answer correctly. The model's high accuracy, combined with the specific nature of the top predictive features, strongly supports this conclusion.

Overall Findings:

1. Does viewing time differ by answer correctness?

--- RQ1: Does viewing time differ by answer correctness? ---

Metric	Correct Mean	Incorrect Mean	P-Value	Effect Size
Total Recording Duration	132.8661	131.3340	0.2689	-0.0403
Total Time of Interest Duration	132.8661	131.3340	0.3060	-0.0403
Total Time of Interest Fixation Count	336.5397	332.3434	0.6067	-0.0499

Answer : No. There is no statistically significant difference in viewing time.
Justification : None of the tested time-based metrics showed a p-value below 0.05 significance threshold. The most promising metric had a p-value of 0.2689.

2. Can we predict correctness from gaze patterns?

--- RQ2: Can we predict correctness from gaze patterns? ---

Prediction Model: RandomForest Classifier
AUC Score: 0.9670

Top 10 Most Important Features for Prediction:

Feature	Importance
2 Hair A	0.077634
2 Face B	0.060244
2 Face A	0.056617
Total Recording Duration	0.055239
5 Sky B	0.047445
5 Sea B	0.041060
Average	0.040969
Total Time of Interest Duration	0.040470
5 Sea A	0.039682
Median	0.038809

Answer : YES, with high confidence.

Justification : The model can predict correctness with 90.4% accuracy (AUC = 0.967), which is significantly better than random chance. This indicates strong, learnable patterns in the eye-tracking data. The most predictive feature was '2 Hair A'.

5.3 Research Question 3

This study investigates potential gender-based differences in gaze behavior when participants are tasked with distinguishing between AI-generated and real images by choosing the AI-generated image in each pair.

A total of 87 participants took part in the gaze experiment. However, after data cleaning and the removal of incomplete entries, the final dataset included 84 unique participants (53 male, 31 female).

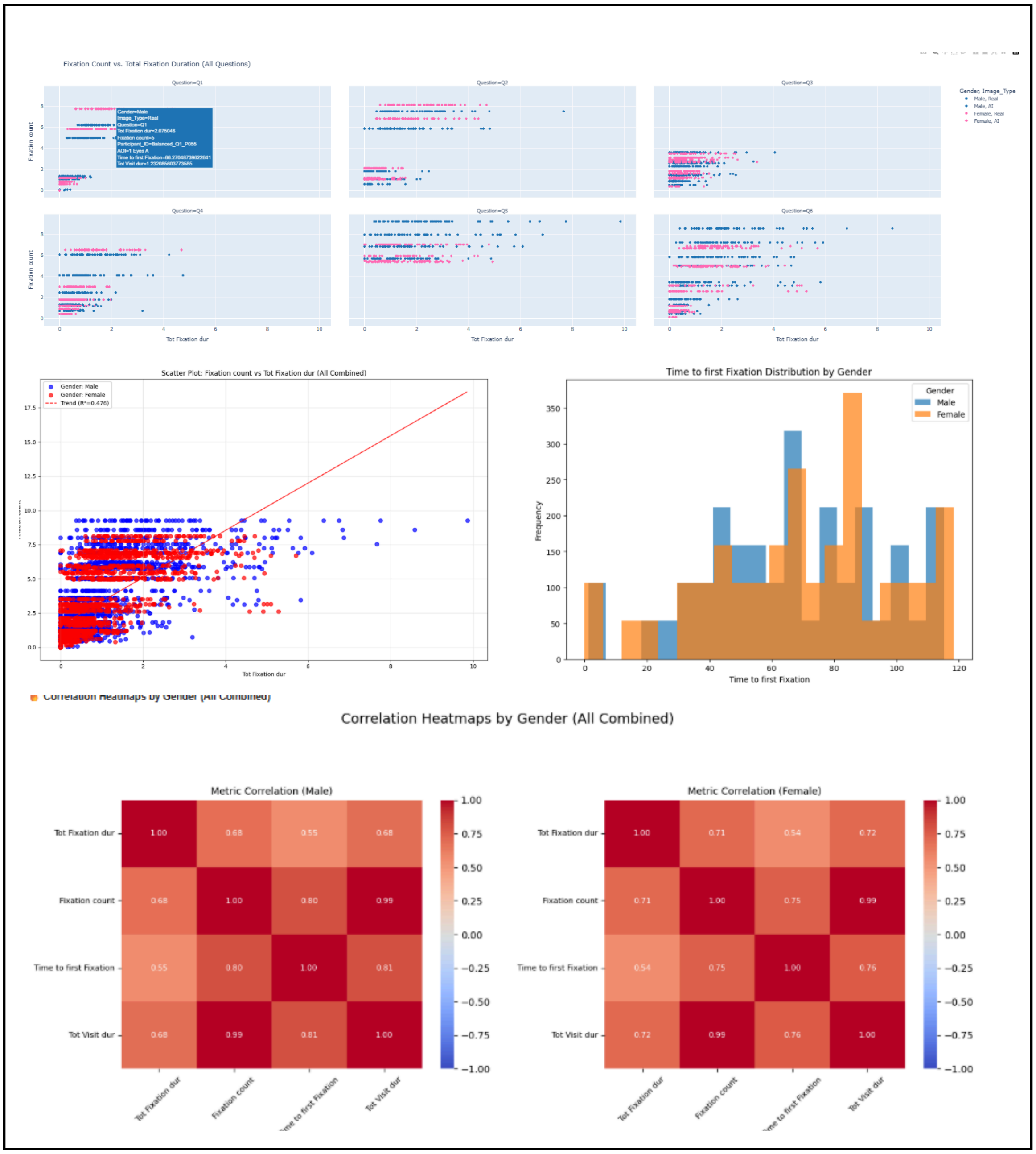
4 key eye-tracking metrics for the analysis:

1. Total Fixation Duration : Total time the viewer's gaze remained fixed on the Area of Interest (AOI)
2. Fixation Count : Total number of fixations made on an AOI
3. Time to First Fixation : Time taken for participants to first look at the AOI.
4. Total Visit Duration : Sum of durations for all visits to an AOI (including revisits)

Normality within gender groups was assessed using the Shapiro-Wilk test. Based on the results, either **independent sample t-tests** or **Mann-Whitney U tests** were applied to compare gaze behavior between genders, with a significance threshold of $p < 0.05$. To address gender imbalance, the dataset was balanced using SMOTE, resulting in 53 male and 53 female(additional synthetic female data points) observations.

Visualization of Results:

[Please refer to Appendix (6) for the full data visualisation of research question 3]



The visual analysis above suggests that while males and females exhibit broadly similar patterns in how they explore AI-generated and real images across the four key eye-tracking metrics, there are instances of gender-specific differences, particularly when examining specific Areas of Interest (AOIs) within individual questions. This indicates that men and women look at AI and real images in mostly the same way overall. These small but consistent differences in specific image areas suggest that gender does affect how people view AI versus real images, even though the overall patterns are similar.

Overall Findings:

	Question	Tot Fixation dur (p)	Fixation count (p)	Time to first Fixation (p)	Tot Visit dur (p)
0	Q1	0.350654	0.444440	0.364560	0.344486
1	Q2	0.104121	0.114204	0.029280	0.098485
2	Q3	0.083303	0.198366	0.022087	0.082087
3	Q4	0.668836	0.910187	0.011678	0.682072
4	Q5	0.025866	0.045993	0.096047	0.023306
5	Q6	0.007694	0.031818	0.133872	0.008593

	Question	Tot Fixation dur (Sig)	Fixation count (Sig)	Time to first Fixation (Sig)	Tot Visit dur (Sig)	Overall Significant
0	Q1	False	False	False	False	False
1	Q2	False	False	True	False	True
2	Q3	False	False	True	False	True
3	Q4	False	False	True	False	True
4	Q5	True	True	False	True	True
5	Q6	True	True	False	True	True

Hypothesis:

There are no significant differences in at least one measure of gaze behavior between male and female participants when distinguishing between AI-generated and real images, potentially varying by image content or specific Areas of Interest (AOIs).

Based on the tables shown above, **the hypothesis is rejected.** There are significant differences in at least one measure of gaze behavior between male and female participants when distinguishing between AI-generated and real images, potentially varying by image content or specific Areas of Interest (AOIs).

Justification:

- Based on the tables shown above, Q1 shows no significant differences in any gaze metric, as all p-values are greater than 0.05.
- However, Q2 to Q6 demonstrate significant gender differences as they have metrics with p-values greater than 0.05
- Q2 to Q4 demonstrate significant gender differences primarily in the metric "Time to First Fixation," which suggests early-stage attention may vary by gender depending on the image content.
- Q5 and Q6 reveal broader differences across multiple gaze metrics, indicating that gender contributes to variations in how intensely and frequently participants engage with visual content.
- Hence, the hypothesis is rejected.

6. Discussion

6.1 Research Question 1

1. Interpretation of Results

- **Longer/more fixations on real images, such as in Q2**: Suggests real images may demand more cognitive processing or attract sustained attention, possibly due to richer details or familiarity.
- **Faster first fixations on real images for Q3, Q4, Q6**: Implies real images might capture initial attention more quickly, potentially due to perceptual priors such as natural textures, lighting.
- **Inconsistent differences in Q5**: The lack of significance in Q5 across metrics could indicate that certain AI-real pairs were perceptually similar, reducing discriminability.

2. Relation to Prior Work

- **Attention to Real vs. Synthetic Images**: Previous research indicates that viewers often respond differently to real and AI-generated (synthetic) images. Our study supports this trend—participants generally showed longer and more frequent fixations on real images, which is in Question 2.
- **Perceptual Sensitivity to AI-Generated Imperfections:** Other literature suggests that even when viewers can't consciously detect flaws, the brain may still register them. This may explain our finding that real images captured attention more quickly

3. Limitations

- **Stimulus specificity**: Results may depend on the selected image pairs.
- **Task constraints**: Forced-choice paradigms (identifying AI images) may bias attention differently than free viewing.
- **Sample homogeneity**: Participants' familiarity with AI imagery could modulate effects, but was unexamined.
- **Temporal dynamics**: Metrics like time to first fixation may conflate bottom-up and top-down attention.

4. Future Directions

- **Expand stimulus set**s: Include diverse generative models and image types to identify consistent artifact-driven patterns.
- **Dynamic tasks**: Test free-viewing or memory-based paradigms to decouple task demands from perceptual differences.
- **Individual differences**: Assess how expertise, like photographers or AI developers, modulates gaze behavior.
- **Longitudinal studies**: Track how gaze patterns evolve as AI image quality improves or public exposure increases.

6.2 Research Question 2

1. Interpretation of Results

- The key finding is that viewing time does not predict success, but gaze location does.
- Participants who answered correctly did not spend more time on the task. Instead, they focused their attention on specific regions of the image, such as 'Hair' and 'Face', which were highly predictive of the correct answer.
- This suggests that successfully identifying AI images is a strategic search task, where success depends on efficiently finding known AI flaws, not on overall effort.

2. Relation to Prior Work

- The importance of the 'Hair' feature connects our behavioral findings directly to the known technical limitations of current AI image generators, such as models like StyleGAN and DALL-E, to produce unnatural textures or inconsistencies in hair rendering (Karras et al., 2019; Ramesh et al., 2021).

3. Limitations

- **Stimulus-Specific**: The findings, particularly the importance of '2 Hair A', are tied to the specific images used in this test.
- **Generalizability**: The participant sample may not represent the general population.
- **Analysis Scope**: The analysis did not include the sequence of eye movements (scanpaths), which could offer deeper strategic insights.

4. Future Directions

- **Diverse Imagery**: Test if these findings hold true across a wider variety of AI-generated images and subjects.
- **Training Study**: Design an experiment to see if we can train people to become better AI detectors by instructing them to focus on common flaw areas.
- **Scanpath Analysis**: Analyze the sequence of fixations to better understand the visual search strategies of successful participants.

6.3 Research Question 3

1. Interpretation of Results

- The analysis reveals that gender has an observable impact on gaze behavior, with differences in how male and female participants attend to visual information
- While Q1 showed no significant differences across any gaze metrics that indicated similar visual behavior between male and female participants, Q2 to Q6 each revealed at least one metric with significant variation by gender.
- In Q2 to Q4, differences were mainly observed in Time to First Fixation ,whereas in Q5 and Q6, differences can be observed across gaze metrics

2. Relation to Prior Work

- Prior eye-tracking studies have highlighted gender-based differences in visual attention patterns, influenced by task type, content familiarity, and cognitive strategies. Our findings align with the research, showing that males and females may adopt different visual strategies, particularly in early-stage attention (Mills et al., 2011).

3. Limitations

- **Small Sample Size:** Even though gender groups were balanced using SMOTE, the overall number of participants was still relatively small, which may reduce the reliability and generalizability of the findings.
- **Context Dependence**: The differences observed may reflect not only gender but also task interpretation or image content, making it difficult to fully isolate the impact of gender.
- **Synthetic Balancing**: Techniques like SMOTE were used to address group imbalance, but synthetic data may not fully reflect the variability found in real human responses.

4. Future Directions

- **Larger and More Diverse Samples**: Future studies should include a greater number of participants from varied backgrounds and ensure balanced representation across genders to improve both generalizability and the reliability of gender-based comparisons.
- **Broader Image Sets**: Using a wider range of image content, styles, and quality levels could help determine whether the observed gender effects are consistent or image-dependent.
- **Tracking Gaze Over Time**: Instead of just looking at where participants look at a certain position for how long, future studies can analyze the order and path of eye movements to help analyze how males and females explore images differently

7. Limitations and Suggestions for Future Work

There are several limitations being identified throughout the project. Firstly, the stimulus set was limited to only a fixed number of image pairs, which limits the generalizability of the findings. The force-choice nature of tasks like participants having to choose between AI-generated and real images, might affect participants' natural viewing behaviour. In order to handle these issues, future work should focus on expanding the diversity of the stimulus set by not just restricting to picking only one image that is AI-generated, and include options where both images are AI-generated images created from different models.

The next limitation is the small sample size and imbalance of gender data. With only 87 participants, all drawn from a similar academic background, the gaze experiment's findings might be restricted to this specific group and may not accurately reflect the broader population. The use of synthetic data using the SMOTE algorithm to address the gender imbalance may not fully capture the complex variability found in real gaze behavior across genders. To address this, future research should prioritize recruiting a significantly larger and more diverse participant pool. This means actively seeking individuals from various demographic backgrounds, including different age groups, genders, professions, educational levels, and cultural backgrounds, not just students from an AI background.

8. Conclusion

This study examined gaze behavior differences between AI-generated and real images, how decision uncertainty influences visual attention, and gender-based variations in gaze patterns during image classification tasks. Results show that participants fixated longer and more frequently on real images, indicating deeper cognitive processing compared to AI-generated ones. Uncertain decisions were associated with more exploratory gaze behavior, reflecting higher cognitive load rather than increased viewing time. Gender differences also emerged, with specific variations in fixation metrics and Areas of Interest (AOI). This suggests that males and females engage with visual content in subtly different ways. These findings contribute to the field by demonstrating that eye-tracking can uncover nuanced patterns of visual engagement shaped by image authenticity, cognitive confidence, and gender, offering valuable insights for improving AI image detection, adaptive user interfaces, and educational or training tools involving synthetic visual content.

9. Acknowledgement

We would like to express our sincere gratitude to all those who contributed to the successful completion of this project.

First and foremost, we would like to thank my supervisor, Dr. Unaizah Hanum Binti Obaidellah, for her continuous support, invaluable guidance, and insightful feedback throughout the research process. Her expertise and encouragement have been crucial in shaping this work.

We also extend our appreciation to the participants who volunteered their time for the eye-tracking sessions. Their involvement made this study possible.

Special thanks to the faculty and staff of the Faculty of Computer Science and Information Technology for providing the necessary resources, facilities, and technical support.

10. References

[1] Krejtz, K. (2021). Statistical analysis of eye tracking data [Workshop slides]. LEAD ME Summer Training School. Universitat Autònoma de Barcelona. https://ddd.uab.cat/pub/presentacions/2021/248206/K.krejtz_workshop.pdf

[2] LEAD-ME COST Action. (2021, July 26). Statistical analysis of eye tracking data [Video]. YouTube. https://www.youtube.com/watch?v=CkoJvnWEOxw

[3] Huang, J., Gopalakrishnan, S., Mittal, T., Zuena, J., & Pytlarz, J. (2024, September 23). Analysis of human perception in distinguishing real and AI-generated faces: an Eye-Tracking based study. arXiv.org. https://arxiv.org/abs/2409.15498

[4] Iskra, Andrej & Gabrijelčič, Helena. (2016). Eye-tracking analysis of face observing and face recognition. Journal of Graphic Engineering and Design. 7. 5-11. doi.org/10.24867/JGED-2016-1-005.

[5] Looser, C. E., & Wheatley, T. (2010). The tipping point of animacy. Psychological Science, 21(12), 1854–1862. https://doi.org/10.1177/0956797610388044

[6] A Style-Based generator architecture for generative adversarial networks. (n.d.). IEEE Conference Publication | IEEE Xplore. https://doi.org/10.1109/CVPR.2019.00453

[7] Feng, Q., Zheng, Y., Zhang, X., Song, Y., Luo, Y., Li, Y., & Talhelm, T. (2011). Gender differences in visual reflexive attention shifting: Evidence from an ERP study. Brain Research, 1401, 59–65. https://doi.org/10.1016/j.brainres.2011.05.041

11. Appendix

Include any supplementary material (e.g., AOI maps, stimuli).

1. AOI Mapping & Definitions for Image Analysis Questions:
   AOI Mapping & Definition

2. Data Analysis Google Colab Project Link (Research Question 1):
   a. Q1: ComparingVisualAttentionQ1.ipynb
   b. Q2: ComparingVisualAttentionQ2.ipynb
   c. Q3: ComparingVisualAttentionQ3.ipynb
   d. Q4: ComparingVisualAttentionQ4.ipynb
   e. Q5: ComparingVisualAttentionQ5.ipynb
   f. Q6: ComparingVisualAttentionQ6.ipynb

3. Data Analysis Google Colab Project Link (Research Question 2):
   Research_Q2.ipynb

4. Data Analysis Google Colab Project Link (Research Question 3):
   GenAIEyeTrackingDatasetRQ3Analysis.py

5. Data Visualisation (Research Question 1)
   Full Visualisations for Research Question 1

6. Data Visualisation (Research Question 3)
   Full Visualisations for Research Question 3

7. Project Presentation Slides
   Presentation Slides Group 4.pdf

8. Video Presentation
   https://youtu.be/8ZCYeGZSuXc