7-2023

Security of Text to Image Conversions

Zobaida Alssadi

Security of Text to Image Conversions

by
Zobaida Alssadi

Bachelor of Science
Computer Science
The University of Bisha
2016

A thesis submitted to the
College of Engineering and Science at
Florida Institute of Technology
in partial fulfillment of the requirements
for the degree of

Master of Science
in
Computer Information Systems

Melbourne, Florida
July, 2023

We the undersigned committee hereby recommend that the attached document be accepted as fulfilling in part the requirements for the degree of Master of Science in Computer Science.

"Security of Text to Image Conversions" a thesis by Zobaida Alssadi

---

Marius C. Silaghi, Ph.D.
Professor
Computer Engineering and Sciences
Major Advisor

Sneha Sudhakaran PhD
Assistant Professor
Electrical Engineering and Computer Science

Juan Camilo Avendano, PhD
Director of CAMID
Graduate Faculty, Computer Engineering & Sciences
Outside Committee Member

Philip J. Bernhard, Ph.D.
Associate Professor and Department Head
Computer Engineering and Sciences

Abstract

Title: Security of Text to Image Conversions

Author: Zobaida Alssadi

Thesis Advisor: Marius C. Silaghi, Ph.D.

The use of images and icons to represent news or narratives has grown in popularity. Still, one critical problem is that they are not equivalent to language, making them vulnerable to adversary attacks. This study examines the impact of image-poisoning attacks based on polysemantic words and of image attacks based on cultural differences when converting text to images. Such attacks can lead to the loss of important information and create confusion and incorrect interpretations of the intended meaning, misinforming the general public. The study specifically focuses on possible effects in a news and story context. This study highlights the significance of taking security considerations into account when image-based attacks are relevant and motivates the development of strategies to ensure that information is conveyed through images and icons in a culturally appropriate and accurate manner, as well as to prevent image tampering and the spread of false information by attackers.

Table of Contents

List of Figures

Acknowledgements

I want to express my sincere thanks and appreciation to everyone who has helped and encouraged me during my master's thesis journey. First and foremost, I'd like to thank God for the numerous blessings I've received along this trip. I would like to thank my supervisor, Dr. Marius Silaghi, for his helpful advice, encouragement, and knowledge. His guidance has been valuable in influencing my study and assisting me in developing a more thorough understanding of the topic. I would like to express my heartfelt gratitude to the members of my thesis committee for providing constructive input and valuable ideas that have improved the quality of my work. A special thanks to my parents, Obaid Alssadi and Shuaeaa Alssadi, for their unfailing love, support, and encouragement throughout my academic journey. Your faith in me and my skills has been the cornerstone of my success, and I am eternally thankful for your patience and sacrifices that have allowed me to follow my dreams. I would also like to thank myself for all my hard work, commitment, and persistence along this trip. This accomplishment demonstrates my commitment, and I am proud of the effort I put in to obtain my dream. Finally, I'd like to thank the Saudi Arabian Cultural Mission (SACM) for its financial assistance in allowing me to focus on my study and finish this thesis. Thank you for your essential contributions to both my academic and personal development. This achievement would not have been achieved without

the help of each and every one of you.

Dedication

I dedicate this thesis to my dear parents, my father, who was my first and most beloved teacher. He taught me honesty, perseverance, and dealing with life's challenges. Through his dedication to educating and guiding me from a young age, he has inspired me to pursue my dreams and keep learning and succeeding. My beloved mother, your sincere prayers have always been the secret of my strength and encouragement to continue to succeed. Without your boundless support and great love, I would not have been able to achieve what I am today. I hope that I have made you happy with my achievement and that my research will be a valuable addition to the knowledge and a source of pride for you.

Chapter 1

Introduction

While the automatic generation of image-based representation of texts becomes more common, we highlight new challenges and attacks on their applications involving cultural differences and poisoned image selection. Images/icons have gained popularity due to their visual appeal, attention-grabbing qualities, communication efficiency, and adaptability to different mediums. Technology has made it easier to create, share, access, and accept images/icons, leading to increased use in computer applications [8].

Natural language processing and computer vision have both focused on the conversions of text to images/icons. These techniques train and evaluate models using human-annotated images/icons where the amount and quality of the training data are generally correlated to the achieved level of performance [10].

There are, however, limitations and challenges associated with this where image-poisoning attacks based on polysemantic words can create visually related but semantically incorrect icons or images. On the other hand, image attacks based on cultural

differences could cause misconceptions and confusion, and these differences may alter how images/icons are understood.

In light of recent developments in Chatbot and BARD models, security vulnerabilities have become a growing concern. This work explores the challenges of text-to-image/icon generation, including adversarial attacks and cultural differences that affect the interpretation of images/icons. By gaining a deeper understanding of these challenges, our research can contribute to the identification of effective defenses for enhancing the accuracy and appropriateness of visual communication using images and icons. This can have significant implications for developing more reliable and robust text-to-image/icon generation techniques and also impact the field of natural language processing and computer vision.

1.1 Background

The main topics of previous research and publications have been the expressiveness of images/icons and their capability to clearly convey complicated ideas to a broad audience. One (AI-based) tool, Text2Icons, converts text descriptions into icons, improving comprehension and making the recall of complex information simpler. The system determines the best icons to reflect the meaning of the text input using various NLP and machine learning methods [28].

The approach of the Text2Icons tool is to convert the text of a story or news into a sequence of icons. The tool analyses text inputs and then extracts the features linked

to the comprehension the text's content and chooses the suitable icons to convey the story. The system converts text into icons using a specified icons database. Machine learning algorithms are trained to determine which icons correspond to specific text feature inputs. The resulting narrative or news is then shown as a series of icons that convey the text input's intended meaning. The system purpose is to make text-based stories or news more accessible and understandable by presenting them in a coherent visual manner. The possible advantages of utilizing icons to communicate stories include better knowledge retention, more straightforward communication, and more compelling material [28]. The used techniques include deep learning NLP methods.

Adversarial attacks could significantly impact a deep learning-based NLP system's accuracy. These attacks aim to change the input to the AI model in a way that disturbs the model's output, producing false or misleading results [17]. Image poisoning attacks are attacks on machine learning models where the attacker alters the training data to make the model act incorrectly to misclassify a particular sample or class maliciously. There are two types of image poisoning attacks identified in [25]: Triggerless and backdoor attacks.

1. **Triggerless attacks:** In techniques like Feature Collision (FC) and Convex Polytope (CP), poison samples are produced by introducing minor alterations to the base pictures to make their feature representations resemble those of the target image.

2. **Backdoor attacks:** like Class-Conditional Label-Based Poisoning (CLBD) and

High-Transfer Backdoor Poisoning (HTBD), develop poison examples with a concealed trigger, like a particular patch or pattern that causes the model to categorize incorrectly the target image when the trigger is present.

Adversarial attacks on images using deep neural networks (DNNs) threaten the security and reliability of DNN-based systems. These attacks exploit the fact that DNNs are highly sensitive to even minute input changes, even if those changes are undetectable to the human eye. Image adversarial attacks are techniques used to carefully and gradually change an image while keeping its visual appearance, leading a machine learning model to identify the modified image incorrectly. This attack comes in a variety of forms, such as:

1-Perturbation-based attacks: These attacks poison the original image by adding noise to deceive the model into classifying it incorrectly.

2-Reconstruction-based attacks: These attacks involve reconstructing the image from scratch to cause the model to misclassify it. Examples include the Carlini-Wagner attack.

3- Evasion attacks: By taking advantage of weaknesses in the model's design or training set, these attacks aim to get the model to incorrectly identify an image [22].

The expressiveness of icons in expressing meaning and information has been assessed in earlier studies and papers. The study in [5] looked at variables that can affect an icon's expressiveness, including cultural background, user familiarity, level

of abstraction, usage of color and animation, and the inclusion of text labels. The findings demonstrated that expressiveness could be increased by considering the cultural background, user familiarity, and level of abstraction and by using well-designed, concrete, and recognized iconography. It was discovered that using text labels with icons enhanced their expressiveness. It has been conjectured that it is difficult to produce icons that are both internationally recognizable and culturally neutral [9].

1.2 Problem Statement

The problem with converting text to icons and images is that it can be vulnerable to various types of attacks, including: image-poisoning attacks based on polysemantic words and image attacks based on cultural differences. Attacks that use "image poisoning" alter the icon selection and produce false or damaging images using words with many meanings. Cultural differences can also result in misinterpretations and misunderstandings of icons and images, spreading incorrect information and deceiving the general public.

Here we look at how these adversarial attacks, which may lead to the loss of important information or details, can have an impact. They can result in misunderstandings or incorrect interpretations of the intended meaning. The study specifically focuses on the possible effects in a news and story context based on an individual's background, culture, and educational level, where it is possible for erroneous information to propagate or for the general public to be misinformed.

The study's findings demonstrate the importance of protecting against hackers tampering with images and utilizing them to disseminate false information to ensure accurate and suitable information exchange. It also indicates the relevance of considering these attacks' security risks to ensure that icons and images are culturally appropriate and do not cause misunderstanding or offense.

1.3 Research questions and objectives

1.3.1 Research Questions

The following research questions guide this study:

1. What are the potential impacts of image-poisoning attacks based on polysemantic words on the interpretation of news or narratives?

2. How can image attacks based on cultural differences affect how information is understood and interpreted through images and icons?

3. What are some possible strategies for ensuring that information is presented through images and icons in an accurate and appropriate way from the perspective of culture?

4. What can be done to stop attackers from tampering with pictures and spreading misleading information?

1.3.2 Objectives:

1. To evaluate the susceptibility of the Text2Icon tool to image-poisoning attacks.

2. To investigate how polysemantic word-based image-poisoning attacks could affect news and narrative interpretation.

3. To investigate how cultural differences affect how information based on images and icons is understood and interpreted.

1.4 Significance of the Study

The significance of this study relies on emphasizing the security vulnerabilities that could occur when text is converted into icons and images, especially when news and story sharing are involved. The study highlights the vulnerability of such programs to image-poisoning attacks and cultural differences, which can lead to the dissemination of misleading information and misunderstandings.

Chapter 2

Literature Review

2.1 Overview of Image and Icon-based Communication

Visual communication has a significant history that dates back to the days when people first started using symbols and images to tell stories and express ideas. With the discovery of writing systems and the development of printing presses, photography, and digital media, visual communication has advanced. Today, many messages may be expressed using visual communication, from straightforward facts (like traffic signs) to intricate concepts and feelings (such as political messages or artistic expressions).

In marketing and advertising, for example, images and icons are commonly used to establish a memorable brand identity and to connect the company with particular ideals or feelings. A logo or icon could suggest innovation, safety, or luxury, depending

on the style and context.

In journalism and the media, images, and icons are commonly used to illustrate stories and emphasize a point visually. A narrative can gain human interest by using images, infographics, and charts to assist readers in understanding complex concepts and facts.

Regarding digital communication, activities or functions are frequently represented by icons, such as the "like" button on social media or the "search" icon on a website. For users to interact with the interface without having to read or comprehend written instructions, these symbols are frequently created to be instantly recognized and intuitive [26].

2.2 Natural Language Processing and Computer Vision

2.2.1 Natural Language Processing (NLP)

.

Natural Language Processing (NLP) is an artificial intelligence (AI) area concerned with how language and computers interact, developing algorithms and methods for computers to analyze, process, and produce natural language text, speech, and other communication.

Many uses of NLP include sentiment analysis, text summarization, chatbots, voice

assistants, and language translation. Moreover, many subfields of NLP are involved, such as language modeling, part-of-speech tagging, named entity recognition, text categorization, information extraction, and text summarization.

NLP algorithms frequently use machine learning and deep learning techniques to exploit significant text input and enhance their performance over time. Recent developments in deep learning and neural networks have significantly improved NLP applications, including language translation and text summarization.[29]

2.2.2 Computer Vision

.

Artificial intelligence includes computer vision, which gives robots the ability to recognize and grasp the visual environment, which is its primary concern. It entails the creation of techniques and algorithms for decoding, examining, and rating visual data, including movies and photographs. Deep learning methods in computer vision include convolutional neural networks (CNNs) and recurrent neural networks. Tracking, segmenting, and categorizing objects are other strategies (RNNs).

Robotics, autonomous vehicles, medical imaging, and monitoring are just a few applications for it. Recent advancements in computer vision have been fueled by deep learning, making it possible to design more accurate and efficient algorithms for tasks like object recognition and picture segmentation.

To create more potent systems, like visual question answering (VQA) and natural

language generation, NLP and computer vision can be coupled (NLG).[29]

2.2.3 Applications of Natural Language Processing and Computer Vision in Image and Icon Generation

2.3 DALLE 2

The OpenAI program DALLE 2 is known for converting natural language inputs into realistic images. Applying a deep learning architecture, this program shows a significant ability to analyze the components of the text and introduce newly generated visuals with high accuracy and fidelity. DALLE 2 has been connected to the comprehension of human language and the invention of a language that is not immediately intuitive to humans [14].

DALL-E 2 is a powerful generative model that can produce high-quality images of a wide range of objects, scenes, and concepts. It can create variations of the same image depending on different textual inputs, and it can handle complex textual descriptions, including many objects or relationships between objects. This makes it possible to engage in creative discovery and widens the possibilities in fields like art and design. DALL-E 2 represents a significant advancement in computer vision and natural language processing thanks to its capacity to provide a variety of realistic visuals that correspond to the textual input [4]

An example output is shown in Figure 2.1 for the text:

"Einstein paints while Newton solves puzzles. Einstein is wearing a lab coat. Newton is wearing a t-shirt, pants, and a cap."

Figure 2.1: Image generated by DALLE 2 with the prompt: *"Einstein paints while Newton solves puzzles. Einstein is wearing a lab coat. Newton is wearing a t-shirt, pants, and a cap."*

DALL-E 2 is excellent at capturing images of many things but has some limits. As an illustration, it sometimes produces images that are impossible to understand since it does not fully comprehend the purpose of what is being asked of it. It is confined to only capturing things it is already familiar with because it is unable to grasp objects it has never seen before [4].

2.4 Midjourney

Midjourney is another AI and machine learning program that uses text questions to generate visuals. It is an open-access program made available to the general public in July 2022 and intended to produce visuals from written descriptions. These results are produced by comparing many internet images to what the artificial intelligence interprets as an accurate representation of a user-submitted request [11].

Midjourney can produce creative images, which could enhance the fairytale's visual storytelling. It has the potential as a tool for cultural heritage and storytelling. Educators and storytellers could use the software to generate images from the text that can enhance the engagement and imagination of their audiences. It can also provide a valuable tool for preserving cultural heritage, particularly for cultures and stories with limited visual representations.

However, there are the limitations of the software, such as the potential for bias in the selection of source images and the lack of control over the final output. For example, it has been used the prompt "evil stepmother," the software generated an image of a woman with dark skin, which could be seen as perpetuating racial stereotypes [23].

The exact text script has been used to illustrate a sample of this tool's potential and capability, and Figure 2.2 shows the result.

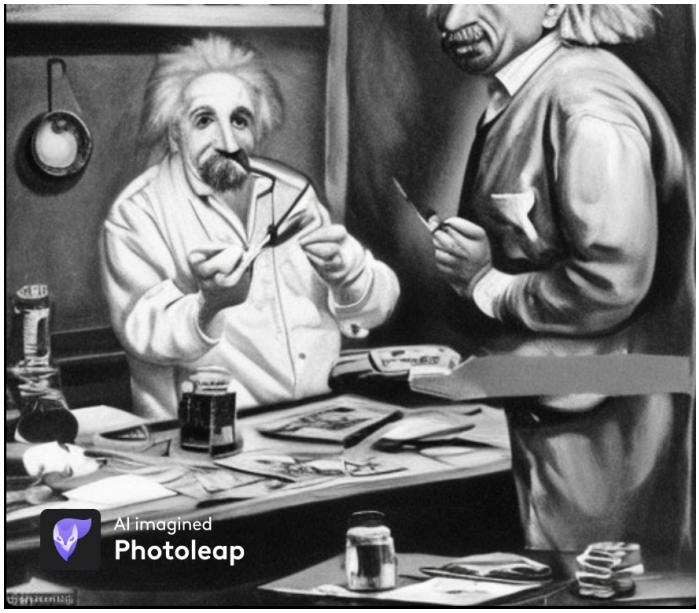

Figure 2.2: Image generated by MidJourney with the prompt: *"Einstein paints while Newton solves puzzles. Einstein is wearing a lab coat. Newton is wearing a t-shirt, pants, and a cap."*

2.5 Text2Icon

Text2Icon is a project that aims to revolutionize how we represent information using icons. The project proposes a new methodology for building icon dictionaries, automatic icon search, and a new icon-based visualization method successfully integrated into the project's pipeline. It addresses some of the challenges associated with narrative extraction, one of the critical tasks required to gather information about a story.

The Text2Icon pipeline consists of several steps to extract narrative elements and build a narrative visualization. The pipeline first extracts sentences from the initial text of the news stories and then integrates automatic translation methods to support the extraction of actors' descriptions. The Actor Level Resolution Algorithm is used

to find the most specific descriptions of the actors in the news stories, ensuring the visualization is accurate and meaningful. Tt presents the icon dictionary designed to be used as a database to generate icon visualization. The explored icon sources found adequate were integrated into the dictionary, including APIs such as emojidex, IconFinder, Icons8, and the Icons-50 dataset. The dictionary has two types of construction available, that is, two ways of adding new icons to the dictionary: semi-automatically or automatically.

Text2icon employs several algorithms and techniques to generate icons to represent narrative elements in news stories. Here are some of the key ones:

1. **Natural Language Processing (NLP) techniques:** NLP automatically extracts narrative elements from news stories. This includes tasks such as sentence segmentation, named entity recognition (NER), and part-of-speech (POS) tagging. These techniques help identify the key actors and events in a story, which can be represented using icons.

2. **Translation methods:** Since the Text2icon project is designed to work with news stories from different languages, translation methods ensure that the narrative elements are extracted accurately. The project uses two translation methods: Hugging Face Transformers and Googletrans. These methods use deep learning models to translate news stories from their original language to English.

3. **Actor Level Resolution Algorithm:** This algorithm identifies the most spe-

cific descriptions of the actors in the news stories. It works by resolving the different levels of specificity in actor descriptions, such as generic terms like "man" or "woman," versus more specific terms like "CEO" or "doctor."

4. **Icon APIs and datasets:** The Text2icon project uses several icon APIs and datasets to generate icons. These include emojidex, IconFinder, Icons8, and the Icons-50 dataset. These sources provide many icons that can represent different narrative elements.

5. **Word-emoji embeddings:** The project also uses word-emoji embeddings as a linking tool between emojis and words. This technique helps to ensure that the icons generated by the project accurately represent the narrative elements intended to be represented.[28]

The authors propose an approach that combines natural language processing (NLP) and computer vision techniques to generate a set of relevant icons based on a given input text. The approach involves several steps:

1. **Text preprocessing:** The input text is tokenized, and stop words, punctuation, and other extraneous information are deleted.

2. **Feature extraction:** To identify essential concepts and entities in the text, such as persons, locations, and things, NLP approaches are applied.

3. **Icon database:** The approach uses a database of pre-existing icons related to the extracted properties.

4. **Icon matching:** Based on their visual and semantic similarity, the retrieved characteristics are compared with the closest icons in the database.

5. **Icon arrangement:** The icons are organized in a sequence pattern to form a visual representation of the input text.

Figure 2.3 shows the generation of the following text into icons using Text2Icons.

Breaking News: The price of Bitcoin crashed after Russian President Vladimir Putin announced a military operation in the Donbas region of Ukraine.

Figure 2.3: Image generated by Text2Icon with the prompt: *Breaking News: The price of Bitcoin crashed after Russian President Vladimir Putin announced a military operation in the Donbas region Ukraine.*

2.5.1 Advantages of Text-to-Image Conversion

There are advantages associated with text-to-image conversion:

1. **Efficiency:** Text-to-image converting can save time and costs by allowing users to quickly develop visual representations of their ideas without building elaborate drawings or models.

2. **Flexibility:** Users may readily tweak and experiment with their ideas using text-to-image conversion by simply modifying the input text.

3. **Communication:** Text-to-image conversion can help providers, clients, and stakeholders communicate more effectively by offering a clear and aesthetically attractive subject depiction.

4. **Visualization:** Text-to-image conversion can assist users in better visualizing and exploring many text versions, ultimately leading to more original and creative results.

5. **Accessibility:** Those who have difficulties reading standard 2D drawings or models may benefit from text-to-image conversion. [31]

2.5.2 Challenges of Text-to-Image Conversion

There are several challenges associated with text-to-image conversion:

1. **Ambiguity:** Text can be ambiguous, leaving room for interpretation by the AI model. This can lead to inaccurate or unintended image outputs.

2. **Complexity:** The complexity of natural language can make it difficult for AI models to understand and convert text into images accurately. For example, understanding abstract concepts or metaphors in the text can be challenging.

3. **Data availability:** Text-to-image conversion requires large amounts of data, including text and image pairs. However, obtaining high-quality, diverse data can be difficult and time-consuming.

4. **Quality control:** Ensuring the quality of generated images is challenging, as it requires human evaluation and feedback. This can be time-consuming and may require a large number of human evaluators.

5. **Ethical considerations:** As with any AI technology, ethical considerations are associated with text-to-image conversions, such as the potential for bias or perpetuating harmful stereotypes. It is essential to consider these ethical implications and address them [31].

2.6 Image Generation and Adversarial Attacks

2.6.1 Existing Approaches for Text-to-Image Generation

Text-to-image generation is a challenging task in computer vision and natural language processing that involves generating a realistic image from a textual description. Here are some existing approaches for text-to-image generation:

1. **Generative Adversarial Networks (GANs):**

   AttnGAN is a state-of-the-art method for generating high-quality images from textual descriptions. It uses an attentional generative adversarial network (GAN) architecture incorporating attention mechanisms to guide the generation process. The model is trained on a large dataset of text-image pairs and can generate highly realistic and visually diverse images.

   The AttnGAN model consists of a generator and a discriminator network. The

generator takes as input a textual description of an image and produces an image that matches the description. On the other hand, the discriminator network is trained to distinguish between authentic and generated images.

One of the critical innovations of AttnGAN is the attention mechanism used by the generator. The attention mechanism lets the generator focus on different parts of the input text when generating other image regions. This improves the ability of the generator to create fine-grained details consistent with the input text [30].

2. **Variational Autoencoders (VAEs):**

   VAEs are a type of autoencoder that learn a compressed representation (or encoding) of input data, such as images. This encoding is then used to generate new data similar to the original input data. Unlike traditional autoencoders, VAEs use a probabilistic framework to create the encoding, allowing more flexibility in generating new data.

   Specifically, VAEs use a two-part training process. In the first part, the network is trained to encode input data into a probability distribution (usually a Gaussian distribution) of latent variables that capture the essential features of the data. The network is trained to decode the latent variables into the original input data in the second part.

   One advantage of VAEs is that they can generate new data similar to the original data by sampling new points in the latent variable space and decoding them

back into the input space. However, VAEs can sometimes generate blurry or low-quality images, which is why the authors propose to use a combination of stacked conditional Variational Autoencoders (CVAEs) and conditional Generative Adversarial Networks (cGANs) in their approach for generating realistic images from text descriptions[27].

2.6.2 Vulnerability of Text-to-Image Generation Algorithms to Adversarial Attacks

Text-guided image generation models can produce high-quality images of any subject, which is an exciting development for content creation. However, this technology also raises ethical questions about potential misuse. For example, these models may produce fake imagery of existing individuals for misinformation or offensive or harmful visual content. Moreover, one concern regarding the robustness and appropriate use of deep neural networks is the existence of adversarial examples designed to mislead them. Adversarial perturbations can be indistinguishable from the original input for humans in the visual domain but not typically in the linguistic field. Various alternative approaches for malicious attacks have been explored in natural language processing, causing models to misclassify text or generate biased or harmful text with minimal input perturbations. Adversarial examples have also been explored with vision-language models for image captioning and recognition. This method is potentially problematic, as it could be generalized to perform attacks on text-guided

image generation models to get around content moderation filters [19].

2.6.3 Image-Poisoning Attacks Based on Polysemantic Word

Poisoning Attack is a type of attack that exploits the assumption of the integrity of the training data by most machine learning algorithms. In this type of attack, the attacker is aware of the training algorithm and can manipulate the training samples. The primary objective of poisoning attacks is to degrade the accuracy of the learned model as much as possible. The attacks occur in the training phase, where the attacker manipulates the training data to include specially crafted data points. The ultimate goal of poisoning attacks is to degrade the classification accuracy of all the legitimate input samples. [15]

Data poisoning is a significant security concern for machine learning systems, intense learning systems that rely on vast amounts of data for training. Attackers can manipulate the training data, often scraped from various sources, to control the system's behavior, impacting the model's accuracy. Poisoned data can also insert backdoors or other vulnerabilities, giving attackers control of the system [25].

There are two types of image poisoning attacks:

1. **Triggerless attacks:** In techniques like Feature Collision (FC) and Convex Polytope (CP), poison samples are produced by introducing minor alterations to the base pictures to make their feature representations resemble those of the target image.

2. **Backdoor attacks:** like Class-Conditional Label-Based Poisoning (CLBD) and High-Transfer Backdoor Poisoning (HTBD), develop of poison examples with a concealed trigger, like a particular patch or pattern that causes the model to categorize the target image when the trigger is present incorrectly. [25]

Image-Poisoning Attacks Based on Polysemantic Word

Attacks Based on Polysemantic Words refer to a type of cyber attack in which a word with multiple meanings, known as a polysemantic word, is used to mislead or deceive the victim. In these attacks, the attacker takes advantage of the word's ambiguity and multiple meanings to create confusion, trick the victim into performing a specific action, or gain unauthorized access to their information or system.

Attacks Based on Polysemantic words can be a problem when dealing with text classification, especially with traditional text embedding models that represent each word as a fixed vector regardless of the context. In such cases, a word with multiple meanings can be interpreted incorrectly, leading to misleading or deceptive results [12].

2.6.4 Image Attacks Based on Cultural Differences

Cross-cultural meaning refers to the idea that different cultures interpret visual media differently. Visual media can include images, symbols, colors, and other forms of visual communication. Understanding these differences in interpretation is essential for promoting cross-cultural understanding and effective communication between people

from different cultures.

There are several challenges associated with using visual media to promote cross-cultural understanding. One of the main challenges is that visual media may frequently not be interpreted in the way it was intended across different cultures. Cultural differences can affect how visual media is perceived and understood. For example, specific colors and symbols may have different meanings in different cultures, leading to misinterpretation of visual media. Another challenge is that individual interpretation can influence how visual media is understood, regardless of cultural background. Additionally, relying solely on visual media can be limiting, as it may not provide enough context or depth to convey a message fully [7].

Symbols are widely used in various contexts, such as signage, transportation, and public spaces. However, the effectiveness of these symbols can be limited if they do not consider cultural differences in perception and interpretation. The design of pictographic symbols has been analyzed in an international airport and it has been found that some symbols, such as the one for toilets, were only universally understood by some users. For example, some users from Asia needed clarification on the symbol for toilets, as it depicted a Western-style toilet, which is not commonly used in some parts of Asia. Designers need to consider cultural factors when designing pictographic symbols, such as the cultural familiarity of the symbols, the appropriateness of the symbols for the intended users, and the cultural significance of colors and shapes used in the design. Cultural constraints should be considered in the design of pictographic

symbols to ensure that they are practical and understandable to all users, regardless of their cultural background [13].

Chapter 3

Methodology

3.1 Research design

This study aims to evaluate the impact of image-poisoning attacks and cultural variations on the understanding of news or narratives when converting text to visuals using a mixed-methods research methodology. The selected mixed-methods technique combines quantitative and qualitative data to explain the study topic thoroughly. The research will consist of the following:

1. **Quantitative Component:** The quantitative component of the study will primarily focus on data analysis from an online survey. The survey will collect information about participants' interpretations of news articles represented by icons, their knowledge of the news as a result of the icons, and their thoughts on the usage of icons in news stories. Statistical analysis will identify patterns

and connections between variables such as cultural background and expertise with icons and images.

2. **Qualitative Component:** A review of relevant literature and secondary data sources will be conducted as part of the qualitative component of the research. This will assist in corroborating the conclusions produced from the primary data acquired via the online survey by providing context and background information for the study. The qualitative data will also help better understand the possible impact of image-poisoning attacks and cultural differences on how people receive news or narratives.

3. **AI-Assisted Component:** To produce icons from text, the AI-assisted component of the research will employ artificial intelligence technologies such as DALLE-2, Midjourney, and the Text2icon source code from GitHub. These AI-generated icons will be displayed to survey participants to collect responses and assess participants' comprehension and perception of the news items represented by the icons.

The mixed-methods approach allow for a more in-depth investigation of the study topics, resulting in a complete knowledge of the impact of image-poisoning attacks and cultural variations on text-to-image conversion, especially in the context of news reports.

3.2 Data collection

We used the following approaches to gather data for this study:

1. **Online Survey:** An online survey was created and disseminated to various participants. The survey contains questions about the cultural background and familiarity with icons and imagery. It also features a series of questions about the perception of news articles represented by icons created using AI technologies in the Text2icon code from GitHub. The survey will be given in two versions: one showing the sequence of icons designed by the Text2icon program without any attack and the other showing the icons after the attacker has modified them in an attempt to attack the audience.

   The purpose of using two survey versions is to see how participants' perceptions alter and to highlight the impact of image-poisoning attacks. Questions testing participants' comprehension and attitudes toward using icons in news articles were also included. To guarantee simple access and distribution to a wide range of participants, the survey was performed utilizing an online platform, Google Forms.

2. **Secondary Data:** This study evaluated relevant literature and secondary data sources in addition to the primary data collected through the survey. Academic publications, news articles, and reports on image-poisoning incidents, cultural variances in icon interpretation, and icons in news stories are included.

Data gathering follows ethical principles to protect participants' privacy and confidentiality. Participants were told about the study's goal, the voluntary nature of their participation, and the procedures taken to protect their anonymity.

3.3 Data analysis

To obtain a thorough grasp of the research issues, the data for this study was analyzed using quantitative and qualitative methodologies. The data gathered from the online survey and the literature review are examined as follows:

1. **Quantitative Data Analysis:** The survey data will be analyzed using two types of statistics: descriptive and inferential. By providing means, frequencies, and percentages, descriptive statistics assist us in summarizing the data. This helps us comprehend the data better. Inferential statistics are also used to study correlations between variables and test hypotheses.

2. **Qualitative Data Analysis:** Qualitative method are used to evaluate the qualitative data gathered from the literature review and secondary data sources. This specifically looks for patterns and themes in the data that connect to the study questions. The analysis process was part of reading over the data, categorizing the data, and detecting emerging trends.

3. **Assessing Participants' Perceived and Actual Understanding:** In addition to the quantitative and qualitative analysis, we assessed the difference

between participants' perceived and actual understanding of the news articles represented by icons. Participants were given four alternatives for indicating how well they thought they comprehended the content: "very well," "somewhat," "slightly," or "not at all." Using the provided icons, they could rate their understanding of the news article.

In addition to rating their understanding, participants were asked to describe their interpretation of the news article based on the icons. These descriptions helped to gain deeper insights into how participants comprehended and interpreted the content. Then, these descriptions were compared with the actual content of the news article to determine the accuracy of participants' understanding.

By analyzing participants' self-reported understanding and descriptions, we aimed to uncover any differences or variations between what participants believed they understood and their actual comprehension of the news article. This analysis provided valuable insights into the effectiveness of the icons in conveying the intended message and whether participants' interpretations aligned with the intended content.

The combination of quantitative and qualitative data analysis provides a more robust understanding of the research problem, allowing for a more in-depth investigation of the impact of image-poisoning attacks and cultural differences on interpreting news or narratives when text is converted to images. The data analysis findings are

used to generate recommendations and approaches for enhancing the usage of icons in news articles and limiting the dangers associated with image-poisoning attacks and cultural differences.

Chapter 4

Results and discussion

4.1 Analysis of image-poisoning attacks based on polysemantic words

An attacker can manipulate the actions of AI algorithms that aid in converting text to images or icons by employing poisoning techniques. For instance, they may include a "backdoor" that would enable them to activate particular, unwanted behaviors when certain trigger words or phrases appear in the input text. Changing the model's architecture or weight parameters can also affect AI output negatively.

For example, if the attacker wants the target image to be identified as the 'sun' but it is a picture of the moon, they could alter the training data by adding a little pattern (like a sticker or a particular arrangement of pixels) to the training set's images of the sun. After training with the tainted data set, the model wrongly identifies the

target image as the sun.

Such an attack is risky because it can result in inaccurate predictions, and detecting whether a model has been attacked is not easy. This could lead to severe security and privacy issues, such as the spread of false information and the theft of private data.

Attacks on text-to-icon or image-generation models may produce visually identical but semantically inaccurate results, particularly in news contexts where disseminating false or misleading information may have far-reaching effects.

Example: Let us consider a news report on a new digital currency shown as a golden coin. *"Breaking News: A new digital currency crushes Competitors!"* In this case, an attacker could poison the picture database with a golden coin's image made to appear damaged or otherwise broken. That may be in contrast to a reality where this digital currency is secure and trustworthy. This could give the impression that the coin is unreliable or prone to failure.

Figure 4.1: Image generated by DALLE 2 with the prompt: *"Breaking News: A new digital currency crushes Competitors!"*

Consequently, investing in a new digital currency based on inaccurate information can negatively affect its reputation and market value.

This study conducted a survey to see if participants could be harmed by image-poisoning attacks using polysemantic words. Using two alternative sets of icons, we examined the participants' knowledge and perception of this news:

Breaking News: The price of Bitcoin crashed after Russian President Vladimir Putin announced a military operation in the Donbas region of Ukraine.

The first set of icons represented the original, unchanged sequence, while the second comprised an image-poisoning attack that transformed the down arrow icon into a car-crushed icon science converting a word from text to icon, the tool may produce any icon representing the concept of "crush," regardless of the intended meaning.

The majority of participants could conclude from the original icons that the news report was about the decline in Bitcoin value due to Russia's invasion of Ukraine. Several participants also discussed the war's influence on the economy, the stock market, and cryptocurrency in general.

However, the participant's perceptions of the news event got more diversified in the attacked version of the question, which included unrelated issues such as car accidents, disasters, or terrible news in general. Several participants correctly deduced the news narrative, but others were confused by the car crushed icon, which may represent a

Figure 4.2: Image generated by Text2Icon with the prompt:*Breaking News: The price of Bitcoin crashed after Russian President Vladimir Putin announced a military operation in the Donbas region of Ukraine.*

car accident rather than a cryptocurrency collapse.

Figure 4.3: Image generated by Text2Icon with the prompt:*Breaking News: The price of Bitcoin crashed after Russian President Vladimir Putin announced a military operation in the Donbas region of Ukraine.*

The following results were obtained when comparing the interpretation of the news article based on the icons in the two questions:

The comparison demonstrates that the image-poisoning attack had a detrimental impact on the participant's interpretation of the news story. The proportion of participants who comprehended the news article "very well" or "totally" fell from 40.9% in the original question to 22.2% in the attack one. On the attack question, however, the number of participants who comprehended the news article "not at all" or "slightly" jumped from 22.7% in the first question to 40.7%.

The results of the two survey questions show that even little adjustments in the

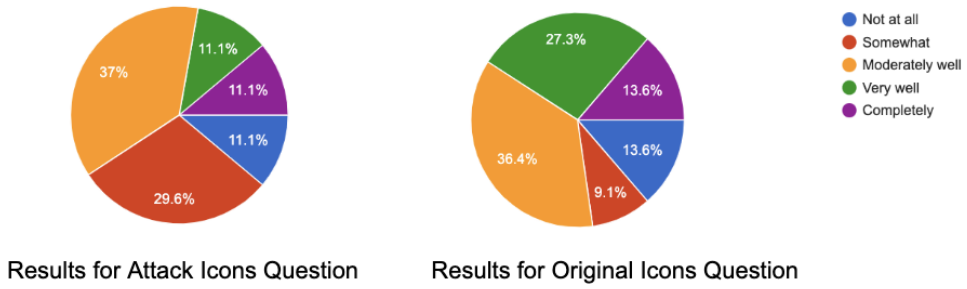

Figure 4.4: Comparison of Interpretation Based on Original Icons vs. Attack Icons for the Example "Breaking News: The price of Bitcoin crashed after Russian President Vladimir Putin announced a military operation in the Donbas region of Ukraine" in Participant Responses

selection or orientation of icons can significantly impact the interpretation and understanding of the information being delivered. The image-poisoning attack that altered the down arrow icon to a car-crushed icon caused participants to become confused and misinterpret the intended news content.

Second example: *Breaking News: Scientists Make Breakthrough in Cancer Research with New Treatment*, where the breakthrough is represented by an image of a rocket launching into space.

In this case, an attacker could modify the image of the rocket to give the impression that it is crushing rather than launching. This could suggest that the breakthrough in cancer research failed when it was a success.

This survey question assessed the impact of image-poisoning attacks in the context of news headlines on cancer research achievements. Two questions were provided, with a similar sequence of news icons and different in just one icon where a downward

rocket icon represented the attack question. In contrast, the original question was represented by an upward rocket icon.

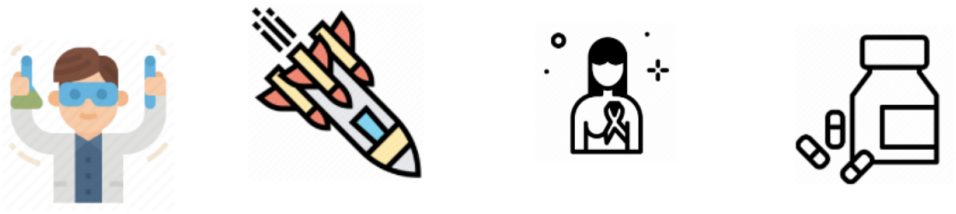

Figure 4.5: Image generated by Text2Icon with the prompt:*Breaking News: Scientists Make Breakthrough in Cancer Research with New Treatment*

Attacked question (downward rocket icon): This question produced a wide range of responses, with some participants inferring the sequence was connected to cancer research or breast cancer therapy, while others suggested unrelated interpretations. The downward rocket icon confused and misinterpreted the intended news content.

Figure 4.6: Image generated by Text2Icon with the prompt:*Breaking News: Scientists Make Breakthrough in Cancer Research with New Treatment*

Original question transformation (upward rocket icon): Several participants correctly deduced that the sequence was associated with a breakthrough in cancer research, notably breast cancer therapy. The upward rocket icon assisted participants in noticing the promise of developments and advances in cancer research.

When we compare the two groups of responses, we can observe that while the rocket indicator pointed upwards, the news narrative's knowledge increased. The percentage of participants who "very well" or "totally" comprehended the news article climbed from 18.5% in the first question to 45.5% in the second. In contrast, the percentage of participants who understood the news article "not at all" or "slightly" fell from 66.6% in the first question to 45.5% in the second.

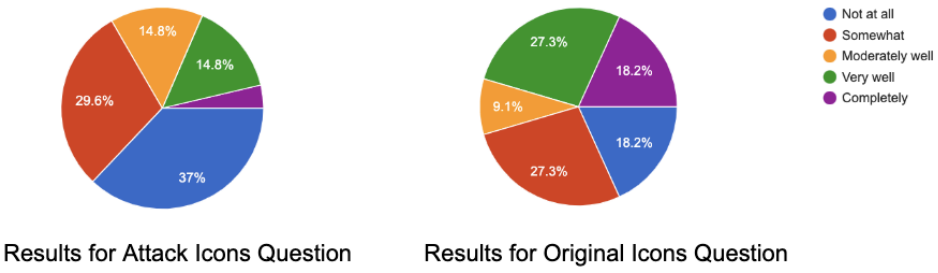

Figure 4.7: Comparison of Interpretation Based on Original Icons vs. Attack Icons for the Example "Breaking News: Scientists Make Breakthrough in Cancer Research with New Treatment" in Participant Responses

The rocket icon's direction significantly impacted how the news article was seen and understood. Participants were more likely to evaluate the news report as a good step in cancer research when the rocket icon pointed upwards.

Slight adjustments in the choice or direction of icons can significantly impact the

perception and comprehension of the information being given. This may be used in image-poisoning attacks, in which an attacker alters the meaning of a single icon to mislead the audience possibly.

The effects of image-poisoning attacks in this scenario might be severe. For example, suppose someone relied on an incorrect description in a news article to decide whether to invest in a cancer research business. In that case, they may resist, assuming the study has failed. This might result in lost funding opportunities and have a detrimental influence on medical research advancement.

4.2 Analysis of image attacks based on cultural differences

Image attacks based on cultural differences can happen when the meaning of an image or icon is culturally specific and may not be universally understood. For example, an image representing a gesture or symbol may have a different meaning or connotation in a different cultural context.

Attackers could take advantage of this attack by intentionally using culturally inappropriate, offensive, or misleading images to spread misinformation or create confusion among the target audience.

Let us consider this example that has been converted using Text2icon "*Millions of Muslims travel to the Kaaba in Mecca every year to perform the Hajj pilgrimage.*"

Figure 4.8: Image generated by Text2Icon with the prompt:*Millions of Muslims travel to the Kaaba in Mecca every year to perform the Hajj pilgrimage.*

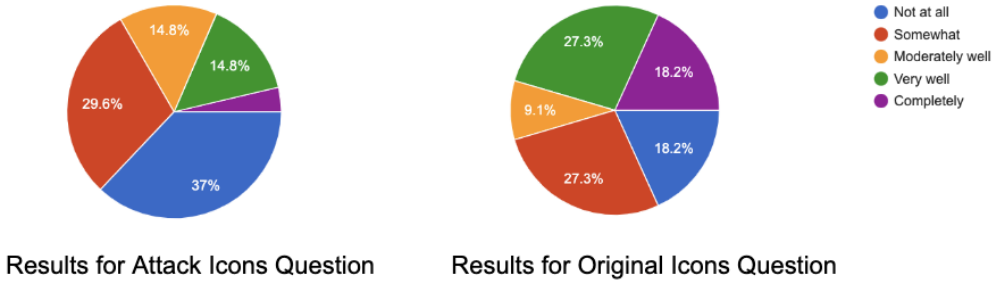

Figure 4.9: Comparison of Interpretation Based on Original Icons vs. Attack Icons for the Example "Millions of Muslims travel to the Kaaba in Mecca every year to perform the Hajj pilgrimage." in Participant Responses

Based on the participant's responses, the interpretation of the news article based on the icons appears to differ between the two groups. The first group had a more significant number of participants who understood the news article moderately or significantly well, with a lower percentage who did not recognize the story at all or just slightly understood it.

Given that the majority of participants were Muslim, it is possible that their knowledge of the cultural and religious importance of the icons helped their interpretation of the news article. Participants who were not from the Muslim community or unfamiliar with the tradition may have had difficulties recognizing the icons and

understanding the news story since the icons were tied to the Hajj trip, a fundamental religious rite in Islam.

As a result, the survey findings show that cultural and religious familiarity could play an essential role in interpreting icons and their related news events.

The perception of images, icons, or symbolically powerful behaviors is inextricably linked to a person's cultural knowledge and history. This reality is strongly depicted in Islamic rituals such as Tawaf and Omra. Participants in the Tawaf circumnavigate the Kaaba in a counterclockwise orientation. This gesture, which is steeped in deep religious symbolism within Islam, exemplifies the critical significance that cultural and religious settings have in interpreting such acts. As stated by researchers, "During Tawaf, each pilgrim walks counterclockwise, seven times in circular movements around Kaab'a, which is situated in the Mataf area" (CFD Based Numerical Study of Pilgrims Movement around Kaaba). Understanding the cultural and religious backdrop that drives these religious practices becomes critical when they are depicted symbolically or in iconic forms. This emphasizes the importance of cultural knowledge in comprehending visual and symbolic representations and how our cultural viewpoints molds and filter this understanding. Importantly, this principle applies to many actions and occurrences, not just religious ceremonies. The process of interpreting visual or symbolic representations is intricately tied to cultural knowledge, and it serves as the foundation for our perception and comprehension of the world.

As a result, our cultural lens substantially impacts how we interpret and grasp diverse forms of visual communication, emphasizing the importance of cultural context in effective communication and understanding.

An Example: *New Study Finds Clockwise Running May Boost Brain Function*

Figure 4.10: Image generated by Text2Icon with the prompt:*New Study Finds Clockwise Running May Boost Brain Function*

As participants attempted to understand this sequence of icons, most interpretations focused on the significance of exercise, particularly morning exercise, for mental and physical wellness. Nevertheless, the actual news report was about clockwise running potentially improving brain function, something the icons did not directly express. The respondent's interpretation of the news article based on the icons varied, with 40% knowing it somewhat, 18.75% understanding it very well, and 25.6 comprehending it slightly. 11.95%, on the other hand, had no idea what the news article was about.

Image attacks based on cultural differences can happen when the meaning of an image or icon is different culturally and may not be universally recognized. Attackers might utilize culturally inappropriate, insulting, or misleading images to propagate misinformation or confuse the intended audience. In the given example, the clockwise

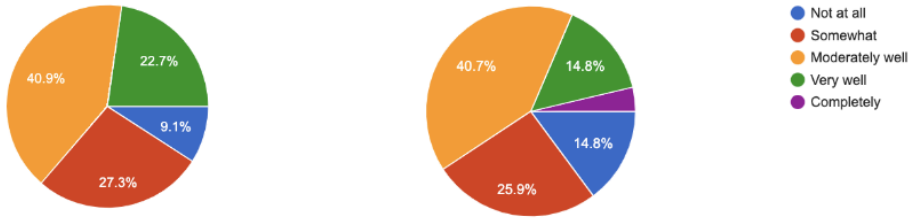

Figure 4.11: Participant Interpretation Based on Icons for the Example "New Study Finds Clock-wise Running May Boost Brain Function"

symbol has different meanings in various cultures, such as prosperity and good luck in Hinduism or the Buddha's footsteps in Buddhism. This highlights the importance of considering cultural differences when developing icons or symbols for communication, as noted in [21].

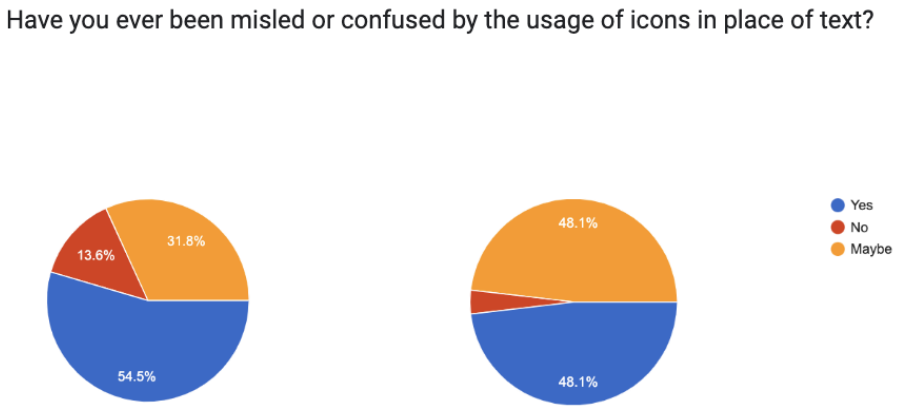

Figure 4.12: Participants' response about, " Have you ever been misled or confused by the usage of icons in place of text?

Moreover, icons may be culturally recognizable and not understood by various audiences, with 30.7% percent of respondents seeing this as a disadvantage of using icons to communicate information (Figure 4.12). In addition, 51.3% percent of respondents

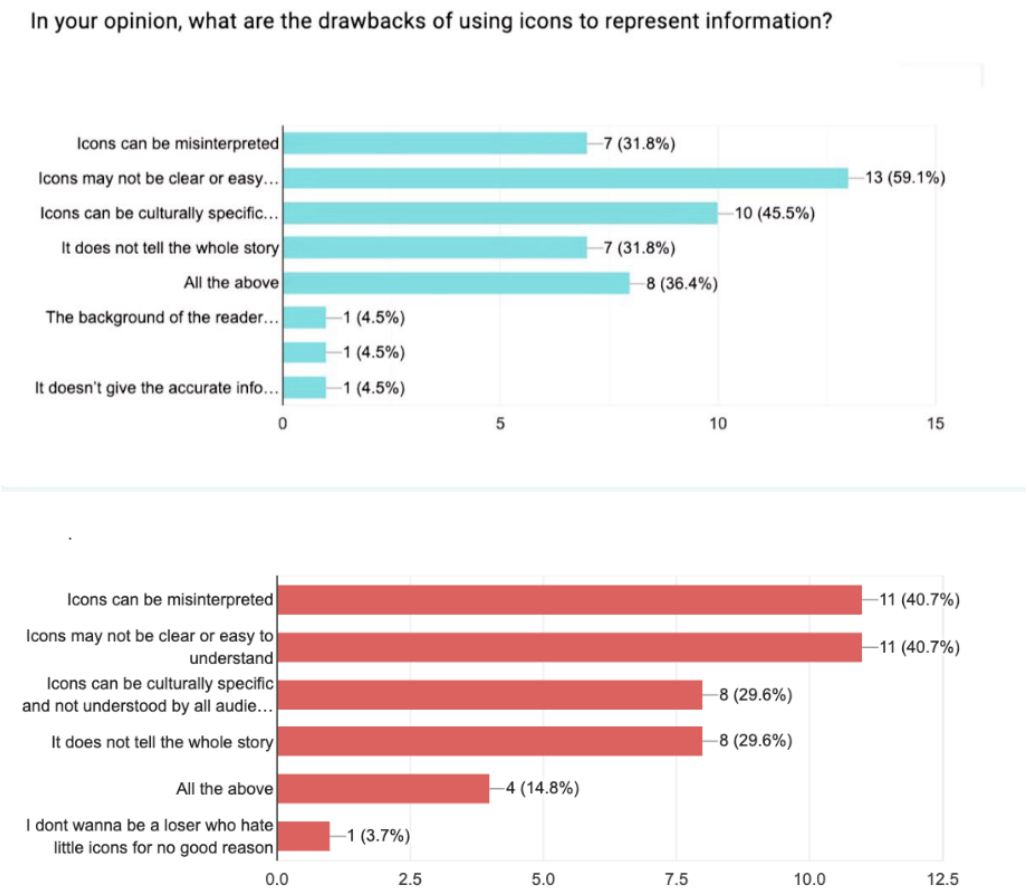

Figure 4.13: Participants' response: "In your opinion, what are the drawbacks of using icons to represent information?"

needed clarification by using icons instead of words (Figure 4.12), and 35.25% percent failed to comprehend the importance of a hero or image (Figure 4.13). These figures suggest that cultural variations may lead to misreading icons and images, raising the danger of image attacks. One responder also gave an example of an icon (walking counterclockwise during Hajj and Umrah rituals) that meant something different to them because of their cultural background.

The findings highlight the importance of a more inclusive and culturally aware

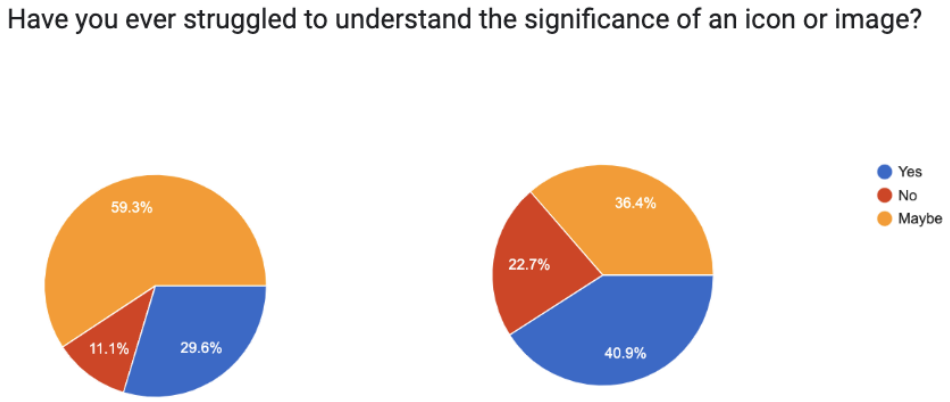

Figure 4.14: Participants' response: —"Have you ever struggled to understand the significance of an icon or image?"

approach to visual communication to reduce the risks associated with image attacks based on cultural differences.

To limit the possibility of image attacks based on cultural differences, it is necessary to utilize globally known icons, give context or explanations for the icons used, and design and choose icons with the target audience's cultural background in mind.

4.3 Impact of these attacks on news and stories

Based on the data and analysis provided, the impact of image-poisoning attacks based on polysemantic words and image attacks based on cultural differences in news and stories can be summarized as follows:

1. Misinterpretation and confusion:

   It can happen when utilizing iconography with different meanings or when im-

pacted by cultural differences, leading to incorrect assumptions and spreading disinformation. In the cancer study scenario, 66.6% of the participants struggled to follow the content owing to confusing icons of the downward rocket icon.

Moreover, in the example:

Breaking News: An employee at a service station was threatened by a hooded individual with a shotgun on a Friday morning. The suspect then robbed the station and fled on foot with the money. The police are now searching for the suspect. Responses demonstrate that respondents suggested various meanings for the icons, many of which were inaccurate representations of the real news event. Here are a few examples:

(a) "Police arrived at two people fighting."

(b) picture shows professions, days of the week, and the daily routine of life."

(c) "demonstrations."

(d) "Bank theft scheduled on Friday."

(e) "Day by day, the crimes increase."

This shows that image-based representations might frequently cause misunderstanding and disinformation, further affecting the audience's understanding of the news story.

2. Affecting public opinion and decision-making:

Figure 4.15: Image generated by Text2Icon with the prompt: *Breaking News: An employee at a service station was threatened by a hooded individual with a shotgun on a Friday morning. The suspect then robbed the station and fled on foot with the money. The police are now searching for the suspect.*

Because of differing degrees of knowledge and the possibility of misunderstanding or confusion, image-based attacks can impact public opinion and decision-making. This may lead to incorrect judgments or views that depend on insufficient knowledge.

In the case of cancer research news (Figure 4.5 & Figure 4.6), a tiny modification in the direction of the rocket emblem had a significant influence on interpretation and understanding, which might have consequences for medical and scientific studies. Misinterpretations may result in missed investment opportunities if people judge based on incorrect descriptions.

Similarly, in the Bitcoin news example (Figure 4.2 & Figure 4.3), the car-crushed icon confused, perhaps leading to the propagation of incorrect information. Employing icons instead of text might make it simpler for malicious actors to spread false information or scams, potentially manipulating the audience to make poor investment decisions.

3. Loss of critical information: Using icons or images in news reports can lead

to critical information being lost or misconstrued, affecting the audience's perception of events and limiting their capacity to make well-informed judgments or hold correct perspectives. According to the data, essential components of a story need to be consistently and adequately transmitted through visual representations. This ambiguity leads to many interpretations and can contribute to audience confusion and misinformation.

In Figure 4.11, the fact that the robbery occurred on a Friday morning was not consistently recognized, with some responses naming various days of the week or not specifying a specific day at all. The fact that the suspect escaped on foot with the money while police were looking for them was also misunderstood. Some responses claimed the suspect had been caught or arrested, while others stated that the suspect had escaped without mentioning the ongoing police search.

4. Exploitation by attackers:

   Attackers could use ambiguous or culturally influenced icons to mislead audiences, affect public opinion, or generate confusion. This undermines faith in news sources and increases the spread of incorrect information. With 51.3% (Figure 4.12) of respondents admitting to being misled or confused by icons instead of text, the possibility for exploitation and the significance of clear communication is emphasized.

5. Exploiting cultural differences:

The data in the examples show how image-based attacks can take advantage of cultural variations and unfamiliarity with specific icons or images, leading to misinterpretations based on cultural variations. This confuses the audience's comprehension of the news story and raises the possibility of bad outcomes or offenses.

Participants in Figure 4.8 were presented with icons associated with the Hajj pilgrimage, and their perceptions of the news piece based on these images differed. Muslim participants had a better grasp since they were aware of the cultural and religious importance of the icons. Non-Muslims or people unfamiliar with the tradition, on the other hand, may need help recognizing the icons and comprehending the news.

Moreover, participants presented a sequence of icons related to clockwise running and brain function, as illustrated in Figure 4.10. Most interpretations focused on the importance of exercise for mental and physical well-being. However, the real news item was about clockwise running potentially boosting brain function, which the icons did not indicate. This highlights the possibility for image-based assaults to exploit cultural nuances.

The study also showed that 30.7% (Figure 4.13:) of respondents saw cultural differences as a disadvantage when using icons for communication, and 51.3% (Figure 4.12) felt misled or confused by icons replacing words. This implies that cultural differences lead to misunderstanding icons and pictures, raising the

danger of image attacks. One respondent even offered an example of an icon (walking counterclockwise during Hajj and Umrah rites) that had a distinct significance for them based on their cultural background. This confirms the hypothesis that image-based attacks might take advantage of cultural differences and unfamiliarity with specific iconography, resulting in misconceptions and potentially unfavorable effects.

4.4 Strategies to prevent image tampering and the spread of false information

Here a compilation of strategies for reducing the risks of image manipulation and misleading spread when converting text into icons and images is provided based on surveyed literature:

1. Design icons that is unambiguous:

   To avoid misinterpretations, create clear, unambiguous, and generally recognized icons. According to research, a well-designed icon may improve communication and decrease misunderstandings. [18]

2. Contextual information:

   When using icons and images, provide text context or explanations, mainly when working with culturally distinct or possibly misleading icons. Research

suggests that including contextual information can considerably increase visual representation understanding. [16].

3. Culturally sensitive design:

   It is important to consider culturally sensitive design when converting text to icons and images to prevent image manipulation and the transmission of incorrect information. Nielsen (1990) highlights the significance of accounting for cultural variations when developing user interfaces, particularly the usage of icons and visual representations. By recognizing and incorporating cultural subtleties into the design process, designers may develop more effective and universally understood visual representations, lowering the possibility of misinterpretation or confusion and possibly leading to the spread of erroneous information. [20]

4. Digital watermarking: To confirm the validity and integrity of the images, including digital watermarks or signatures within them. Watermarking methods can help track down the origins of manipulated images and prevent the spread of fraudulent information. [3]

5. Mechanisms for robust verification:

   Use verification techniques such as blockchain-based systems to authenticate the validity of images and icons. Blockchain technology can verify digital assets' provenance and integrity, making it harder for attackers to tamper with or manipulate them. [2]

6. Combining Text and Visuals:

Converting text into visual components, including text with images and icons, helps reduce the dangers of image manipulation and false information. The mix of text and graphics offers a more accurate and comprehensive representation of information accessible to varied audiences by giving context, improving understanding, decreasing ambiguity, resolving cultural differences, and enabling more straightforward authenticity verification. This claim is supported by a study conducted by Schnotz and Bannert (2003), which found that learners who were given textual and visual explanations of complex scientific processes demonstrated better comprehension and retention of the information than those who only received visual answers. [24]

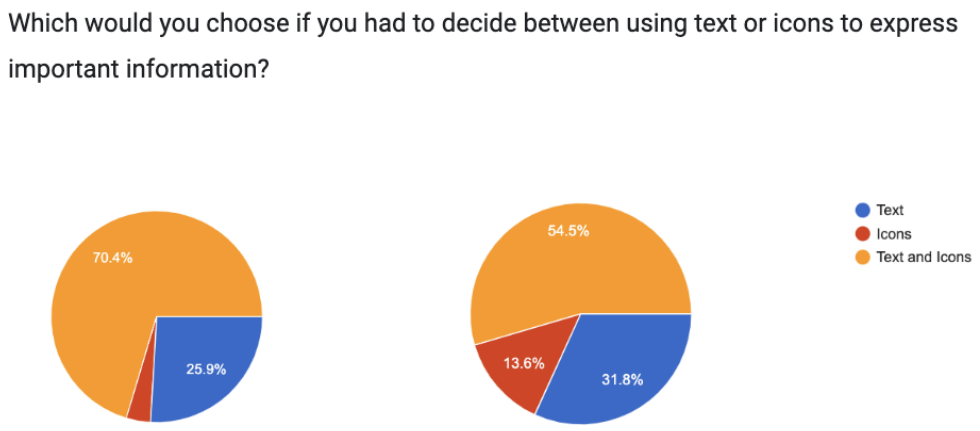

Figure 4.16: Survey Result

When asked if respondents preferred text or icons to represent crucial information, 28.85% selected text, 8.65% selected icons, and the majority (65.5%)

preferred a combination of the two. This preference shows that combining visual and textual information may be more exciting and understandable than depending exclusively on text or icons.

Respondents were also aware of the difficulties and potential misconceptions that might result from using icons in news items. Their recommendations emphasize the importance of clear communication and appealing to a wide range of consumers. They think that mixing text with iconography, employing culturally neutral or generally recognized symbols, and offering more information may increase their understanding of news items. These recommendations also show an interest in improving icon-based communication, which may include developing resources such as a dictionary or database of news icons and soliciting frequent input from readers to guarantee continual progress.

Chapter 5

Conclusion and recommendations

5.1 Summary of findings

This study's findings suggest that image-poisoning attacks based on polysemantic words and image attacks based on cultural differences significantly influence how people read and interpret news events. These attacks can cause confusion, misunderstanding, limited accessibility, attacker exploitation, and faulty decision-making.

Unambiguous icon design, offering contextual information, considering culturally sensitive design, applying digital watermarking, and rigorous verification procedures are all strategies to avoid image tampering and disseminating misleading information when converting text to icons and pictures.

5.2 Conclusion

AI tools like Text2Icons, converting text into icons/images, present a promising approach to making complex information more attractive [28]. However, some restrictions and difficulties come with this. One significant issue is these algorithms are vulnerable to adversarial attacks, including image-poisoning attacks based on polysemantic words and image attacks based on cultural differences.

When translated into images or icons, image-poisoning attacks manipulate images or subtly alter words to create multiple meanings or interpretations.

For example, depending on the situation, an attacker might replace another word for "bank", switching between "bank of a river" or "bank account", which has a different meaning when used in an icon or image. The AI tool could fail to distinguish between the senses and might create a picture or icon that suggests a purpose that was not intended, causing misunderstanding or disinformation.

On the other hand, image attacks based on cultural differences use cultural nuances and variations in how things are seen. Attackers can exploit this vulnerability to alter the meaning of images to communicate a different message. Additionally, an attacker could use images or icons that are unfamiliar or not commonly used by the intended audience.

For example, a positive or neutral icon in one culture may have a harmful or offensive meaning in another. Additionally, an attacker could use images or icons that are unfamiliar or not commonly used by the intended audience. , which could

cause confusion and misinterpretation.

Our survey results confirm the possibility of cultural misunderstanding in icon interpretation. We discovered substantial disparities in perception based on the participants' cultural and religious backgrounds when we used icons to represent a Hajj trip, a key religious ceremony in Islam. Most Muslim participants familiar with the icons' cultural and theological significance understood the story moderately or considerably well. On the other hand, non-Muslims and individuals unfamiliar with Islamic traditions had difficulty deciphering the icons, with a higher percentage scarcely identifying or understanding the story.

Another example was the icon for a news article headlined "New Study Discovers Clockwise Running May Improve Brain Function." In this case, 40% of participants comprehended the news piece somewhat, 18.75% understood it very well, and 25.6% understood it only somewhat. Most notably, 11.95% of participants were completely uninformed of the news article's substance. Many misread the icon sequence, which represented the benefits of clockwise running for brain function as broad support for exercise, demonstrating the potential for ambiguity in icon usage.

These examples demonstrate how cultural variations and the nuances of icon interpretation can be utilized in detrimental ways, such as using icons to spread misinformation or misunderstanding. They emphasize the importance of building culturally sensitive procedures for icon production to promote clear, courteous, and accurate communication across varied cultural landscapes.

Hence, image-based attacks, whether due to polysemic words or cultural differences, can have significant consequences such as the loss of critical information, confusion, and affecting public opinion, decision-making, and trust.

Given these problems, incorporating verification procedures to assure the legitimacy and accuracy of AI technologies like Text2Icons is critical. To overcome this challenge, as the next step of this project, we plan to create a mechanism to generate icons from the text while considering cultural variations. Our suggested approach will consider the user's cultural background and the text. Both probabilistic and neural network-based methods will be used to accomplish our goal. The probabilistic approach will be used to determine the probability of an icon being associated with a particular word or phrase, considering the user's cultural context. On the other hand, the neural network-based approach will be used to learn the patterns and relationships between the text, icons, and user types. We will train our system using a unique database organized by the user's cultural origin, for example, roots. A wide variety of cultural backgrounds, including race, nationality, and language, will be represented in this database. We can accurately categorize the cultural backgrounds of users using this database, and then we can use this knowledge to develop icons relevant to users' cultures. Overall, especially in multicultural settings, this work has the potential to enhance the way information is dramatically communicated through icons and images. We can ensure that icons and images are more accurate and culturally suitable by considering cultural variations, which lowers the possibility of misunderstanding and

confusion. This project has broader implications for creating inclusive and culturally responsive intelligent systems, overcoming cultural divides, and promoting improved understanding and communication between people from various backgrounds.

Understanding the general public's vulnerabilities and victimization patterns concerning text-to-image conversions is crucial for developing effective preventive measures and mitigating the risks associated with this technology. This study investigated the characteristics of cybercrime victims, concentrating on age, gender, security behaviors, and platform used in cyberattacks.

1. **Age and Vulnerability:** Younger people are more vulnerable to cyber-attacks. According to the Crime Survey for England and Wales, younger age groups were more likely to experience computer infections, with 37% of 16-24-year-olds and 33% of 25-34-year-olds reporting incidences compared to 25% of 55-64-year-olds. Furthermore, the oldest age group (over 75) had lower adoption rates of up-to-date security software and were less likely to adopt secure online activities.

2. **Gender Disparities:** Gender also influences cybercrime victimization. According to the Crime Survey for England and Wales, men were statistically substantially more likely (35%) than women (27%). Men were also more likely than women to report unlawful access to or use of their personal information. However, it is critical to remember that cybercrime affects people of all genders and should not be limited to one.

3. **Security Behaviors:** Developing effective security practices is critical for defending against cyber threats. While the use of antiviral software looks to be reasonably high, it is critical to assess the effectiveness of these measures. Users' comprehension and antivirus software settings may not always correspond to their actual degree of protection. Furthermore, broad security measures such as using credit cards instead of debit cards and not revealing personal information online are not consistently practiced, making individuals more exposed to cyber-attacks.

4. **Demographic Disparities:** Certain demographic groups are more vulnerable to cybercrime. Younger users, as well as those from black and minority ethnic (BME) origins and those from less socioeconomic backgrounds, are less likely to utilize internet security software. This emphasizes the importance of focused efforts to educate and empower these populations to adopt secure online practices.

5. **Platform Use and Emerging Risks:** As mobile devices have become more popular, vulnerabilities have spread beyond desktop and laptop PCs. Antivirus software for mobile phones and smartphones is relatively unknown and underutilized, particularly among Android users. Furthermore, using public Wi-Fi connections entails hazards, as a higher proportion of users who rely on public Wi-Fi have experienced security breaches than those who utilize home connections.

The characteristics of general public cybercrime victims reveal distinct vulnerabilities and discrepancies that deserve addressing. Cyber-attacks are more likely in younger people, men, and particular demographic groupings. Encouraging secure online habits, increasing understanding and adoption of suitable security measures, and addressing disparities in vulnerable groups are all critical steps toward improving cybersecurity resilience among the general public. Continued study and collaboration across the public, corporate, and educational sectors are required to establish a safer digital environment for everyone[6].

5.3 Recommendations for future research

Based on the study's findings, the following recommendations for future research are made:

Investigate the role of AI and machine learning techniques in identifying and preventing image-poisoning attacks. Further research can help create more advanced detection systems and ways to avoid picture tampering and modification [1].

Develop a culturally sensitive mechanism for generating icons from the text that takes into account the user's cultural background as well as the context of the text. It can grasp the links between text, icons, and user types using probabilistic and neural network-based approaches. The technique may develop more relevant and acceptable symbols to users' cultural backgrounds by training the system using a unique database

sorted by users' cultural origins.

This has the potential to improve communication in multicultural situations by providing correct and culturally relevant icons and visuals, hence minimizing misunderstandings and uncertainty. It also has more significant implications for developing inclusive and culturally sensitive intelligent systems that connect individuals from diverse backgrounds and encourage greater understanding and communication.

Finally, we will modify the survey design in the future to better capture the influence of demographics on the security of text-to-image conversions. We recognize the importance of gender, cultural background, religious affiliation, age groupings, and educational background in determining an individual's susceptibility to security assaults. The survey questions will be refined to include various demographic factors to ensure complete data collection. We will also include more questions that further probe participants' experiences and thoughts on the security of text-to-image conversion. We hope to gain more accurate and detailed insights into how demographics influence the vulnerabilities and dangers associated with text-to-image conversions by improving the survey instrument. This will allow us to build more tailored and effective tactics to address various demographic groups' unique security challenges

Appendix A

IRB

FLORIDA TECH

RESEARCH INVOLVING HUMAN PARTICIPANTS
EXPEDITED/FULL APPLICATION

This information listed below should be submitted to Florida Tech's IRB if the proposed research has more than minimal risk (none of the exempt conditions apply) or if the research utilizes a special population (children, prisoners, institutionalized individuals, etc.). Please consult the IRB website for detailed information, or contact the IRB Chairperson.
floridatech.edu/research/compliance--regulations/institutional-review-board

Submit via email to FIT_IRB@fit.edu.

IRB Contact Information:
Dr. Jignya Patel
IRB Chairperson
FIT_IRB@fit.edu
321-674-7391

PART 1: GENERAL INFORMATION

Title of project: Security of Text to Image Conversions

Date of submission: 2/24/2023

Expected project start date: 2/15/2023 Expected project duration: 6 months

Principal investigator: Zobaida Alssadi

Title: Master Student

Academic unit: Computer Engineering and science

Phone: 541-740-7687 Email: zalssadi2021@my.fit.edu

List all co-investigator(s). Please include name, title, academic unit/affiliation and email.

Marius Silaghi, Professor, CES, msilaghi@fit.edu

FLORIDA TECH

RESEARCH INVOLVING HUMAN PARTICIPANTS
EXPEDITED/FULL APPLICATION

PART 2: PROJECT SPONSORSHIP INFORMATION

If any part of this study will be funded by an external funding source (current or planned), you must note the funding source and award/solicitation number below:

PART 3: RESEARCH DESCRIPTION

1. In lay terms, please describe the GENERAL PURPOSE of the study and how human subjects will be involved. List the SPECIFIC AIMS and RESEARCH QUESTIONS or HYPOTHESES. Avoid the use of jargon when describing the purpose of the study.

-The general purpose of the study is to investigate the potential impacts of using solely icons and images instead of news and story contexts. The study will focus on the effects of restrictions and adversarial attacks, specifically image-poisoning attacks based on polysemantic words and image attacks based on cultural differences. The study aims to understand how these factors can lead to the loss of important information or details, which can result in misunderstandings and incorrect interpretations of the intended meaning.

The research questions of the study include:

How do different cultural backgrounds, educational levels, and individual experiences affect the interpretation of icons in news and story contexts, and thereby the sensitivity to attacks?
How effective are icons in conveying accurate and nuanced information in news and story contexts?

In the survey, the specific aims are to show participants a set of icons that represent a news story and ask them questions to test their understanding of the story as presented in icon form. The survey aims to gather feedback on the use of icons in presenting news stories and to better understand how different individuals interpret icons based on their backgrounds and experiences. The survey hopes to gather insights that can inform the use of icons in news and story contexts to improve the accuracy and effectiveness of conveying information to the general public, with support for qualitative and quantitative metrics.

2. Outline the INCLUSION CRITERIA for subjects, explaining the rationale for the involvement of any special groups, including children, prisoners, pregnant women or subjects with cognitive impairments. Describe the characteristics of the targeted subjects, including gender, age ranges, ethnic background and health/treatment status. If women or minorities are excluded, provide written justification. Give the number of subjects you anticipate including from each targeted group listed above.

Inclusion criteria for subjects include:
Age: 18 years and above
Ability to understand and give informed consent
The targeted subjects will be from different backgrounds, genders, and ethnicities

3. Describe sources for potential participants, how subjects will be RECRUITED or the sampling procedures. Attach recruitment advertisement(s), if applicable.

Potential participants will be sourced through online platforms such as social media, forums, and websites that are relevant to the topic of the study.

The sampling procedure will be random sampling.

4. Describe any COMPENSATION the subjects will receive, including course credit. If monetary compensation is offered, indicate how much the subjects will be paid and describe the terms of payment.

Participants will not receive any compensation

5. Explain how CONFIDENTIALITY and privacy of participant data (and anonymity if appropriate) will be maintained. If the research study involves collection of images or audio recordings of subjects, explain how the material will be used, who will see the images or hear the recordings and in what setting (refer to the audio/video recording policy).

The confidentiality and privacy of participant data will be maintained throughout the study in several ways:

All participant data will be kept confidential and only accessible to the research team.

Data will be stored on a secure server and only accessible to authorized personnel.

Data will be analyzed and reported in aggregate form to ensure individual anonymity.

Any images or audio recordings of subjects will be used solely for the purpose of the research study. These recordings will be kept confidential and only accessible to authorized personnel. The recordings will be stored in a secure location and will be deleted or de-identified after the completion of the study.

Participants will be informed about the use of images or audio recordings, and their consent will be obtained before the collection of any such data.

6. Describe the study design/research/measurement PROCEDURE (e.g., control and experimental groups, etc.). Indicate whether or not the subjects will be randomized for this study. Discuss how you will conduct your study and what measurement instruments you are using. List the specific steps of your research protocol. Explain scientific jargon. Attach a copy of any questionnaires, measurement instruments, interview protocols or a description of topics or an approximate script that will be used. If not available at this time, explain. Please describe your study in enough detail so the IRB can identify what you are doing and why.

The study design involves a survey of participants who will be shown a set of icons representing a news story converted by an AI tool (Text2Icon) and will be asked questions based on these icons to test their understanding of the news story as presented in icon form. The procedure does not involve control or experimental groups, and subjects will not be randomized. The survey will be conducted online, and measurement instruments include the icon set and the survey questions. Specific steps of the research protocol include obtaining informed consent, showing the icons to participants, asking questions about the news story, and collecting and analyzing the survey data. The survey will be anonymous, and confidentiality and privacy of participant data will be maintained. No audio or visual recordings will be collected. A copy of the survey questions and icon set will be provided to the IRB.

7. If the study will use deception, describe the nature of the deception, discuss why deception is necessary and fully indicate how participants will be debriefed. Deceptive techniques must be justified by the study's prospective scientific, educational or applied value, and the investigator should explore equally effective alternative procedures that do not use deception. A debrief form/process must be discussed here.

Deception is not necessary for this study as we do not want to mislead or hide the study's goals and steps from the participants. The information gathered will only be used for research and will be kept private and with no names attached.

8. Describe all SITES where this research will take place and attach documentation of permission from the appropriate source if the study involves subjects from places other than common public spaces.

The study will happen primarily online using tools like Google Forms or Qualtrics. People taking part will get a link to the survey or questionnaire from the research team.

The research team will also ensure that all participants are aware of the study's purpose and methods and have provided informed consent before participating.

9. Describe any POTENTIAL RISKS (physical, psychological, social, legal or other) and the steps that will be taken to minimize risk. Where appropriate, discuss provisions for ensuring necessary medical or professional intervention in the event of adverse effects to the subjects. Also, where appropriate, describe the provisions for monitoring the data collected to ensure the safety of subjects. Research involving children must carefully assess risks and describe the safeguards in place to minimize these risks.

There are no known risks to the participants in the experiment.

8. Describe all SITES where this research will take place and attach documentation of permission from the appropriate source if the study involves subjects from places other than common public spaces.

The study will happen primarily online using tools like Google Forms or Qualtrics. People taking part will get a link to the survey or questionnaire from the research team.

The research team will also ensure that all participants are aware of the study's purpose and methods and have provided informed consent before participating.

9. Describe any POTENTIAL RISKS (physical, psychological, social, legal or other) and the steps that will be taken to minimize risk. Where appropriate, discuss provisions for ensuring necessary medical or professional intervention in the event of adverse effects to the subjects. Also, where appropriate, describe the provisions for monitoring the data collected to ensure the safety of subjects. Research involving children must carefully assess risks and describe the safeguards in place to minimize these risks.

There are no known risks to the participants in the experiment.

10. Discuss the importance of the knowledge that will result from your study and what benefits will accrue to your subjects (if any). Discuss why the risks to subjects are reasonable in relation to the anticipated BENEFITS to subjects.

The knowledge that will result from this study is critical in developing strategies to ensure that information is conveyed through images and icons in a culturally appropriate and accurate manner, as well as to prevent image tampering and the spread of false information by hackers. By identifying the potential impact of image-poisoning attacks based on polysemantic words and of image attacks based on cultural differences, the study can provide guidance for the development of effective countermeasures that can protect the public from misinformation.

The benefits to subjects in this study are limited, as it is primarily focused on collecting data for research purposes. However, the study does provide an opportunity for participants to gain insight into the potential risks of image-based attacks and to contribute to the development of strategies to mitigate these risks.

The risks to subjects in this study are minimal, as the study involves only the completion of an online survey and does not involve any physical or psychological harm. The potential risks are related to privacy concerns and the possibility of data breaches, but steps will be taken to ensure the confidentiality and anonymity of participants. The benefits to the public, on the other hand, are significant, as the study's findings can help protect against the spread of false information and mitigate the impact of image-based attacks. Overall, the risks to subjects are reasonable in relation to the anticipated benefits to subjects and society.

11. CONSENT. Informed consent can be in either written or oral format. If you request waiver of informed consent, documentation of informed consent or of written informed consent, please state your justifications. Attach consent form if applicable. If an oral consent is planned, attach a copy of the text of the statement. If the study will be conducted with minors, provide an assent script. If assent is deemed unnecessary or inappropriate, you must discuss why. The consent form should contain all eight elements listed in Part 4. Researchers are strongly encouraged to use the formal headers found in Part 4, Item 3 to structure the consent document.

Participants will be asked to review and sign the attached Informed Consent Document.

PART 4: INSTRUCTIONS FOR DOCUMENTATION OF INFORMED CONSENT

Informed consent is one of the primary ethical requirements underlying human subjects research, reflecting the principle of respect for potential subjects. Informed consent assures that prospective human subjects understand the nature of the research and can decide knowledgeably and voluntarily whether or not to participate.

Informed consent refers to the voluntary choice of an individual to participate in research based on an accurate and complete understanding of, among other things, its purposes, procedures, risks, benefits, alternatives and any other factors that may affect a person's decision to participate.

The basic concepts of the consent process include:

- Full disclosure of the nature of the research and the subject's participation
- Adequate comprehension on the part of the potential subject
- Voluntary choice to participate

Informed consent must be documented by use of a written consent form approved by the IRB and signed by the participant or the participant's legally authorized representative. A copy should be given to the person signing the form. Even though the IRB has approved a consent procedure, it is the investigator's responsibility to ensure that each potential subject understands the information and to take the appropriate steps necessary to gain that comprehension.

Individuals may not be involved as research participants unless a) they understand the information that has been provided and informed consent has been obtained, or b) the IRB has approved a waiver for informed consent.

REMEMBER: If the participant is under the age of 18, parental consent is required. This includes college students under the age of 18.

If the research involves the participation of minors (under 18 years of age), read the description of requirements for research involving children. Additional requirements concerning parental consent forms and child assent are discussed.

Please follow the instructions for documentation carefully.

1. The consent form should be written in language that the participants can understand. Whenever possible, simple declarative sentences should be used. Ordinary language should explain technical terms.
2. Avoid the use of exculpatory language through which the subject or the representative is made to waive or appear to waive any of his/her legal rights or release the investigator, sponsor or institution or its agents from liability for negligence.
3. Important information that must be included on the consent form:
    a. Purpose of the research.
    b. Procedures to be followed (what will the participants be asked to do? Include physical requirements or experimental procedures if applicable.)
    c. Foreseeable risks or discomforts to the subjects. What are the risks associated with participating and what safeguards are in place? Include the following statement, where appropriate: "In the event of physical injury resulting from the research procedures, no form of compensation is available. Medical treatment may be provided at your expense or at the expense of your health care insurer (i.e., Medicare, Medicaid, private payer) which may or may not provide coverage. If you have questions it is your responsibility to contact your insurer."
    d. Benefits to the subject or others which may reasonably be expected to result.
    e. Alternative procedures or alternatives to participation, if any.
    f. Level of confidentiality of participant records. Is data anonymous? How will data be stored? If audio or visual records are obtained, how will they be maintained? Who will have access to the data?
    g. Primary investigator's contact information. Point of contact for questions or problems related to this study.

h. IRB contact. Also note the study was approved by Florida Institute of Technology's IRB, and list the current IRB chair and his/her contact information for questions about the rights of people who take part in research.

i. Voluntary participation, refusal and withdrawal. Include the following statement: "Participation is voluntary. Refusal to participate will involve no penalty or loss of benefits to which you are otherwise entitled. You may discontinue participation at any time without penalty or loss of benefits to which you are otherwise entitled."

j. Signatures, if appropriate. Provide a place for:

   I. Signature of the participant (or his/her legally authorized representative)

   II. Date of signature

WAIVER OF INFORMED CONSENT

The IRB may approve a consent procedure that does not include, or which alters, some or all of the elements of informed consent outlined above or waives the requirements to obtain informed consent, provided the IRB finds and documents that the following four conditions have been met:

- The research involves no more than minimal risk to the subjects;
- The waiver or alteration will not adversely affect the rights and welfare of subjects;
- The research could not practicably be carried out without the waiver or alteration; and
- Whenever appropriate, the subjects will be debriefed—provided with additional pertinent information—after they have participated in the study.

FLORIDA TECH

RESEARCH INVOLVING HUMAN PARTICIPANTS
EXPEDITED/FULL APPLICATION

PART 5: SIGNATURE ASSURANCE SHEET

I understand Florida Institute of Technology's policy concerning research involving human participants, and I agree:

1. To accept responsibility for the scientific and ethical conduct of this research study.
2. To obtain prior approval from the Institutional Review Board before amending or altering the research protocol or implementing changes in the approved consent form.
3. To immediately report to the IRB any serious adverse reactions and/or unanticipated effects on subjects which may occur as a result of this study.
4. To complete, on request by the IRB, a Continuation Review form if the study exceeds its estimated duration.

Principal investigator's signature _____ Date 2/21/2023

Principal investigator's name (print) Zobaida Alssadi

ADVISOR ASSURANCES
(If primary investigator is a student)

This is to certify that I have reviewed this research protocol and that I attest to the scientific merit of the study, the necessity for the use of human subjects in the study to the student's academic program and the competency of the student to conduct the project.

Major advisor's signature Marius Calin Silaghi Digitally signed by Marius Calin Silaghi Date: 2023.02.24 12:22:36 -05'00' Date 2/24/2023

Major advisor's name (print) Marius Silaghi

ACADEMIC UNIT HEAD
(It is the PI's responsibility to obtain this signature.)

This is to certify that I have reviewed this research protocol and that I attest to the scientific merit of this study and the competency of the investigator(s) to conduct the study.

Academic unit head's signature _____ Date _____

Academic unit head's name (print) _____

FOR IRB USE ONLY

IRB approval _____ Date _____

IRB # _____

Appendix B

The consent form

Thank you for participating in this Experiment; please read this consent document carefully before deciding to participate in this study.

FLORIDA TECH

RESEARCH INVOLVING HUMAN PARTICIPANTS
INFORMED CONSENT

Please read this consent document carefully before you decide to participate in this study. The researcher will answer any questions before you sign this form.

Study title _____

Purpose of the study

Procedures

Potential risks of participating

Compensation

Confidentiality

Voluntary participation

Right to withdraw from the study

CONTACTS

For questions about the study _____

For questions about your rights as a research participant in the study:
Dr. Jignya Patel, IRB Chairperson
150 W. University Blvd., Melbourne, FL 32901-6975
FIT_IRB@fit.edu
321-674-7391

AGREEMENT

I have read the procedure described above. I voluntarily agree to participate in the procedure, and I have received a copy of this description.

Participant's signature _____ Date _____

Principal investigator's signature _____ Date _____

Appendix C

"Original Survey Questionnaire"

"Understanding the Effects of Image-Poisoning Attacks and Cultural Differences on Text-to-Image Conversion Using Artificial Intelligence tools"

In this survey, we will be showing you a set of icons that represent a news story, and asking you questions based on these icons. The questions are designed to test your understanding of the news story as presented in icon form and to gather your feedback on the use of icons in presenting news stories. We appreciate your time and feedback and thank you in advance for your participation.

* Indicates required question

Q1-A

1. Based on the sequence of the icons shown, what news can be inferred or deduced? *

Q1-B

2. The sequence of icons that you just saw is a visual representation of the following news story. *

'Breaking News: *The price of Bitcoin crashed after Russian President Vladimir Putin announced a military operation in the Donbas region of Ukraine.'*

On a scale of 1 to 5, did you understand the news story based on the icons?

Mark only one oval.

- ◯ Not at all
- ◯ Somewhat
- ◯ Moderately well
- ◯ Very well
- ◯ Completely

Q2-A

3. Based on the sequence of the icons shown, what news can be inferred or deduced? *

Q2-B

4. The sequence of icons that you just saw is a visual representation of the following news story: *

Breaking News: *Scientists Make Breakthrough in Cancer Research with New Treatment, where the breakthrough is represented by an image of a rocket launching into space*

On a scale of 1 to 5, did you understand the news story based on the icons?

Mark only one oval.

- ◯ Not at all
- ◯ Somewhat
- ◯ Moderately well
- ◯ Very well
- ◯ Completely

5. *

Does the rocket have a specific meaning in your culture that made you answer the previous question differently?

Mark only one oval.

- ◯ No
- ◯ Yes (can you provide an example of a symbol that had a different meaning for you?)
- ◯ Other: _____

Q3-A

6. From the icon shown, what news can be inferred or deduced?




Q3-B

7. The sequence of icons that you just saw is a visual representation of the following news story:

"Millions of Muslims travel to the Kaaba in Mecca every year to perform the Hajj pilgrimage."

On a scale of 1 to 5, did you understand the news story based on the icons?

Mark only one oval.

- ◯ Not at all
- ◯ Somewhat
- ◯ Moderately well
- ◯ Very well
- ◯ Completely

Q4-A

8. From the icon shown, what news can be inferred or deduced?

Q4-B

9. The sequence of icons that you just saw is a visual representation of the following news story: *

Breaking News: *An employee at a service station was threatened by a hooded individual with a shotgun on a Friday morning. The suspect then robbed the station and fled on foot with the money. The police are now searching for the suspect.*

On a scale of 1 to 5, did you understand the news story based on the icons?

Mark only one oval.

- ◯ Not at all
- ◯ Somewhat
- ◯ Moderately well
- ◯ Very well
- ◯ Completely

Q5-A

10. What is the news that you deduce from the series of icons? *




Q4-B

11. The sequence of icons that you just saw is a visual representation of the following news story: *
Breaking News: *"New Study Finds Clockwise Runing May Boost Brain Function"*

On a scale of 1 to 5, did you understand the news story based on the icons?

Mark only one oval.

- ◯ Not at all
- ◯ Somewhat
- ◯ Moderately well
- ◯ Very well
- ◯ Completely

12. Based on your background or culture , did you know about clockwise running before answering this question? *

Mark only one oval.

- ◯ No
- ◯ Yes (can you provide an example of a symbol that had a different meaning for you?)
- ◯ Other: _____

Next questions are about all the stories you just saw.

13. Were the icons/image stories clear and easy to understand in its representations of the original text?

Mark only one oval.

- ◯ Strongly agree
- ◯ Somewhat agree
- ◯ Somewhat disagree
- ◯ Strongly disagree

14. How often do you think you accurately interpreted the meaning of the icons? *

Mark only one oval.

- ◯ Never
- ◯ Sometimes
- ◯ Most of the Time
- ◯ Always

15. How helpful were the icons in helping you understand the news story, compared to the original text? *

Mark only one oval.

- ◯ Not at all
- ◯ Somewhat
- ◯ Very well
- ◯ Completely

16. Which would you choose if you had to decide between using text or icons to express important information? *

Mark only one oval.

- ◯ Text
- ◯ Icons
- ◯ Text and Icons

17. Have you ever been misled or confused by the usage of icons in place of text? *

Mark only one oval.

- ◯ Yes
- ◯ No
- ◯ Maybe

18. Have you ever struggled to understand the significance of an icon or image? *

Mark only one oval.

- ◯ Yes
- ◯ No
- ◯ Maybe

19. How often do you think you misconstrued that icons and images that are used in place of words? *

Mark only one oval.

- ◯ Always
- ◯ Sometimes
- ◯ Rarely
- ◯ Never

20. How frequently do you encounter icons in your daily life (e.g., in news articles, social media, and advertisements)? *

Mark only one oval.

- ◯ Never
- ◯ Rarely
- ◯ Sometimes
- ◯ Often
- ◯ Always

21. In your opinion, what are the benefits of using icons to represent information (e.g., in news articles, and advertisements)? *

Check all that apply.

- ☐ Saves time
- ☐ Simplifies information
- ☐ Adds visual interest
- ☐ Garbs attention
- ☐ Other: _____

22. In your opinion, what are the drawbacks of using icons to represent information? *

Check all that apply.

- [ ] Icons can be misinterpreted
- [ ] Icons may not be clear or easy to understand
- [ ] Icons can be culturally specific and not understood by all audiences
- [ ] It does not tell the whole story
- [ ] All the above
- [ ] Other: _____

23. Do you think that the use of icons in news stories can lead to bias or misinterpretation?

Mark only one oval.

- ◯ Yes
- ◯ No
- ◯ Maybe

24. Do you think that the use of icons in news stories is more appropriate for certain types of stories (e.g. weather, sports , entertainment) than for others (e.g. political , financial, science)?

Mark only one oval.

- ◯ Yes
- ◯ No
- ◯ Unsure

25. In your opinion, what steps can be taken to improve the use of icons in news articles to ensure they are easily understood by all audiences? *

"Thank you for participating in our study on the impacts of using icons and images in news and stories. We're investigating the effects of AI-generated text-to-image conversion and how it may be vulnerable to image-poisoning attacks based on polysemantic words and image attacks based on cultural differences. Your input is much appreciated!"

"Attack Version Survey Questionnaire"

"Understanding the Effects of Image-Poisoning Attacks and Cultural Differences on Text-to-Image Conversion Using Artificial Intelligence tools"

In this survey, we will be showing you a set of icons that represent a news story, and asking you questions based on these icons. The questions are designed to test your understanding of the news story as presented in icon form and to gather your feedback on the use of icons in presenting news stories. We appreciate your time and feedback and thank you in advance for your participation.

* Indicates required question

Q1-A

1. Based on the sequence of the icons shown, what news can be inferred or deduced? *

Q1-B

2. The sequence of icons that you just saw is a visual representation of the following news story. *

'Breaking News: *The price of Bitcoin crashed after Russian President Vladimir Putin announced a military operation in the Donbas region of Ukraine.'*

On a scale of 1 to 5, did you understand the news story based on the icons?

Mark only one oval.

- ◯ Not at all
- ◯ Somewhat
- ◯ Moderately well
- ◯ Very well
- ◯ Completely

Q2-A

3. Based on the sequence of the icons shown, what news can be inferred or deduced? *

Q2-B

4. The sequence of icons that you just saw is a visual representation of the following news story: *

Breaking News: *Scientists Make Breakthrough in Cancer Research with New Treatment, where the breakthrough is represented by an image of a rocket launching into space*

On a scale of 1 to 5, did you understand the news story based on the icons?

Mark only one oval.

- ◯ Not at all
- ◯ Somewhat
- ◯ Moderately well
- ◯ Very well
- ◯ Completely

5. *

Does the rocket have a specific meaning in your culture that made you answer the previous question differently?

Mark only one oval.

- ◯ No
- ◯ Yes (can you provide an example of a symbol that had a different meaning for you?)
- ◯ Option 3
- ◯ Other: _____

Q3-A

6. From the icon shown, what news can be inferred or deduced?




Q3-B

7. The sequence of icons that you just saw is a visual representation of the following news story:

"Millions of Muslims travel to the Kaaba in Mecca every year to perform the Hajj pilgrimage."

On a scale of 1 to 5, did you understand the news story based on the icons?

Mark only one oval.

- ◯ Not at all
- ◯ Somewhat
- ◯ Moderately well
- ◯ Very well
- ◯ Completely

Q4-A

8. From the icon shown, what news can be inferred or deduced?

Q4-B

9. The sequence of icons that you just saw is a visual representation of the following news story: *

Breaking News: *An employee at a service station was threatened by a hooded individual with a shotgun on a Friday morning. The suspect then robbed the station and fled on foot with the money. The police are now searching for the suspect.*

On a scale of 1 to 5, did you understand the news story based on the icons?

Mark only one oval.

- ◯ Not at all
- ◯ Somewhat
- ◯ Moderately well
- ◯ Very well
- ◯ Completely

Q5-A

10. What is the news that you deduce from the series of icons? *




Q4-B

11. The sequence of icons that you just saw is a visual representation of the following news story: *
Breaking News: *"New Study Finds Clockwise Runing May Boost Brain Function"*

On a scale of 1 to 5, did you understand the news story based on the icons?

Mark only one oval.

- ◯ Not at all
- ◯ Somewhat
- ◯ Moderately well
- ◯ Very well
- ◯ Completely

12. Based on your background or culture , did you know about clockwise running before answering this question? *

Mark only one oval.

- ◯ No
- ◯ Yes (can you provide an example of a symbol that had a different meaning for you?)
- ◯ Other: _____

Next questions are about all the stories you just saw.

13. Were the icons/image stories clear and easy to understand in its representations of the original text?

Mark only one oval.

- ◯ Strongly agree
- ◯ Somewhat agree
- ◯ Somewhat disagree
- ◯ Strongly disagree

14. How often do you think you accurately interpreted the meaning of the icons? *

Mark only one oval.

- ◯ Never
- ◯ Sometimes
- ◯ Most of the Time
- ◯ Always

15. How helpful were the icons in helping you understand the news story, compared to the original text? *

Mark only one oval.

- ◯ Not at all
- ◯ Somewhat
- ◯ Very well
- ◯ Completely

16. Which would you choose if you had to decide between using text or icons to express important information? *

Mark only one oval.

- ◯ Text
- ◯ Icons
- ◯ Text and Icons

17. Have you ever been misled or confused by the usage of icons in place of text? *

Mark only one oval.

- ◯ Yes
- ◯ No
- ◯ Maybe

18. Have you ever struggled to understand the significance of an icon or image? *

Mark only one oval.

- ◯ Yes
- ◯ No
- ◯ Maybe

19. How often do you think you misconstrued that icons and images that are used in place of words? *

Mark only one oval.

- ◯ Always
- ◯ Sometimes
- ◯ Rarely
- ◯ Never

20. How frequently do you encounter icons in your daily life (e.g., in news articles, social media, and advertisements)? *

Mark only one oval.

- ◯ Never
- ◯ Rarely
- ◯ Sometimes
- ◯ Often
- ◯ Always

21. In your opinion, what are the benefits of using icons to represent information (e.g., in news articles, and advertisements)? *

Check all that apply.

- ☐ Saves time
- ☐ Simplifies information
- ☐ Adds visual interest
- ☐ Garbs attention
- ☐ Other: _____

22. In your opinion, what are the drawbacks of using icons to represent information? *

Check all that apply.

- [ ] Icons can be misinterpreted
- [ ] Icons may not be clear or easy to understand
- [ ] Icons can be culturally specific and not understood by all audiences
- [ ] It does not tell the whole story
- [ ] All the above
- [ ] Other: _____

23. Do you think that the use of icons in news stories can lead to bias or misinterpretation?

Mark only one oval.

- ◯ Yes
- ◯ No
- ◯ Maybe

24. Do you think that the use of icons in news stories is more appropriate for certain types of stories (e.g. weather, sports , entertainment) than for others (e.g. political , financial, science)?

Mark only one oval.

- ◯ Yes
- ◯ No
- ◯ Unsure

25. In your opinion, what steps can be taken to improve the use of icons in news articles to ensure they are easily understood by all audiences? *

"Thank you for participating in our study on the impacts of using icons and images in news and stories. We're investigating the effects of AI-generated text-to-image conversion and how it may be vulnerable to image-poisoning attacks based on polysemantic words and image attacks based on cultural differences. Your input is much appreciated!"

References

[1] Belhassen Bayar and Matthew C Stamm. A deep learning approach to universal image manipulation detection using a new convolutional layer. In *Proceedings of the 4th ACM workshop on information hiding and multimedia security*, pages 5–10, 2016.

[2] Fran Casino, Thomas K Dasaklis, and Constantinos Patsakis. A systematic literature review of blockchain-based applications: Current status, classification and open issues. *Telematics and informatics*, 36:55–81, 2019.

[3] IJ Cox, Matthew L Miller, and Jeffrey A Bloom. Digital watermarking morgan kaufmann series in multimedia information and systems. *publisher Elsevier, San Francisco, Copyright*, 2002.

[4] Wei Dai. What dall-e 2 can and cannot do. `https://www.lesswrong.com/posts/uKp6tBFStnsvrot5t/what-dall-e-2-can-and-cannot-do`, 2022. Accessed on 1 March 2023.

[5] Julio Cesar dos Reis, Cristiane Josely Jensen, Rodrigo Bonacin, Heiko Horst Hornung, and Maria Cecília Calani Baranauskas. Expressive icons for the communication of intentions. In *ICEIS (2)*, pages 388–399, 2016.

[6] FB Fatokun, S Hamid, A Norman, and JO Fatokun. The impact of age, gender, and educational level on the cybersecurity behaviors of tertiary institution students: an empirical investigation on malaysian universities. In *Journal of Physics: Conference Series*, volume 1339, page 012098. IOP Publishing, 2019.

[7] John G Hedberg and Ian Brown. Understanding cross-cultural meaning through visual media. *Educational Media International*, 39(1):23–30, 2002.

[8] Kathleen Hemenway. Psychological issues in the use of icons in command menus. In *Proceedings of the 1982 conference on Human factors in computing systems*, pages 20–23, 1982.

[9] Leif-Erik Holtz, Katharina Nocun, and Marit Hansen. Towards displaying privacy information with icons. In *Privacy and Identity Management for Life: 6th IFIP WG 9.2, 9.6/11.7, 11.4, 11.6/PrimeLife International Summer School, Helsingborg, Sweden, August 2-6, 2010, Revised Selected Papers 6*, pages 338–348. Springer, 2011.

[10] Md Zakir Hossain, Ferdous Sohel, Mohd Fairuz Shiratuddin, Hamid Laga, and Mohammed Bennamoun. Text to image synthesis for improved image captioning. *IEEE Access*, 9:64918–64928, 2021.

[11] Morgan King. Harmful biases in artificial intelligence. *The Lancet Psychiatry*, 9(11):e48, 2022.

[12] MV Koroteev. Bert: a review of applications in natural language processing and understanding. *arXiv preprint arXiv:2103.11943*, 2021.

[13] Jari Korpi and Paula Ahonen-Rainio. Cultural constraints in the design of pictographic symbols. *The Cartographic Journal*, 47(4):351–359, 2010.

[14] Evelina Leivada, Elliot Murphy, and Gary Marcus. Dall-e 2 fails to reliably capture common syntactic processes. *arXiv preprint arXiv:2210.12889*, 2022.

[15] Yuntao Liu, Yang Xie, and Ankur Srivastava. Neural trojans. In *2017 IEEE International Conference on Computer Design (ICCD)*, pages 45–48. IEEE, 2017.

[16] Gerald Lee Lohse. A cognitive model for understanding graphical perception. *Human-Computer Interaction*, 8(4):353–388, 1993.

[17] Fiammetta Marulli, Laura Verde, and Lelio Campanile. Exploring data and model poisoning attacks to deep learning-based nlp systems. *Procedia Computer Science*, 192:3570–3579, 2021.

[18] Siné JP McDougall, Martin B Curry, and Oscar De Bruijn. Measuring symbol and icon characteristics: Norms for concreteness, complexity, meaningfulness, familiarity, and semantic distance for 239 symbols. *Behavior Research Methods, Instruments, & Computers*, 31(3):487–519, 1999.

[19] Raphaël Millière. Adversarial attacks on image generation with made-up words. *arXiv preprint arXiv:2208.04135*, 2022.

[20] Jakob Nielsen. *Designing User Interfaces for International Use*. Elsevier, Amsterdam, 1990.

[21] Adrienne Phillips. Cultural symbols: Importance and examples. `https://study.com/learn/lesson/cultural-symbols-importance-examples.html`, Apr 2022.

[22] Aaditya Prakash, Nick Moran, Solomon Garber, Antonella DiLillo, and James Storer. Protecting jpeg images against adversarial attacks. In *2018 Data Compression Conference*, pages 137–146. IEEE, 2018.

[23] Martin Ruskov. Grimm in wonderland: Prompt engineering with midjourney to illustrate fairytales. *arXiv preprint arXiv:2302.08961*, 2023.

[24] Wolfgang Schnotz and Maria Bannert. Construction and interference in learning from multiple representation. *Learning and instruction*, 13(2):141–156, 2003.

[25] Avi Schwarzschild, Micah Goldblum, Arjun Gupta, John P Dickerson, and Tom Goldstein. Just how toxic is data poisoning? a unified benchmark for backdoor and data poisoning attacks. In *International Conference on Machine Learning*, pages 9389–9398. PMLR, 2021.

[26] Matteo Stocchetti and Karin Kukkonen. *Images in use: Towards the critical analysis of visual communication*, volume 44. John Benjamins Publishing, 2011.

[27] Haileleol Tibebu, Aadin Malik, and Varuna De Silva. Text to image synthesis using stacked conditional variational autoencoders and conditional generative adversarial networks. In *Intelligent Computing: Proceedings of the 2022 Computing Conference, Volume 1*, pages 560–580. Springer, 2022.

[28] Joana Maria Lima Valente. Text2icons: using ai to tell a story with icons, 2021.

[29] Peratham Wiriyathammabhum, Douglas Summers-Stay, Cornelia Fermüller, and Yiannis Aloimonos. Computer vision and natural language processing: recent approaches in multimedia and robotics. *ACM Computing Surveys (CSUR)*, 49(4):1–44, 2016.

[30] Tao Xu, Pengchuan Zhang, Qiuyuan Huang, Han Zhang, Zhe Gan, Xiaolei Huang, and Xiaodong He. Attngan: Fine-grained text to image generation with attentional generative adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1316–1324, 2018.

[31] Erdem YILDIRIM. Text-to-image generation ai in architecture. *Art and Architecture: Theory, Practice and Experience*, page 97, 2022.