

THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

Investigating Potential Effects of Confirmation Bias on Participants' Discernment of Real versus
AI-Generated Political Images

CLAIRE CHEN
SPRING 2026

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Psychology
with honors in Psychology

Reviewed and approved* by the following:

Rick O. Gilmore
Professor of Psychology
Thesis Supervisor & Honors Adviser

Karen Gasper
Professor of Psychology
Faculty Reader

* Electronic approvals are on file.

ABSTRACT

Artificial intelligence (AI) has unprecedented technical abilities to create vast amounts of sophisticated media quickly and has been used in the United States to create fake images and stories. When viewing images online, Americans must now remain vigilant and discern fake images from real ones. Previous literature has discovered that participants are less discerning and more trusting when fake news headlines align with their pre-existing political beliefs, an effect of confirmation bias, but evidence remains inconclusive for studies utilizing AI-generated videos. This study examines the potential effects of confirmation bias on participants' abilities to discern between AI-generated and real images and hypothesized that participants would more likely believe an image was real if it aligned with their political views, and political affiliation would significantly affect participants' accuracy with images showing support for different ideologies. 126 participants were shown various (real and AI-generated) images that displayed left-leaning and right-leaning politicians in positive and negative contexts and asked to conclude if the image was real or AI-generated. Though participants were more likely to believe an image was real if it aligned politically, overall image classification accuracy remained unaffected. Older age was found to be a significant predictor of better classification accuracy, leading to new discussions about future AI education and training.

TABLE OF CONTENTS

LIST OF TABLES.....	iii
ACKNOWLEDGEMENTS.....	iv
Introduction.....	1
Misinformation in Digital and Social Media.....	2
Visual Information Processing and Misinformation.....	3
Cognitive Evaluation and Confirmation Bias.....	4
Hypotheses.....	6
Methods.....	8
Participants.....	8
Materials.....	8
Procedure.....	9
Results.....	11
Discussion.....	18
Lack of Political Affiliation Effect.....	18
Effects of Image Congruency on Evaluation Accuracy.....	19
Potential Biases Uninfluential on Objective Classification Performance.....	20
Decreased Accuracy of Real Image Evaluations.....	21
Increased Accuracy with Age.....	22
Limitations and Future Directions.....	22
Conclusion.....	24
Appendix A General Result Data Analysis.....	26
Appendix B Image Stimuli Presented to Participants.....	28
Appendix C Political Evaluation and Demographic Questions.....	57
BIBLIOGRAPHY.....	61

LIST OF TABLES

Table 1 - Overview of Participant-Image Congruency	6
Table 2 - Participant Demographics.....	11
Table 3 - Summary of Image Accuracy	12
Table 4 - Average Accuracy Percentages by Image Condition and Political Affiliation	13
Table 5 - Mixed-Effects Logistics Regression of Perceived Image Realness	14
Table 6 - Mixed Effects Logistic Regression Predicting Accuracy.....	15
Table 7 - Demographic ANOVA Results	16
Table 8 - Image Evaluation Accuracy by Age Group.....	16

ACKNOWLEDGEMENTS

Thank you so much to Dr. Karen Gasper and Dr. Rick Gilmore for all the time, support, and feedback they provided for this project. Their research experience and insights proved invaluable to this project and are extremely appreciated. Dr. Gasper's excitement and expert feedback have sparked desires to further explore this project; her dedication and creative mindset are valued.

An incredible thank you to Sarah Richards for her support throughout the thesis and Schreyer journey as a whole. Her support in thesis operations and academics does not go unnoticed. This project would not be possible without her advice.

Thank you to Erica Chen for her continued encouragement. Her positivity, self-assurance, analytical mind, curiosity, and intellect will produce impactful research and academic contributions.

Finally, the biggest thank you to my parents, Gary Chen and Catherine Hu. Their unwavering support and advice were crucial throughout this process; I am consistently reminded how lucky I am to have their encouragement and expertise. For always fostering an environment that promoted academic appreciation and a curious mindset, thank you.

Introduction

Artificial intelligence (AI) has emerged as the dominant new technology in the United States (US), and its introduction and continued improvement have had a drastic impact on the lives of Americans. Generative AI models can automate and rapidly spread intentional and tailored misinformation campaigns (Goldstein et al., 2023). AI-generated images proliferate social media, especially political conversations, stirring fake news stories and providing fabricated “proof” for nonexistent events (Yang et al., 2023). Coupled with the immense reach of social media algorithms, AI-generated images and media can contribute to large-scale misinformation (Hicks, 2026). Confirmation bias, the tendency for people to seek information that aligns with their beliefs, expectations, and experiences, also contributes to the spread of misinformation, especially in online communities (Klayman, 1995; Knobloch-Westerwick et al., 2014; Del Vicario et al., 2017). Literature has outlined the impact of confirmation bias in believing fake news, with participants more likely to trust headlines that align with their views (Calvillo et al., 2021, Del Vicario et al., 2017). Despite extensive literature on misinformation, misinformation in digital media remains understudied, with no research conducted on the impact of confirmation bias on fake images (Yang et al., 2023). With image creation and alteration becoming drastically advanced in recent years, examining the impact of confirmation bias in detecting AI-generated images will provide new, tech-focused perspectives to the study of misinformation.

Misinformation in Digital and Social Media

With AI advancing exponentially, image generation abilities have evolved to produce ever more realistic images based on user prompts. Individuals living in the US, especially social media users, are likely to encounter AI-generated images due to the sheer volume of production. 34 million AI-generated images are created daily, and 71% of images shared on social media are AI-generated (Marr, 2025). AI-generated images can emulate real images, causing viewers to believe false events, stories, and realities.

Digital media has remained the most popular source of news for several years, with 43% of adults accessing news through social media and the web (with the next most popular source being television at 32%). Social media trails news websites and apps by 7% as the second most popular source of political news, with Americans ages 18-29 being most likely (48%) to use social media as their primary news source (Mitchell et al., 2020).

AI-generated images are not the first “fake” images to permeate social media. Over 20% of Facebook image posts about United States politics contain misinformation, which includes images containing misleading text, falsely labeled images, and edited photos depicting something false (Yang et al., 2023). The overwhelming volume of fake images online has signaled alarms for many Americans, with 70% of adults finding “it hard to trust information online” as they cannot differentiate “content [that] is real or AI generated” (AI Misinformation Control 2025 | Statista, 2025). Individuals often struggle to analyze digital media for authenticity due to the complex processing of visual information.

Visual Information Processing and Misinformation

Innate processing of digital information, like images, contributes to belief in the content viewed. People process images much faster and more efficiently than written or verbal messages, producing stronger emotional responses that grab the viewer's attention and aid in memory recall (Barry, 1997; Graber, 1990; Newhagen, 1998). Images also compel audiences more effectively due to their similarity to reality, vividness, and immediacy of memory (Wittenberg et al., 2021; Yadav et al., 2011). In the context of news coverage, images also produce stronger framing effects than text or audio messages (Powell, 2015). Images of candidates can heavily influence voters' perceptions and affect their votes, exemplifying the impact visuals can have on opinion and behavioral intentions. (Ahler et al., 2016; Rosenberg et al., 1986). Exploring participants' ability to identify fake images provides important insight for preventing the spread of misinformation.

While little research surrounding the identification of AI-generated images exists, previous studies have concluded that participants generally cannot discern edited images from originals. Even when participants believe an image has been edited, they seldom identify what has changed (Nightengale, 2017). Participants' ability to detect doctored images correlates positively with greater disruption of the underlying pixel structure of the photos, indicating that a higher volume of edits suggests image manipulation (Nightengale, 2017). However, AI-generated media produces entirely new images, without any pixel disruption present. Individuals cannot rely on traditional strategies of identifying image falsehoods and must rely on logic and beliefs.

Cognitive Evaluation and Confirmation Bias

The process of determining the veracity of an image involves cognitive evaluation processes, and the dual process model from Stanovich (1999) and Kahneman (2011) has been widely examined and accepted throughout literature, especially when evaluating belief in misinformation (Moravec et al., 2019). The dual process model outlines two separate cognitive processes, System 1 and System 2, that comprise all cognitive evaluation. System 1 represents automatic cognition that happens instinctively and runs continuously to deliver automatic, involuntary conclusions. System 2 requires conscious choice and effort, but can be easily overwhelmed (Kahneman, 2011). Most people attempt to minimize cognitive effort and accept the conclusions of System 1 over attempting to invoke System 2, typically evaluating events in alignment with their existing beliefs (Taylor & Fiske, 1978). The reliance on System 1 of the dual process model explains the existence and prevalence of confirmation bias in cognition.

Confirmation bias denotes the systematic inclination to preferentially seek out, interpret, and remember information in ways that reinforce an individual's preexisting beliefs and expectations (Klayman, 1995; Knobloch-Westerwick et al., 2014; Del Vicario et al., 2017). Individuals tend to avoid information that causes discomfort; when people read headlines inconsistent with their beliefs, Moravec et al. (2018) found evidence of suppression in the alpha band (8-12 Hz), which previous literature associated with an increase in cognitive load. Increases in cognitive load explain why bias infiltrates the selection of information, compelling individuals to look at familiar, comfortable information in decision-making (Rajsic et al., 2016). Confirmation bias acts naturally and influences everyday information processing and judgment; increased polarization of opinion and spreading of information can be attributed to the cognitive phenomenon (Del Vicario et al., 2017).

Researchers have observed the effects of confirmation bias in political misinformation, finding that participants show less discernment for fake news that originates from or supports their political affiliation (Beauvais, 2022; Ludwig & Sommer, 2024). After the 2020 election, people with more conservative ideologies and approval of Donald Trump showed greater belief in fake headlines claiming election fraud (Calvillo et al., 2021). Ludwig & Sommer (2024) found that Democrats and Republicans are more likely to accept news congruent with their beliefs and are less adept at detecting favorable fake news. Michael & Breaux (2021) similarly concluded that political affiliation influenced participants' belief in the integrity of news sources, with participants more likely to believe a source had "fake news" if it was incongruent with their beliefs.

While extensive literature on misinformation and confirmation bias with textual stimuli exists, the effects in digital media remain understudied. In a literature review conducted by Yang et al. (2023) for articles on "disinformation, propaganda, fact-checking and misinformation" published between 2017 and 2021, only nine out of 135 articles focused on images. Though Ebner et al. (2025) sought to expand research on political confirmation bias in visual media, their primary visual focus was on deepfake videos and found no significant conclusions in support of confirmation bias effects.

This present study seeks to extend the study of confirmation bias to images, an under-examined branch of misinformation literature. Due to recent advancement of AI image generation powers, identifying AI-generated images has not been studied in the context of confirmation bias. This study examines whether political confirmation bias influences participants' ability to distinguish between real and AI-generated images.

Hypotheses

Prior research has shown that confirmation bias has had an impact on participants' ability to distinguish between real and fake political news headlines (Ludwig & Sommer, 2024; Michael & Breaux, 2021). This study will extend the study of confirmation bias beyond text and look to determine if political affiliation influences accuracy when evaluating politically congruent versus politically incongruent images. Congruency will occur when participants are evaluating images that align with their political views.

Table 1 - Overview of Participant-Image Congruency

Participant	Image Type	Congruency
Democrat	Pro-Liberal	Congruent
Democrat	Pro-Conservative	Incongruent
Republican	Pro-Conservative	Congruent
Republican	Pro-Liberal	Incongruent

This paper will explore two hypotheses evaluating response bias and performance bias. Both hypotheses together will determine the effects of confirmation bias on participant image evaluation. Additionally, this paper will explore a hypothesis predicting that other participant variables (besides political affiliation) will not have a significant impact on image evaluation.

Hypothesis 1: Main Effect of Congruency

I hypothesize that participants will be more likely to classify politically congruent images as real. I hypothesize that image congruency will significantly affect participants' probability of

classifying an image as real, producing a main effect of congruency. Confirmation bias suggests that participants are more likely to believe things they agree with, leading participants to believe politically congruent images are real.

Hypothesis 2: Interaction Between Political Affiliation and Image Congruency

I hypothesize that different political groups will respond differently to the same types of images, producing a crossover interaction. Participants will be more accurate when evaluating images that align with their beliefs and less accurate when evaluating images that are not aligned with their beliefs. Confirmation bias suggests that if the two political participant groups evaluate the same image with preconceptions influencing the decision, the images will be judged differently.

Hypothesis 3: Impact of Other Variables

I hypothesize that all other variables (age, gender, AI usage, and social media usage) will not have a significant impact on image evaluation by participants. Though variations are expected with expected exposure to AI (ex. Participants with AI experience will evaluate more accurately, participants who are younger will perform more accurately), I do not predict the variations to be statistically significant.

Methods

Participants

126 study participants were recruited on Prolific and were paid \$2 for completion of the survey. Of the 126 participants, 41% were male ($n=52$), 53% were female ($n=67$), 4% were non-binary ($n=5$), and 2% ($n=2$) identified as another option (represented as “other”). All study participants were American by nationality and resided in the United States. 21% of participants were ages 18 to 29 ($n=26$), 30% of participants were ages 30 to 39 ($n=38$), 22% of participants were ages 40 to 49 ($n=28$), 14% of participants were ages 50 to 59 ($n=18$), 7% of participants were ages 60 to 69 ($n=9$), and 6% of participants were 70 years of age or older ($n=7$).

Materials

This study used 30 images total, both real and AI-generated. All images have been sourced legally from the web, news sources, or social media (for AI-generated images). 20 of the images featured a political figure from the United States, someone who ran for presidential or vice-presidential office from 2012 to 2026. The political images featured right-leaning politicians (J. D. Vance and Donald Trump) and left-leaning politicians (Kamala Harris, Joe Biden, and Barack Obama) in both a positive and negative context. The context of the image influenced its classification as pro-liberal (more left-leaning) or pro-conservative (more right-leaning). Right-leaning politicians in negative contexts (ex. pictured with criminals) were classified as pro-liberal, and left-leaning politicians in negative contexts (ex. getting arrested) were classified as pro-conservative. Including positive and negative images, 10 total images representing pro-

liberal ideology (congruent with left-leaning participants' beliefs) and 10 total images featuring pro-conservative ideology (congruent with right-leaning participants' beliefs) are included.

Participants viewed images and answered corresponding questions through a Qualtrics survey posted on Prolific in March 2026.

Procedure

I conducted a within-subjects experiment, where all participants were given a set of 30 real and AI-generated images. For each image, participants were asked to determine if they believed the image was real or AI-generated, and their confidence in the decision. To assess the impact of confirmation bias in detecting AI-generated images, I established one independent variable of image congruence based on three sub-independent variables: image truth (AI-generated and real images), image type (pro-left and pro-right), and participant political affiliation (very liberal, liberal, moderate, conservative, and very conservative). This study measured AI image detection performance and assessed the potential impact of confirmation bias on detection performance, measured by congruence effects and the interaction between participant political affiliation and image political valence on detection accuracy.

Participants were given 30 images to evaluate, with 10 pro-conservative images (five real, five AI-generated), 10 pro-liberal images (five real, five fake), and 10 control images of portraits and setting images of both AI-generated people (5) and real people (5). For each of the images, each participant was asked whether they believed the image is real or AI-generated, their confidence in the decision (evaluated on a 5-point Likert scale ranging from “not confident at all” to “extremely confident”), and if they had seen the image before (yes or no). All images

were presented in a randomized order for each participant to prevent pattern detection that would impact image authenticity decision-making.

After viewing each image and answering the related questions, each participant was asked to answer six questions to evaluate their political affiliation. Questions include self-reports of political affiliation, party affiliation, and favorability of specific conservative and liberal political figures, evaluated on a 5-point Likert scale ranging from “strongly disagree” to “strongly agree”. The determination of political affiliation through questions on topics was emulated from Ebner et al. (2025). Questions determining alignment with right-leaning ideals were scored from -20 to +20 based on responses from the Likert scale, with a difference of 10 between each choice. Questions determining alignment with left-leaning ideals were scored from +20 to -20 based on the responses from the Likert scale, in order to ensure that left leaning ideology was captured by a negative number and right-leaning ideology was captured by a positive number. All responses to political questions were summed to determine the participant’s overall political affiliation.

Finally, participants were asked to fill out a survey of demographic information, including age, gender, cultural background, disposition to trust, familiarity with AI, and social media usage.

Results

128 participants responded to the survey, with two responses removed from the dataset due to the participants' failure to pass attention checks asking them to select "disagree" for question 89 and "extremely confident" for question 120. A participant size of $N=126$ was analyzed for this study, with very liberal ($n=44$, 35%), liberal ($n=25$, 20%), moderate ($n=28$, 22%), conservative ($n=15$, 12%), and very conservative ($n=14$, 11%) participants represented.

Table 2 - Participant Demographics

Characteristic	Category	N	Percentage
Race	White	98	77.8%
	Black/African American	15	11.9%
	Asian	7	5.6%
	Other	6	4.8%
Gender	Female	67	53.2%
	Male	52	41.3%
	Non-binary	5	4.0%
	Other	2	1.6%
Age Group	18-29	26	20.6%
	30-39	38	30.2%
	40-49	28	22.2%
	50-59	18	14.3%

	60-69	9	7.1%
	70+	7	5.6%
Political Standing	Very Liberal	44	34.9%
	Liberal	25	19.8%
	Moderate	28	22.2%
	Conservative	15	11.9%
	Very Conservative	14	11.1%

As shown in Table 3, on average, participants evaluated 20.42 of 30 images correctly, with the most accurate participant correctly evaluating 29/30 images and the least accurate participant evaluating just under half of the images correctly (14/30).

Table 3 - Summary of Image Accuracy

N	Mean	Std Dev	Min	Max	Variance
126	20.42063492	2.915759042	14	29	8.501650794

Table 4 displays the average accuracy percentage of image evaluation for each of the political affiliations, with an accuracy breakdown for each image truth, image type, and overall accuracy. Both left-leaning (very liberal and liberal) and right-leaning (very conservative and conservative) participants had similar overall accuracy scores for political and control images combined, as well as just political images. However, a chi-squared test highlighted a significant difference in accuracy ($p=0.011$) by political affiliation for the real left-leaning (pro-liberal) images (Appendix A). All political affiliations were tested against each other using various two-

proportion Z-tests, and very liberal participants performed significantly better than conservative participants ($p < 0.01$) and very conservative participants ($p < 0.05$), liberal participants performed significantly better than conservative participants ($p < 0.05$), and moderate participants performed significantly better than conservative participants ($p < 0.05$). Overall, the chi-squared and Z-tests showed that conservative participants were significantly less accurate when evaluating real pro-liberal images compared to other participant groups.

Table 4 - Average Accuracy Percentages by Image Condition and Political Affiliation

Political Affiliation	Pro-Liberal Real	Pro-Conservative Real	Pro-Liberal Fake	Pro-Conservative Fake	Political Image Accuracy	Apolitical Real	Apolitical Fake	Overall Accuracy
Very Liberal	82.8	56.8	72.4	76.8	72.2	61.0	63.8	68.94
Liberal	81.6	64.8	72.0	74.4	73.2	61.6	56.8	68.53
Moderate	82.8	58.6	68.6	67.2	69.3	57.2	62.8	66.19
Conservative	70.6	58.6	66.6	72.0	66.95	62.6	60.0	68.22
Very Conservative	77.2	67.2	70.0	72.8	71.8	60.0	61.4	68.10

Note. All numbers are percentages. “Fake” means the image was AI-generated. Apolitical refers to the control images featuring unknown people. Bolded number is statistically significant.

To examine the role of political bias in participants’ evaluations of images, two mixed-effects logistic regression analyses were conducted to analyze responses based on congruence and test the two hypotheses. Mixed-effects logistic regression was utilized to account for the repeated-measures structure of the data, and random intercepts were included for participants.

Perceived realness (whether participants decided an image was real) was analyzed to assess response bias (Hypothesis 1), and classification accuracy (whether participants were correct in their image classification) was analyzed to assess if biases had an impact on overall performance. The analyses were expected to produce significant results to reflect the presence of bias.

Hypothesis 1: Main Effect of Congruency

As shown in Table 5, mixed-effects logistic regression revealed a marginal effect of congruency on participants' perceived image realness, $\beta = -0.224$, $SE = 0.116$, $z = -1.93$, $p = .054$. Participants were marginally less likely to classify politically incongruent images as real than politically congruent images.

Table 5 - Mixed-Effects Logistics Regression of Perceived Image Realness

Predictor	β (Estimate)	SE	z	p
Intercept	-0.089	0.108	-0.828	.408
Congruency (Incongruent)	-0.224	0.116	-1.929	.054

Note. DV = Perceived Realness (1 = classified as real). Negative coefficients indicate a lower likelihood of labeling images as real. Congruent images serve as the reference category.

Hypothesis 2: Interaction Between Political Affiliation and Image Congruency

A mixed-effects logistic regression model was conducted, predicting accuracy from image ideology (pro-liberal vs. pro-conservative), political affiliation, image truth (real vs. AI-generated), and their interactions.

As shown in Table 6, the analysis did not reveal a significant interaction between image ideology and political affiliation, indicating that the effect of image ideology on accuracy did not differ reliably across political groups. However, a significant main effect of image truth was

observed, $\beta = -1.79$, $SE = 0.44$, $z = -4.04$, $p < .001$, suggesting that participants were less accurate when evaluating real images. Pro-conservative real images were judged less accurately overall by participants (Table 4).

Table 6 - Mixed Effects Logistic Regression Predicting Accuracy

Predictor	β	SE	z	p
Intercept	2.148	0.383	5.615	< . .001
Image Ideology (Pro-Liberal)	-0.563	0.486	-1.159	.247
Political Standing (Liberal)	0.034	0.486	0.070	.944
Political Standing (Moderate)	-0.076	0.470	-0.162	.872
Political Standing (Very Conservative)	-0.741	0.491	-1.507	.132
Political Standing (Very Liberal)	0.428	0.466	0.919	.358
Image Truth (Real)	-1.792	0.444	-4.038	< . .001
Image Ideology $\times$ Image Truth	1.101	0.597	1.845	.065
Political Standing $\times$ Image Truth (Very Conservative)	1.111	0.594	1.872	.061

Note. DV = Accuracy (1 = correct). Reference categories: Pro-conservative images, Conservative political standing, and AI-generated images. Only key interaction terms are shown. Bolded p-values are statistically significant.

Hypothesis 3: Impact of Other Variables

To determine if age, gender, AI exposure (previous AI image exposure, AI usage, frequency of AI viewing), and social media usage had any effect on participant judgements of

image accuracy, ANOVAs were conducted for each variable. Contrary to the original hypothesis, age group stood out as a statistically significant predictor ($p < 0.05$) of image evaluation accuracy.

Table 7 - Demographic ANOVA Results

Variable	F-Statistic	df	η^2	$p < 0.05$
Age Group	4.74	5,120	0.165	0.0007
Gender	1.22	3,122	0.029	0.305
Seen AI Images	0.48	1,124	0.004	0.49
AI Tools Usage	1.71	4,121	0.054	0.152
AI Content Frequency	0.45	3,122	0.011	0.716
Social Media Frequency	1.18	3,122	0.028	0.321

Note: Critical F-value at $\alpha = 0.05$ varies by degrees of freedom. Bolded variable statistically significant.

Additionally, a linear accuracy trend emerged by age group, with older participants demonstrating slightly higher image detection accuracy. Table 8 displays detection accuracy by age group, with each older age group performing marginally better than the previous.

Table 8 - Image Evaluation Accuracy by Age Group

Age Group	N	Mean	SD	Min	Max
18-29	26	19.35	2.67	14	25
30-39	38	19.76	2.54	15	25
40-49	28	20.68	2.89	16	26

50-59	18	21.28	3.12	14	27
60-69	9	22.33	2.96	18	28
70+	7	23.14	3.44	19	29

Discussion

Previous studies examining political confirmation bias with text concluded that confirmation bias affects how participants evaluate news and which headlines they are more likely to believe. However, previous studies exposing participants to visual media did not find any significant evidence of confirmation bias when participants were evaluating political videos. The present study examined whether confirmation bias influences individuals' ability to distinguish between real and AI-generated images, focusing on both response tendencies (belief bias) and objective accuracy (performance bias). The results provide minimal support for confirmation bias, revealing slightly biased judgments in perceived realness of images, but the effect was not strong enough to affect classification accuracy. Though the presence of confirmation bias is not fully supported, additional findings create interesting avenues for further exploration.

Lack of Political Affiliation Effect

The significant effect found for real left-leaning images suggests that conservatives were less accurate than very liberal, liberal, and moderate participants when evaluating real images that favored a liberal position. This pattern is consistent with the idea that motivated reasoning or prior beliefs may selectively impair performance when the content conflicts with one's political orientation, though the effect size is small (*Cramér's $V = 0.16$*). Additionally, no significant group differences were found for real right-leaning images, fake images of either orientation, or overall political image accuracy. The logistic regression results reinforce the conclusion that

political affiliation matters only in interaction with specific image types and does not affect overall performance, and image type and truth remain the strongest determinants of accuracy.

Effects of Image Congruency on Evaluation Accuracy

Consistent with the first hypothesis, participants demonstrated political bias in their image evaluations, as data revealed they were more likely to believe an image was real when it was politically congruent. Though the main effect was marginally significant, the direction of effect still aligned with the predictions derived from confirmation bias. This finding suggests that personal beliefs can impact judgments (especially of authenticity), even in cases of visual stimuli. When evaluating visual stimuli like images, participants may still rely on heuristic processing and accept belief-congruent information as more truthful and bestow less skepticism. The greater participant belief in politically congruent images aligns with previous confirmation bias research, where participants were presented with text, and suggests that similar cognitive shortcuts are still used when presented with images. When viewing political content (through text or visually), one must remain wary of falsified material, especially when it aligns with our own political beliefs, as we tend to become less skeptical when information aligns with our preconceptions.

Though a main effect was observed, the effect was marginal, indicating that the influence of confirmation bias remained minimal. Participants' knowledge of the study design and their desire to answer correctly may have encouraged more cautious or analytical responses, dampening potentially stronger effects. Despite any caution applied by participants throughout

the study, marginal main effects of congruence were still observed, suggesting that participants were slightly more likely to believe an image was real if it aligned with their pre-existing beliefs.

Participants across political groups were less accurate when judging real pro-conservative images, especially when pro-conservative image accuracy is compared to other political images (Table 4). Uncertainty about the validity of real pro-conservative images could be due to the high volume of AI-generated images posted by one of the most well-known conservative politicians, current President Donald Trump. Since 2022, Trump has posted 62 AI-generated images or videos on his Truth Social account, both promoting him and his campaign and disparaging opponents; even in office, Trump has posted 36 AI-generated images and videos to Truth Social (Thompson, 2025; Tuquero, 2025). The frequency of AI-generated visual content from Trump may have caused participants to associate right-leaning images with AI, causing them to incorrectly believe the real images are AI-generated.

Potential Biases Uninfluential on Objective Classification Performance

Contrary to the second hypothesis, the data produced no significant interaction between political affiliation and image ideology on image classification accuracy. Participants' abilities to correctly distinguish between real and AI-generated images were not systematically influenced by whether the content aligned with their political beliefs. Although participants exhibited a tendency to perceive politically congruent images as real (confirmed by the results of Hypothesis 1), this bias did not translate into reduced accuracy. The lack of alignment between belief and accuracy indicates participants may hold biased perceptions at a subjective level without those biases necessarily impairing their ability to make correct classifications. While a participant's

initial judgments are potentially influenced by prior beliefs, they may engage in more deliberative processing when making final decisions, allowing them to correct for potential biases (demonstrated by the lack of difference in overall performance). The study structure, where participants were informed that their evaluations could be incorrect and were required to make explicit decisions, likely led participants to make a more analytical response utilizing higher-order cognition.

The dissociation between response bias and accuracy indicates that confirmation bias may operate at an initial evaluative stage without necessarily impairing ultimate task performance, suggesting that increased cognitive load can help individuals move past initial bias and provide better aid in evaluative situations. When faced with political content, knowledge that information can be false and employing higher-order cognition can potentially prevent individuals from believing falsified content.

Decreased Accuracy of Real Image Evaluations

Participants were less accurate when evaluating real images compared to AI-generated images, demonstrating a significant main effect of image truth. This pattern suggests a general bias toward identifying images as AI-generated, which may reflect increased skepticism or uncertainty due to unfamiliarity with rapidly advancing AI technology. Such a tendency could have reduced the likelihood that confirmation bias would manifest in accuracy, as participants may have defaulted to a more cautious response strategy across conditions to compensate for a lack of knowledge. AI-generated images come in many levels of realism, leaving participants to question whether the real images were extremely realistic AI and to choose accordingly.

Increased Accuracy with Age

Despite all other external variables having no significant impact, age played a significant role, with older participants evaluating images more accurately than younger participants. While all participants 60+ had seen AI-generated visual content before, and most had frequently interacted with AI output or tools, the successful identification of the older participants could be due to longer lifetime exposure to real faces and less familiarity with digitally modified faces. Younger individuals, who use social media more, would likely be exposed to more edited and filtered faces, potentially impacting their schema of “normal” faces and inadvertently adjusting their authenticity threshold to include digitally modified faces. However, further exploration of this finding with a larger sample size is required to draw conclusions.

Limitations and Future Directions

One of the major limitations of this study was the small sample size, limited by study funding. Though this study found age to be a statistically significant predictor of image evaluation accuracy, the sample size for older participants was extremely small (ages 70+ $N=7$, ages 60-69 $N=9$, and ages 50-59 $N=18$). The present study should be run with a larger age 50+ participant pool to determine if age remains a significant predictor of classification accuracy. The current age findings contradict general beliefs that younger individuals are better than older individuals at understanding AI; further research exploring if older individuals are better at detecting AI in other forms (ex. written text) could potentially contradict long-held beliefs about human interaction with technology. Investigating the cognitive processes that underlie age accuracy differences could also shape AI education approaches.

Investigating potential patterns with image analysis accuracy and time taken on the study could determine whether more deliberative processing helped participants in assessing image type. While participant response time was collected, it was not analyzed for this study; the negligible impact of confirmation bias suggests that a future replication can examine response time effects further.

Additionally, further replications of this study with more images and a larger, more diverse participant pool could examine whether the same effects and patterns are found in a more diverse sample size. While the AI materials for the present study were sourced online, replications with AI-generated images from the most advanced platforms would reflect participants' detection abilities with images they are more likely to encounter naturally, as AI technology has advanced since the images were created and sourced.

The second limitation of the study was that political affiliation questions were mainly self-reported, and asking participants their alignment with liberalism/conservatism or the Democratic/Republican parties may not have captured their true political stance. Additionally, image-political affiliation congruency would only be greatly aligned if participants had a favorable opinion of the politicians of their party. Future replication studies with AI-images generated specifically for the study will strengthen the current findings.

Conclusion

The effects of confirmation bias on individuals' abilities to discern between real and AI-generated political and apolitical images remain negligible. Despite a tendency to perceive politically congruent images as real, participants did not show reduced accuracy in overall image classification. The marginal response bias having no significant impact on participants' overall image classification opens avenues for further research. The confirmation bias could have operated at an initial stage that was later controlled by cognition and did not affect performance accuracy. As participants knew some images were real and some were AI-generated, they likely were more aware of potential biases and took more time to evaluate images. Though participants of all political affiliations had similar image accuracy scores, all participants performed worse on real pro-conservative images. The performance deficit at just the one image condition could be explained by right-leaning politicians posting more AI, leading participants to become more doubtful of pro-conservative images; however, this phenomenon requires further research.

Importantly, an effect of age emerged in participant image evaluation data. As age increased, participants became more accurate with their image classifications, contrary to general assumptions about AI knowledge and detection ability.

As AI-generated content becomes more sophisticated and ubiquitous, understanding the individual differences in detection ability will provide crucial insight into AI education and training. This study demonstrated that overall, people do not let political biases impact their judgment of images, knowing that there is the potential that an image is AI-generated.

Awareness of the possibility that an image is AI-generated can help prevent individuals from

believing false images, demonstrating support for AI-generated post labels on social media and news platforms. Understanding how individuals cognitively mitigate biases in judgment and decision-making can foster greater collaboration, communication, and understanding, especially in an era where all media has the potential to be fabricated.

Appendix A

General Result Data Analysis

Table 9 - Overall Performance Accuracy by Political Affiliation

Political Affiliation	Liberal Real	Conservative Real	Liberal Fake	Conservative Fake	Apolitical Real	Apolitical Fake	Number Correct	Accuracy (%)
Very Liberal	4.14	2.84	3.62	3.84	3.05	3.19	20.68	68.94
Liberal	4.08	3.24	3.60	3.72	3.08	2.84	20.56	68.53
Moderate	4.14	2.93	3.43	3.36	2.86	3.14	19.86	66.19
Conservative	3.53	2.93	3.33	3.60	3.13	3.00	20.47	68.22
Very Conservative	3.86	3.36	3.50	3.64	3.00	3.07	20.43	68.10

Note. Fake means the image was AI-generated. Apolitical refers to the control images.

Table 10 – Chi-Squared Test Results

Column	χ^2	p-value
Pro-Liberal Real Images	12.9	0.011
Pro-Conservative Real Images	6.4	0.17
Pro-Liberal AI-generated Images	4.8	0.31
Pro-Conservative AI-generated Images	3.9	0.41

Political Image Accuracy	1.8	0.77
--------------------------	-----	------

Appendix B

Image Stimuli Presented to Participants

Pro-Liberal Real Images:

1. Image 1:



2. Image 2:



3. Image 3:



4. Image 4:



5. Image 5:



Pro-Conservative Real Images:

1. Image 1:



2. Image 2:



3. Image 3:



4. Image 4:



5. Image 5:



Pro-Liberal AI-generated Images:

1. Image 1:



2. Image 2:



3. Image 3:



4. Image 4:



5. Image 5:



Pro-Conservative AI-generated Images:

1. Image 1:



2. Image 2:



3. Image 3:



4. Image 4:



5. Image 5:



Control Real Images:

1. Image 1:



2. Image 2:



3. Image 3:



4. Image 4:



5. Image 5:



Control AI-generated Images:

1. Image 1:



2. Image 2:



3. Image 3:



4. Image 4:



5. Image 5:



Appendix C
Political Evaluation and Demographic Questions

Political Questions:

1. Generally speaking, I identify as a liberal
 - a. Strongly disagree
 - b. Disagree
 - c. Neither agree nor disagree
 - d. Agree
 - e. Strongly agree
2. Generally speaking, I identify as a conservative
 - a. Strongly disagree
 - b. Disagree
 - c. Neither agree nor disagree
 - d. Agree
 - e. Strongly agree
3. Generally speaking, my views align with the current Democratic party
 - a. Strongly disagree
 - b. Disagree
 - c. Neither agree nor disagree
 - d. Agree
 - e. Strongly agree
4. Generally speaking, my views align with the current Republican party
 - a. Strongly disagree

- b. Disagree
 - c. Neither agree nor disagree
 - d. Agree
 - e. Strongly agree
5. I have a favorable opinion of Donald Trump
- a. Strongly disagree
 - b. Disagree
 - c. Neither agree nor disagree
 - d. Agree
 - e. Strongly agree
6. I have a favorable opinion of Kamala Harris
- a. Strongly disagree
 - b. Disagree
 - c. Neither agree nor disagree
 - d. Agree
 - e. Strongly agree

Demographic and AI Familiarity Questions:

1. Select your age
- a. 18 – 29
 - b. 30 – 39
 - c. 40 – 49
 - d. 50 – 59

- e. 60 – 69
 - f. 70+
2. Select your gender identity
- a. Male
 - b. Female
 - c. Non-binary
 - d. Other
3. How frequently do you come across AI content on the Internet?
- a. Never
 - b. Rarely
 - c. Sometimes
 - d. Often
 - e. Always
4. How frequently do you use AI tools?
- a. Never
 - b. Rarely
 - c. Sometimes
 - d. Often
 - e. Always
5. Have you come across AI-generated images or videos before?
- a. Yes
 - b. No
6. How frequently do you use social media?

- a. Never
- b. Rarely
- c. Sometimes
- d. Often
- e. Always

BIBLIOGRAPHY

- Ahler, D. J., Citrin, J., Dougal, M. C., & Lenz, G. S. (2016). Face value? Experimental evidence that candidate appearance influences electoral choice. *Political Behavior*, 39(1), 77–102.
<https://doi.org/10.1007/s11109-016-9348-6>
- AI misinformation control 2025 | Statista. (2025). Statista.
<https://www.statista.com/statistics/1615571/ai-misinformation-impact-management/>
- Barry, A. M. (1997). *Visual intelligence : Perception, image, and manipulation in visual communication*. State University Of New York Press, Cop.
- Beauvais, C. (2022). Fake news: Why do we believe it? *Joint Bone Spine*, 89(4).
<https://doi.org/10.1016/j.jbspin.2022.105371>
- Calvillo, D. P., Rutchick, A. M., & Garcia, R. J. B. (2021). Individual differences in belief in fake news about election fraud after the 2020 U.S. election. *Behavioral Sciences*, 11(12), 175. <https://doi.org/10.3390/bs11120175>
- Del Vicario, M., Scala, A., Caldarelli, G., Stanley, H. E., & Quattrociocchi, W. (2017). Modeling confirmation bias and polarization. *Scientific Reports*, 7(1).
<https://doi.org/10.1038/srep40391>
- Ebner, J., Lutz, B., & Neumann, D. (2025). Deep fakes and political attitudes: An analysis of confirmation bias and cognitive dissonance using neurophysiological measurements. *Lecture Notes in Information Systems and Organisation*, 299–311.
https://doi.org/10.1007/978-3-031-71385-9_26

Goldstein, J. A., Sastry, G., Musser, M., DiResta, R., Gentzel, M., & Katerina Sedova. (2023).

Generative language models and automated influence operations: Emerging threats and potential mitigations. ArXiv. <https://doi.org/10.48550/arxiv.2301.04246>

Graber, D. A. (1990). Seeing is remembering: How visuals contribute to learning from television news. *Journal of Communication*, 40(3), 134–156. <https://doi.org/10.1111/j.1460-2466.1990.tb02275.x>

Hicks, J. (2026). AI-powered software is helping misinformation spread online after disasters. NPR. <https://www.npr.org/2026/01/01/nx-s1-5645183/ai-powered-software-is-helping-misinformation-spread-online-after-disasters>

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Klayman, J. (1995). Varieties of confirmation bias. *Psychology of Learning and Motivation*, 32(1), 385–418. [https://doi.org/10.1016/s0079-7421\(08\)60315-1](https://doi.org/10.1016/s0079-7421(08)60315-1)

Knobloch-Westerwick, S., Johnson, B. K., & Westerwick, A. (2014). Confirmation bias in online searches: Impacts of selective exposure before an election on political attitude strength and shifts. *Journal of Computer-Mediated Communication*, 20(2), 171–187. <https://doi.org/10.1111/jcc4.12105>

Ludwig, J., & Sommer, J. (2024). Mindsets and politically motivated reasoning about fake news. *Motivation and Emotion*, 48. <https://doi.org/10.1007/s11031-024-10067-0>

Marr, B. (2025, March 10). 15 mind-blowing AI statistics everyone must know about now. *Forbes*. <https://www.forbes.com/sites/bernardmarr/2025/03/10/15-mind-blowing-ai-statistics-everyone-must-know-about-now/>

- Michael, R. B., & Breaux, B. O. (2021). The relationship between political affiliation and beliefs about sources of “fake news.” *Cognitive Research: Principles and Implications*, 6(1).
<https://doi.org/10.1186/s41235-021-00278-1>
- Mitchell, A., Jurkowitz, M., Oliphant, J. B., & Shearer, E. (2020, July 30). Americans who mainly get their news on social media are less engaged, less Knowledgeable. Pew Research Center’s Journalism Project; Pew Research Center.
<https://www.pewresearch.org/journalism/2020/07/30/americans-who-mainly-get-their-news-on-social-media-are-less-engaged-less-knowledgeable/>
- Moravec, P. L., Minas, R. K., & Dennis, A. R. (2019). People believe what they want to believe when it makes no sense at all. *MIS Quarterly*, 43(4), 1343-A13. JSTOR.
<https://doi.org/10.2307/26848107>
- Newhagen, J. E. (1998). TV news images that induce anger, fear, and disgust: Effects on approach-avoidance and memory. *Journal of Broadcasting & Electronic Media*, 42(2), 265–276. <https://doi.org/10.1080/08838159809364448>
- Nightingale, S. J., Wade, K. A., & Watson, D. G. (2017). Can people identify original and manipulated photos of real-world scenes? *Cognitive Research: Principles and Implications*, 2(30). <https://doi.org/10.1186/s41235-017-0067-2>
- Powell, T. E., Boomgaarden, H. G., De Swert, K., & de Vreese, C. H. (2015). A clearer picture: The contribution of visuals and text to framing effects. *Journal of Communication*, 65(6), 997–1017. <https://doi.org/10.1111/jcom.12184>
- Rajsic, J., Taylor, J. E. T., & Pratt, J. (2016). Out of sight, out of mind: Matching bias underlies confirmatory visual search. *Attention, Perception, & Psychophysics*, 79(2), 498–507.
<https://doi.org/10.3758/s13414-016-1259-4>

- Rosenberg, S. W., Bohan, L., McCafferty, P., & Harris, K. (1986). The image and the vote: The effect of candidate presentation on voter preference. *American Journal of Political Science*, 30(1), 108. <https://doi.org/10.2307/2111296>
- Stanovich, K. E. (1999). *Who is rational?* Psychology Press.
- Taylor, S. E., & Fiske, S. T. (1978). Salience, attention, and attribution: Top of the head phenomena. *Advances in Experimental Social Psychology*, 11, 249–288. [https://doi.org/10.1016/s0065-2601\(08\)60009-x](https://doi.org/10.1016/s0065-2601(08)60009-x)
- Wittenberg, C., Tappin, B. M., Berinsky, A. J., & Rand, D. G. (2021). The (minimal) persuasive advantage of political video over text. *Proceedings of the National Academy of Sciences*, 118(47), e2114388118. <https://doi.org/10.1073/pnas.2114388118>
- Yadav, A., Phillips, M. M., Lundeberg, M. A., Koehler, M. J., Hilden, K., & Dirkin, K. H. (2011). If a picture is worth a thousand words is video worth a million? Differences in affective and cognitive processing of video and text cases. *Journal of Computing in Higher Education*, 23(1), 15–37. <https://doi.org/10.1007/s12528-011-9042-y>
- Yang, Y., Davis, T., & Hindman, M. (2023). Visual misinformation on Facebook. *Journal of Communication*, 73(4). <https://doi.org/10.1093/joc/jqac051>