

*Note: the format of this report with paragraphs in bullet points is due to professor preferences. It is a report of research results rather than a formally written research paper which I can provide upon request.*

## **AI Marketing Awareness - A/B Testing Report**

- a. For our user study, we conducted A/B testing to evaluate if being aware of human versus AI-made messaging affects people's responses to a given advertisement. Our advertisement is for a cleaning sponge called SparkleScrub generated by ChatGPT. The advertisement was the same in both versions of our A/B test, except one version had a disclaimer that it was AI-generated. The presence or absence of this statement is the independent variable in our experiment. The variables we measured are purchase likelihood, perceived price justification, recommendation likelihood, advertisement persuasiveness, and emotional appeal. These variables relate to advertisement effectiveness and were collected through post-exposure survey questions. Each of our hypotheses looks at whether the disclaimer statement causes any difference in the responses for each variable or the correlation between each pair of dependent variables. The research design is between-subjects and we conducted independent samples t-tests and computed the correlation matrix to test our null hypotheses.
- b. We conducted five independent samples t-tests for our analysis. The independent samples t-tests were conducted once for each dependent variable. Across all five tests, there were no statistical differences between the version of the advertisement with the disclaimer and the version without (t-values < 1.48 and p-values were > 0.17, df= 58).

In addition to the independent samples t-tests, we computed a correlation matrix between all pairs of the five dependent variables. The correlation matrix reveals a moderately positive statistically significant relationship between purchase likelihood and price justification ( $r(N=60) = .448, p < .001$ ). It reveals a strong positive statistically significant relationship between purchase likelihood and recommendation likelihood ( $r(N=60) = .718, p < .001$ ). Between purchase likelihood and emotional appeal, the correlation matrix reveals a moderately strong statistically significant positive relationship ( $r(N=60) = .522, p < .001$ ). Between purchase likelihood and advertisement persuasiveness, the correlation matrix reveals a strong positive statistically significant relationship ( $r(N=60) = .687, p < .001$ ). Between price justification and recommendation likelihood, the correlation matrix reveals a moderately strong statistically significant relationship ( $r(N=60) = .521, p < .001$ ). Between price justification and emotional appeal, the correlation matrix reveals a moderately positive statistically significant relationship ( $r(N=60) = .343, p = .007$ ). Between price justification and advertisement persuasiveness, the correlation matrix reveals a moderate statistically significant relationship ( $r(N=60) = .494, p < .001$ ). Between recommendation likelihood and emotional appeal, the correlation matrix reveals a strong positive statistically significant relationship ( $r(N=60) = .731, p < .001$ ). Between recommendation likelihood and advertisement persuasiveness, the correlation matrix reveals a moderately strong positive statistically significant relationship ( $r(N=60) = .517, p < .001$ ). Finally, between emotional appeal and advertisement persuasiveness, the correlation matrix reveals a moderately positive statistically significant relationship ( $r(N=60) = .427, p < .001$ ).

- c. We fail to reject the null hypotheses for all of the independent samples t-tests as all of our p-values are greater than .17 (lowest was .174). The null hypotheses are: “The presence of the AI disclaimer statement has no effect on purchase likelihood”, “The presence of the AI disclaimer statement has no effect on the degree of perceived price justification”, “The presence of the AI disclaimer statement has no effect on the likelihood of recommending the product to others”, “The presence of the AI disclaimer statement has no effect on how much a person feels persuaded by the advertisement”, and “The presence of the AI disclaimer statement has no effect on the degree of emotional appeal of the advertisement”. By failing to reject these null hypotheses, we might be making a type II error. Our effect size (d) was .38, which was calculated from our highest t-score (1.376) and corresponds to 30 participants in each group. As our corresponding power was .42 our chances of detecting a true effect were only 42%. Based on the power formula  $1 - \beta$ , this means that our beta value is .58 and that there is a 58% chance we are making a type II error.

For all of our hypotheses for correlation between our dependent variables, we can reject the null hypotheses as all of our corresponding p-values are lower than or equal to 0.007. These are the null hypotheses we can reject: “There is no correlation between purchase likelihood and price justification”, “There is no correlation between price justification and advertisement persuasiveness”, “There is no correlation between purchase likelihood and advertisement persuasiveness”, “There is no correlation between purchase likelihood and emotional appeal”, “There is no correlation between price justification and recommendation likelihood”, “There is no correlation between price justification and advertisement persuasiveness”, “There is no correlation between price justification and emotional appeal”, “There is no correlation between recommendation likelihood and advertisement persuasiveness”, “There is no correlation between recommendation likelihood and emotional appeal”, and “There is no correlation between advertisement persuasiveness and emotional appeal.”

By looking at the corresponding correlation coefficients, we can determine that all relationships were positive and that the majority of the relationships between our dependent variables are moderate ( $0.5 > r > 0.3$ ) or strong ( $r > 0.5$ ). Across all of our correlation hypotheses, we risk making a type I error - i.e. incorrectly rejecting the null hypothesis. The maximum probability we make a type I error is equal to the alpha error probability which is .05 in our case; however, we can look at the p-values to determine the likelihood that we observe the null hypotheses. For the hypotheses between all pairs of dependent variables except price justification and emotional appeal, the p-value is  $p < .001$  which means that the probability of observing a result as extreme or more extreme than this, assuming the null hypothesis is true, is 0.1%. For price justification and emotional appeal,  $p = .007$  which means that the probability of observing a result as extreme or more extreme than this, assuming the null hypothesis is true, is 0.7%. This means that the probability we make a type I error for price justification and emotional appeal is 0.7% and for all other pairs of dependent variables it is 0.1%.

Note: As our results do not detect any statistically significant results for the effect of our manipulation of the independent variable, it is important to note that we are rejecting the null

hypotheses that there is no correlation between each pair of dependent variables. We observe these relationships regardless of the independent variable rather than due to its effect.

- d. We conducted the power analysis to answer c above. As mentioned above, for the independent samples t-tests, our effect size (Cohen's d) was .38 and our power was .42. For correlation, we conducted power analysis based on the relationship between purchase likelihood and price justification ( $r(N=60) = .448, p < .001$ ). The power is .956 and the critical r is -.25 (lower) and .25 (upper). The effect size for each correlation is equal to the r for each correlation.
- e. Our results indicate that while we observe no statistically significant difference between the version of our advertisement with the disclaimer and the version without it, the dependent variables - purchase likelihood, perceived price justification, recommendation likelihood, advertisement persuasiveness, and emotional appeal - all have positive statistically significant moderate to strong relationships. We selected these five variables as we wanted to look at different aspects of what made an advertisement effective - i.e. factors that contribute to a customers positive associations of a product and in turn may generate revenue. Our results suggest that the variables we selected are positively correlated meaning we might expect them all to be important aspects of what makes an advertisement perform well. As an example, the strongest relationship we observed was between emotional appeal and recommendation likelihood ( $r(N=60) = .731, p < .001$ ). This means that if the advertisement successfully emotionally appeals to the consumer, the consumer is more likely to recommend the advertised product to someone else.

While we did not observe statistically significant differences between the version of our advertisement that included the disclaimer statement and the version that did not on any of the outcome variables, there are several limitations that may help us understand the meaning of these results. First and foremost, with 30 participants in each group, our power was .42. To produce results with less risk of making a type II error, we would have needed a larger participant size to get greater power and increase our likelihood of finding true effects. Similarly, it is important to note that while participants were randomly assigned to the disclaimer or no disclaimer conditions, we recruited participants from classmates, family, and friends. Especially for family or close friends, there may be a bias towards positive ratings in responses regardless of the condition assignment which could act as a confound in our research. If our study had random sampling in addition to random assignment, the responses may have been more neutral. We did not collect any biographic or demographic data from participants; however, given the nature of our recruitment strategy, our participants are unlikely to accurately represent the population. For example, our target population was young adults and older. To reduce bias, our participants should accurately reflect all age groups in this population; however, it is not unlikely that younger adults were more represented than older adults. If younger adults are more positive towards AI, this could act as a potential confound when drawing conclusions about a larger population if the lack of statistical differences is driven by attitude differences between age groups. Additionally, as one of our group members is from Norway and recruited Norwegian family members, this may introduce potential cultural differences in perceptions of AI that would not be accurately tested for in our study.

It is also important to note that at the time of this study, the use of AI has become increasingly widespread within the past year. Due to widespread exposure, people may have accepted the use of AI in the media. In light of this, the results may have been different if this study had been conducted in 2023 rather than in 2024. It may also be important to consider the context in which people may disapprove or approve of the use of AI in the media. Negative responses to AI-generated content may not be caused by an awareness or negative perception of AI, but rather something for instance being poorly written, using unnatural language or unnatural facial expressions. A more complex and nuanced research design may be required to answer these questions. For example, if users were exposed to multiple advertisements either categorized as high quality human written, low quality human written, high quality AI written, and low quality AI written and had to detect whether each one was AI then answer our advertisement survey questions, we might have had a greater chance of detecting statistically significant results.

## f. APPENDIX

### Part 2

1. For our user study, we will conduct A/B testing to evaluate if being aware of human versus AI-made messaging will affect people's responses to a given advertisement. Our advertisement is for a cleaning sponge called SparkleScrub generated by ChatGPT. The advertisement will be the same in both versions of our A/B test, except one version will have a disclaimer that it was AI-generated. The presence or absence of this statement is the independent variable in our experiment. The variables we will measure are purchase likelihood, perceived price justification, recommendation likelihood, advertisement persuasiveness, and emotional appeal. These variables relate to advertisement effectiveness and will be collected through post-exposure survey questions. Each of our hypotheses looks at whether the disclaimer statement causes any difference in the responses for each variable or the correlation between each pair of dependent variables. The research design will be between-subjects and we will conduct independent samples t-tests and compute the correlation matrix to test our null hypotheses.

2. The dataset will look something like this:

group	purchase_likelihood	price_justification	recommendation	persuasion	emotion
"disclaimer"	4	2	4	3	1
"no_disclaimer"	7	4	5	7	6

- We are doing a between-subjects A/B test design with two levels to the independent variable. The IV is categorical with two levels, defined as the presence or absence of the

statement: “Disclaimer: This advertisement was made using AI.” The disclaimer group receives this statement and the control group does not.

- Our dependent variables are continuous on a Likert-type scale from 1-7. They are:
  - Purchase likelihood: how likely the person is to purchase the sponge on a scale of 1-7.
  - Price justification: to what degree a person feels as though the price of the product is justified on a scale from 1-7.
  - Recommendation likelihood: how likely a person is to recommend the product to someone else on a scale from 1-7.
  - Advertisement persuasiveness: how successful the advertisement was in persuading the individual of the product's benefits on a scale from 1-7.
  - Emotional appeal: the degree to which the advertisement emotionally appeals to a person on a scale from 1-7.

3. When we thought about what questions to ask in our survey, we intentionally thought about how to make our data collection as clean as possible. Since all of our questions can each be applied to Likert scales, our data cleaning will be very minimal and, if everything goes as planned, nonexistent, as every data point for every DV is a number mapped to a set scale of 1-7. Additionally, we should be able to require an answer to every question for a user to proceed which should ensure that we do not have any missing values. Consequently, we should not need to handle data imputation. However, should requiring answers to every question for some reason not be possible within Qualtrics, we will handle missing values through data imputation as needed. To do so, we may use a simple imputation method such as mean or median imputation; however, we believe looking at the data and seeing which method describes the central tendency best across both groups will be important to finalize this strategy.

4. For our analysis, we are conducting both independent samples t-tests and computing a correlation matrix. A two-tailed independent samples t-test will allow us to measure the impact of the manipulated independent variable on each dependent variable since we are doing a between-subjects A/B test design yielding a t-score and subsequent p-value. Based on these scores, we can determine whether or not we can reject each subsequent null hypothesis. Further, because we have multiple dependent variables that all relate to the concept of effective advertisement, we are interested in seeing the correlation between each dependent variable with the correlation matrix and subsequent correlation coefficients. Below are the hypotheses for every test statistic.

#### Independent samples t-test:

- We will obtain a t-value from the independent samples t-test to assess the effect of the independent variable on the dependent variable of purchase likelihood.
  - $H_a$ : The presence of an AI disclaimer statement decreases the likelihood of purchase.
  - $H_0$ : The presence of the AI disclaimer statement has no effect on purchase likelihood.

- We will obtain a t-value from the independent samples t-test to assess the effect of the independent variable on the dependent variable of price justification.
  - $H_a$ : The presence of the AI disclaimer statement negatively impacts the degree of perceived price justification.
  - $H_0$ : The presence of the AI disclaimer statement has no effect on the degree of perceived price justification.
- We will obtain a t-value from the independent samples t-test to assess the effect of the independent variable on the dependent variable of recommendation likelihood.
  - $H_a$ : The presence of the AI disclaimer statement negatively impacts the likelihood of recommending the product to others.
  - $H_0$ : The presence of the AI disclaimer statement has no effect on the likelihood of recommending the product to others.
- We will obtain a t-value from the independent samples t-test to assess the effect of the independent variable on the dependent variable of advertisement persuasiveness.
  - $H_a$ : The presence of an AI disclaimer statement negatively impacts how much a person feels persuaded by the advertisement.
  - $H_0$ : The presence of the AI disclaimer statement has no effect on how much a person feels persuaded by the advertisement.
- We will obtain a t-value from the independent samples t-test to assess the effect of the independent variable on the dependent variable of emotional appeal.
  - $H_a$ : The presence of an AI statement negatively impacts the degree of emotional appeal of the advertisement.
  - $H_0$ : The presence of the AI disclaimer statement has no effect on the degree of emotional appeal of the advertisement.

Correlation matrix:

Across all variables we hypothesize that the effect of the IV on each DV will be correlated with the effect on all other DVs. Our overall null hypothesis is that there is no correlation between the effect of the IV on each DV.

- We will obtain a correlation coefficient to assess the relationship between purchase likelihood and price justification.
  - $H_a$ : The relationship between purchase likelihood and price justification is positively correlated.
  - $H_0$ : There is no correlation between purchase likelihood and price justification.
- We will obtain a correlation coefficient to assess the relationship between purchase likelihood and recommendation likelihood:
  - $H_a$ : The relationship between purchase likelihood and recommendation likelihood is positively correlated.
  - $H_0$ : There is no correlation between purchase likelihood and recommendation likelihood.
- We will obtain a correlation coefficient to assess the relationship between purchase likelihood and advertisement persuasiveness:

- $H_a$ : The relationship between purchase likelihood and advertisement persuasiveness is positively correlated.
- $H_0$ : There is no correlation between purchase likelihood and advertisement persuasiveness.
- We will obtain a correlation coefficient to assess the relationship between purchase likelihood and emotional appeal:
  - $H_a$ : The relationship between purchase likelihood and emotional appeal is positively correlated.
  - $H_0$ : There is no correlation between purchase likelihood and emotional appeal.
- We will obtain a correlation coefficient to assess the relationship between price justification and recommendation likelihood:
  - $H_a$ : The relationship between price justification and recommendation likelihood is positively correlated.
  - $H_0$ : There is no correlation between price justification and recommendation likelihood.
- We will obtain a correlation coefficient to assess the relationship between price justification and advertisement persuasiveness:
  - $H_a$ : The relationship between price justification and advertisement persuasiveness is positively correlated.
  - $H_0$ : There is no correlation between price justification and advertisement persuasiveness.
- We will obtain a correlation coefficient to assess the relationship between price justification and emotional appeal:
  - $H_a$ : The relationship between price justification and emotional appeal likelihood is positively correlated.
  - $H_0$ : There is no correlation between price justification and emotional appeal.
- We will obtain a correlation coefficient to assess the relationship between recommendation likelihood and advertisement persuasiveness:
  - $H_a$ : The relationship between recommendation likelihood and advertisement persuasiveness is positively correlated.
  - $H_0$ : There is no correlation between recommendation likelihood and advertisement persuasiveness.
- We will obtain a correlation coefficient to assess the relationship between recommendation likelihood and emotional appeal:
  - $H_a$ : The relationship between recommendation likelihood and emotional appeal is positively correlated.
  - $H_0$ : There is no correlation between recommendation likelihood and emotional appeal.
- We will obtain a correlation coefficient to assess the relationship between advertisement persuasiveness and emotional appeal:
  - $H_a$ : The relationship between advertisement persuasiveness and emotional appeal is positively correlated.
  - $H_0$ : There is no correlation between advertisement persuasiveness and emotional appeal.

## Part 1

1. A short summary:
2. User Study Plan:
  - a. For our user study, we will conduct A/B testing to evaluate whether the awareness of human versus AI-made messaging will affect people's response to a given advertisement. In both versions of our A/B test, the advertisement will be the exact same with the exception of a disclaimer that it was AI made in one version. In this way, the presence or absence of this statement is a predictor or independent variable in our experiment and the variables we collect or measure are purchase likelihood, perceived price justification, recommendation likelihood, advertisement persuasiveness, and emotional appeal. These variables all relate to advertisement effectiveness and will be collected through post exposure survey questions.
    - i. This is our advertisement, which we generated with the help of chatGPT:



i.

*"Say goodbye to stubborn grime with SparkleScrub, the ultimate cleaning sponge designed to make every surface shine. Its durable yet gentle material tackles the toughest messes without scratching, keeping your home spotless with ease. Perfect for kitchens, bathrooms, and beyond, SparkleScrub makes cleaning quicker and more efficient than ever. Plus, it's eco-friendly and reusable, saving you money and reducing waste. Get your sparkle back with SparkleScrub – because a clean home is a happy home!"*

- b. Our study is experimental with a between-subjects A/B test design. We will randomly assign our participants into two separate groups where one receives the advertisement as is and the other with a disclaimer statement that it was made by AI. The questions following exposure to the advertisement will also be the exact same to measure advertisement effectiveness in both scenarios and compare for differences.
- c. Our IV is the presence or absence of an AI disclaimer statement. Our DVs are purchase likelihood, perceived price justification, recommendation likelihood, emotional appeal, and advertisement persuasiveness.
- d. For our independent variable, we will operationalize or manipulate it by adding a statement that the advertisement was AI-made in only one version of the advertisement while keeping everything else, including our survey questions and setup, identical.



Therefore our operationalized independent variable is the presence or absence of the statement; “Disclaimer: This advertisement was made using AI.”

Our dependent variables are purchase likelihood, perceived price justification, recommendation likelihood, emotional appeal, and advertisement persuasiveness. All of our dependent variables are measured on likert scales ranging from 1-7.

- We operationalize purchase likelihood as how likely the person is to purchase the sponge on a scale of 1-7.
  - Related question: On a scale of 1 to 7, how likely are you to purchase this product?
    - (1 = Not likely at all, 7 = Extremely likely)
- Perceived price justification is operationalized as to what degree a person feels as though the price of the product is justified on a scale from 1-7.
  - Related question: The price of this product is \$3.99. Does the value of the product justify the price of the product?
    - Not at all (1), Very Slightly (2), Slightly (3), Not sure (4), Somewhat (5), Very Much (6), Extremely (7)
- Recommendation likelihood is operationalized as how likely a person is to recommend the product to someone else on a scale from 1-7.
  - Related question: How likely are you to recommend this product to someone else?
    - Scale 1-7, 1 = extremely unlikely, 7 = extremely likely
- Advertisement persuasiveness is operationalized as how successful the advertisement was in persuading the individual of the product's benefits on a scale from 1-7.
  - Related Question: How persuasive was the ad in convincing you of the product's benefits?
    - (1 = Not persuasive at all, 7 = Extremely persuasive)
- Emotional appeal is operationalized as the degree to which the advertisement emotionally appeals to a person on a scale from 1-7.
  - Related Question: On a scale of 1 to 7, how much did the advertisement for this product emotionally appeal to you?
    - (1 = Not at all, 7 = Very much)
- e. Our population target is young adults and older. We made our advertisement for a cleaning sponge to make the product widely applicable to adult lifestyles and preferences. The wide applicability and availability of the population is important in reaching our participant goal since we plan on recruiting participants through students in this class as well as family and friends. We plan on recruiting 60 people for our study, with 30 participants in each version.

### 3. Analysis plan:

- a. Our research questions include: **(1)** “Does the presence of an AI disclaimer statement on an advertisement impact the likelihood of purchase?” **We hypothesize that the presence of an AI disclaimer statement decreases the likelihood of purchase.** **(2)** “Does the presence of an AI disclaimer statement on an advertisement impact the degree of which the product's price feels

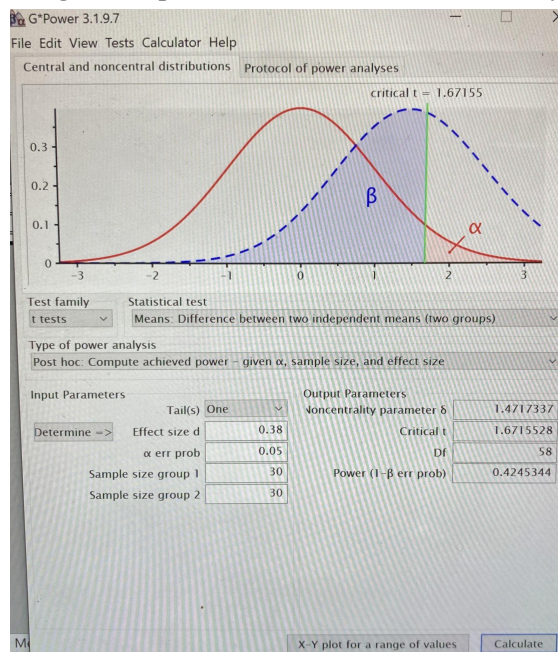
justified?” **We hypothesize that the presence of an AI disclaimer statement negatively impacts the degree of perceived price justification. (3)** “Does the presence of an AI disclaimer statement on an advertisement impact the likelihood of recommending the product to others?” **We hypothesize that the presence of an AI disclaimer statement negatively impacts the likelihood of recommending the product to others. (4)** “Does the presence of an AI disclaimer statement impact how much a person feels persuaded by the advertisement?” **We hypothesize that the presence of an AI disclaimer statement negatively impacts how much a person feels persuaded by the advertisement. (5)** “Does the presence and absence of an AI disclaimer statement impact the emotional appeal of the advertisement?” **We hypothesize that the presence of an AI statement negatively impacts the emotional appeal of the advertisement.**

Additionally, we are looking at the correlation between each pair of the dependent variables, where our questions are, “**is  $DV_i$  correlated with  $DV_j$ ?**” and our null hypotheses are that each pair of dependent variables are not correlated. For each hypothesis for the relationship between the independent variable and each dependent variable, our null hypothesis is that the presence or absence of the AI disclaimer statement has no effect on the given dependent variable. Since all our questions will yield data on different aspects of how a person is impacted by the advertisement mentioned in the operationalization above, analyzing our data between the different groups should enable us to determine whether or not we can reject our null hypothesis/hypotheses. All our data will be self-reported through our survey questions on a scale from 1-7 as mentioned in the operationalization section above. For example, for purchase likelihood, we will get data on a scale from 1-7 on how likely a person feels they are to purchase the product, where 7 is extremely likely. While we rely on self-reported data, making our survey questions specific helps ensure the validity of our data and results. Since we will randomly assign participants and our design only differs on the disclaimer statement where if we can meet our participant target, our experiment should not have any confounds. Consequently, if we can reject one or more of our null hypotheses based on our associated p-value, it will mean there is a statistically significant relationship between the presence or absence of the AI disclaimer statement and that dependent variable. If there is, based on the results of our statistical tests, we can infer whether the presence or absence of the AI disclaimer statement causes any difference in these variables. Additionally, if we can reject one or more of the null hypotheses for the correlation between pairs of the dependent variables, it will mean that these variables have a significant relationship.

- b. To analyze our data we will conduct independent samples t-tests for each dependent variable and compute a correlation matrix. This is because we have a categorical independent variable and continuous dependent variables. Since each of our dependent variables will be on the exact same likert scale (1-7), computing the correlation matrix will allow us both to see whether there is any correlation between our dependent variables. Additionally, we will conduct the independent samples t-test once for each dependent variable. A two tailed independent samples t-test will allow us to measure the impact of the manipulated independent variable on each dependent variable since we are doing a between-subjects A/B test design.

## APPENDIX 2

**Image 1:** G\*power screenshot from t-test analysis



**Image 2:** G\* power screenshot from correlation

## DSCI 517: Project Part 3

Group: Tomine Bergseth, Sydney Leister, Tammi Sison, Elisa Xia

