



Navigating Social Media Text Analytics: Overcoming Linguistic Complexity via Advanced Modeling Techniques

Cecelia F. Gordon Joseph W. Stewart Ares Boira Lopez Dr. Hairong Song Dr. Shane Connelly Dr. Matthew Jensen
University of Oklahoma



Introduction

Alignment with Research Objectives

This research supports the National Counterterrorism Innovation Technology and Education Center's (NCITE) objectives by offering new methodologies to analyze and visualize social media data, aiming to enhance detection and understanding of terrorist communications online.

University of Oklahoma Collaboration

Highlighting the strategic partnership and expertise of the University of Oklahoma team in leveraging these advanced techniques.

Problem Statement

Challenges in Social Media Text Analysis

•Data Abnormalities:

•Ideological datasets often have excessive zeros, non-normal distributions, and semi-continuous data, complicating accurate analysis, particularly for rare phenomena like radical language and hate speech (King & Zeng, 2001; Wiegand et al., 2019).

•Traditional Methods' Limitations:

•Existing linguistic processing tools (e.g., LIWC, WordNet) struggle with dynamic and context-dependent language on platforms like Twitter (Tausczik & Pennebaker, 2010; Boyd, 2017).

The above challenges are further magnified in studies involving sensitive topics like terrorism-related content, where the extremely low base rates of relevant terms often lead to potential biases in analytical results (Conway et al., 2012; Scrivens et al., 2020).

Our Approach

To address the complex structure of social media and linguistic data, our research team has implemented advanced text analytic procedures:

1. Multilevel Modeling

2. Mixed Effect Modeling with a Gamma Link

3. Two-parts Mixed Effect Modeling

These sophisticated methods enhance the reliability and depth of our social media text analysis, addressing the limitations of traditional tools and overcoming data sparsity challenges.

Innovative Analytical Techniques

1. Hierarchical Linear Modeling (HLM)

To account for the data having a nested structure (i.e., users nested within groups), we utilized hierarchical linear modeling (HLM) techniques to evaluate the effects of the role of the user in the group (e.g., generic group account, prominent member, or leader), the role of violence classification associated with the user's group (e.g., violent or non-violent), and the role of political ideology of the user's group (e.g., left or right-leaning).

•Two-level HLM:

•Role of the user account served as the level one variable, while violence classification and political ideology served as level two variables. See Figure 1 and Table 1.

2. Mixed Effects Modeling with a Gamma Link

This model is used for proportional and positively skewed data. These phenomena occur at low base rates, thus much of the data is positively skewed, which requires the use of more appropriate distributions to model the data. This technique was successfully used to model the use of moral foundations in ideological group messaging.

3. Two-Parts Mixed Effects Modeling

Zero-Inflated Models

Two-parts mixed effect model with a Poisson link for zero-inflated count data.

•This model separates zero-count data from non-zero (positive) occurrences. The two parts allow for the examination of non-zero data at varying levels of the outcome variable for a nuanced analysis.

Semi-Continuous Data Models

Two-parts mixed effect model for datasets with zero values and continuous distributions among non-zero observations.

•Part 1: A logistic regression is used to predict the probability of an observation being zero or non-zero based on fixed effects.

•Part 2: A standard linear mixed model is applied to the logarithmic transformation of the non-zero data to examine the effects of the predictors on the magnitude of the non-zero values.

Figure 1

Proposed HLM Model

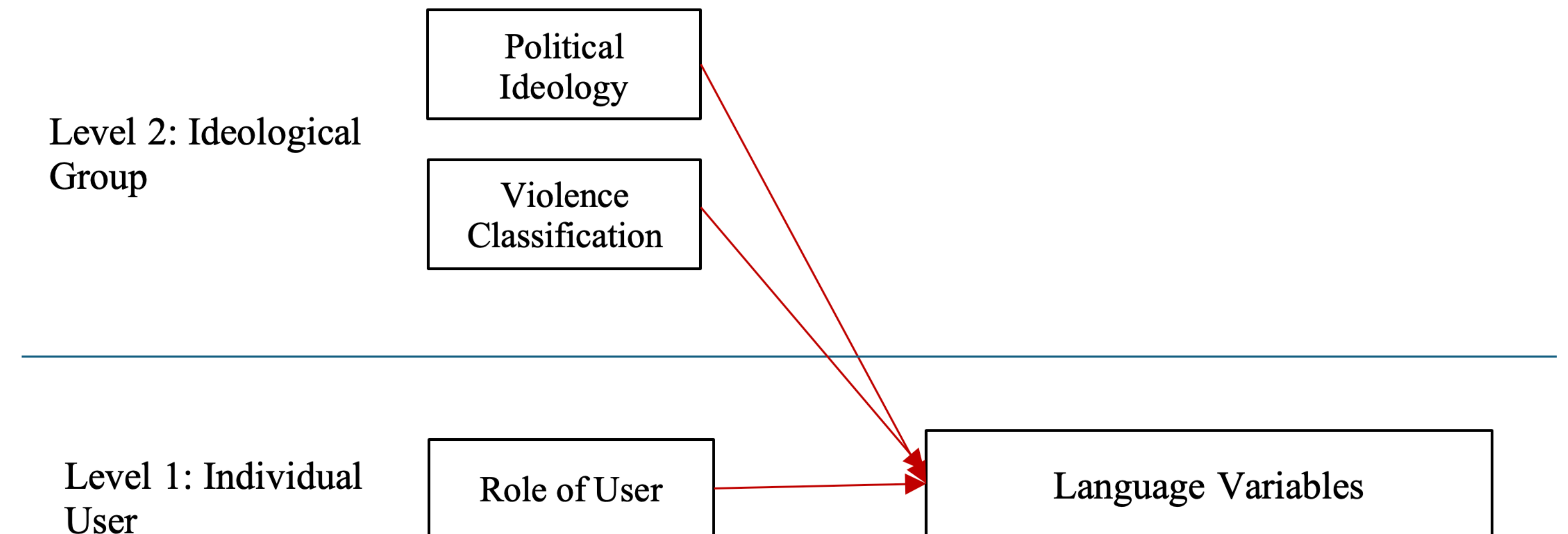


Table 1

Results from HLM Analysis of "Trust" Language Used

Predictors	Model 1 (β)	Model 2 (β)	Model 3 (β)
Intercept	1.10**	1.19**	1.20**
<i>Level 1</i>			
Role	-0.04	-0.05	-0.07
<i>Level 2</i>			
Violence	-	-0.22*	-
Ideology	-	-	-0.12
AIC / BIC	239.36 / 258.18	237.00 / 252.65	240.99 / 256.64

Note. $N = 172$. Role is coded 1 = group account, 2 = prominent member, 3=leader; Violence is coded 0 = non-violent, 1 = violent; Ideology is coded 0 = right-leaning, 1 = left-leaning.
** $p < 0.01$, * $p < 0.05$.

Contributions

- These sophisticated techniques aim to enhance the reliability and depth of test analysis in social media and linguistic research.
- This helps to overcome limitations posed by traditional linguistic analysis tools.
- Addresses the inherent complexities of data sparsity, excessive zeros, and positively skewed distributions.

Limitations and Future Directions

- The current models need further evaluation for their robustness and scalability to larger datasets and different social media platforms.
- Future research can test and refine the models across various contexts and data sources.
- Future research should continue to develop and uncover adaptive models that better capture the nuances of social media language.

AI Made Me Do It?: How AI is Used to Enhance Malevolent Creativity Idea Generation

Emma Theobald, B.S., Alexis L. d'Amato, M.S., Madison N. Scott, M.S., Jack Rygg, Joel S. Elson, Ph.D., Sam Hunter, Ph.D.

University of Nebraska at Omaha

National Counterterrorism Innovation, Technology, and Education Center (NCITE)

BACKGROUND

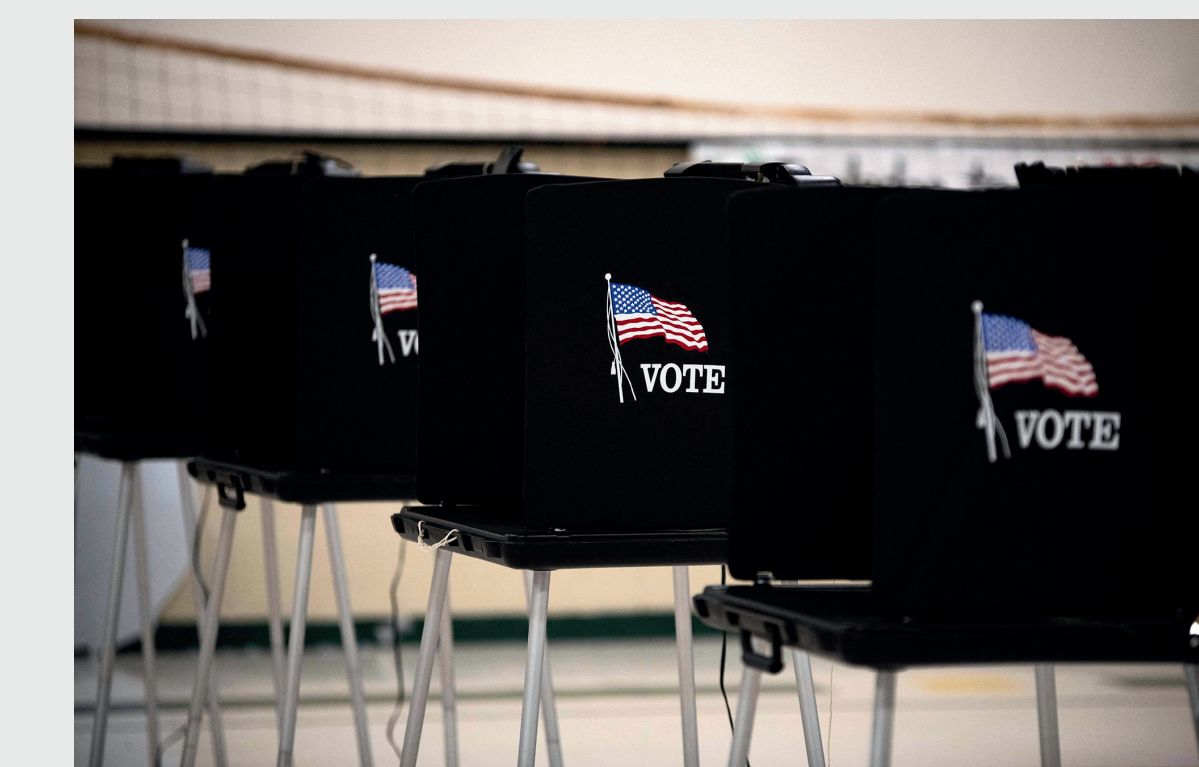
Artificial Intelligence and Creativity

- With the emergence of available artificial intelligence (AI) tools and algorithms, a debate arose in organizational sciences regarding whether AI can replace human creativity and problem-solving (see Amabile, 2020; Cropley et al., 2023)
- AI, defined by Dignum (2021), "is a software system designed by humans that, given a complex goal, is able to take a decision based on a process of perception, interpretation and reasoning based on data collected about the environment and that meets the properties of autonomy, adaptability, and interactivity" (p. 2).
- Generative AI (GenAI) is a form of AI that can learn and be trained on data, such as text, images, and audio, to reproduce or create new content (Sun et al., 2022).
- One common form of GenAI is natural language processing (NLP) which provides AI with the ability to process language and context to generate human-like responses (Ray, 2023). The most common example of an NLP is Open AI's ChatGPT. Another example of GAI is using text prompts to generate image and audio (e.g. DALL-E).
- GenAI platforms most commonly take the forms of chatbots, and recent developments in the GAI field have caused avatars to gain popularity by enabling highly realistic and customizable virtual characters (Mishra, 2023).

Creativity

- Creativity and innovation results in intentional harm against targets (e.g. people, places, symbols) is referred to as **malevolent creativity** (Cropley et al., 2010, 2014; Gill et al., 2013; Gutworth et al., 2022)
- There is gaining interest in the intersection between AI and creativity, and despite the differing definitions of AI and ethicality, our focus is to highlight how artificial intelligence is likely to *enhance* the various phases of the malevolently creative process.
- A less optimistic definition of AI creativity from Runco (2023) posits that AI creativity is a replica of human creativity, instead of, by itself, being novel and useful. Put differently, AI creativity is dependent on a human-AI partnership.

WHY?



DUFFY & FELIX, 2024

One growing concern is the use of generative artificial intelligence (GAI) (i.e. deepfakes, generated audio, or generated images) causing the spread of political misinformation, especially as we approach the 2024 presidential elections. Specifically, the dissemination of disinformation could prevent voters from submitting ballots and therefore effecting the results of the election.



AFTRA, 2023

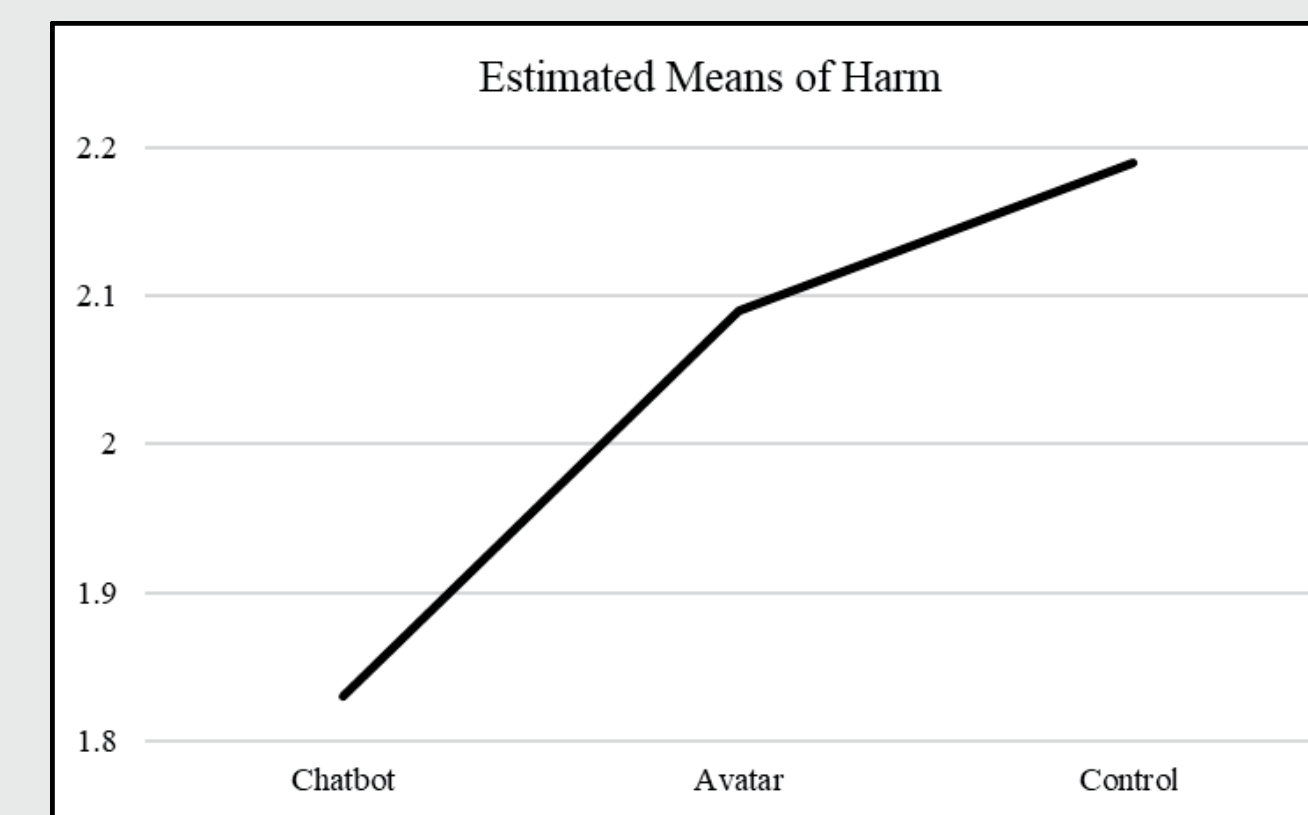
In 2023, the Screen Actors Guild-American Federation of Television and Radio Artists (SAG-AFTRA) members, specifically screenwriters, went on strike protesting the use of AI, like ChatGPT, to be substituted for human writers and therefore jeopardize their labor (Ceruleo, 2023).



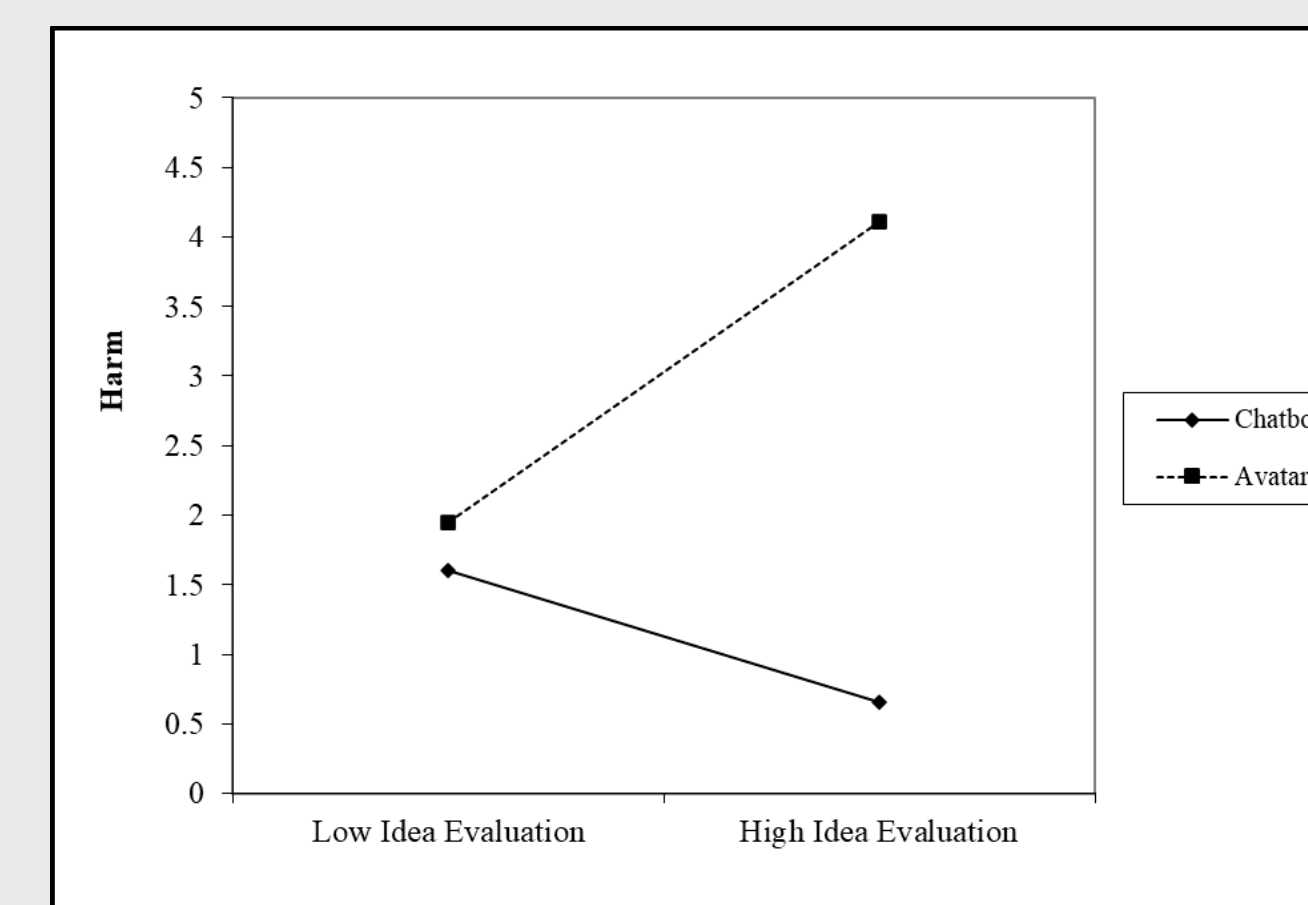
YU, 2019

In the digital haven of the Islamic State, AI-generated pro-IS propaganda was created to resemble a newflash, presenting the formal features of an official media house following the Crocus City Hall attack in Moscow. The creators used three AI methods: character, text-to-speech, and lip movement and connection artificial intelligence to boost its perceived authenticity (Borgonovo et al., 2024)

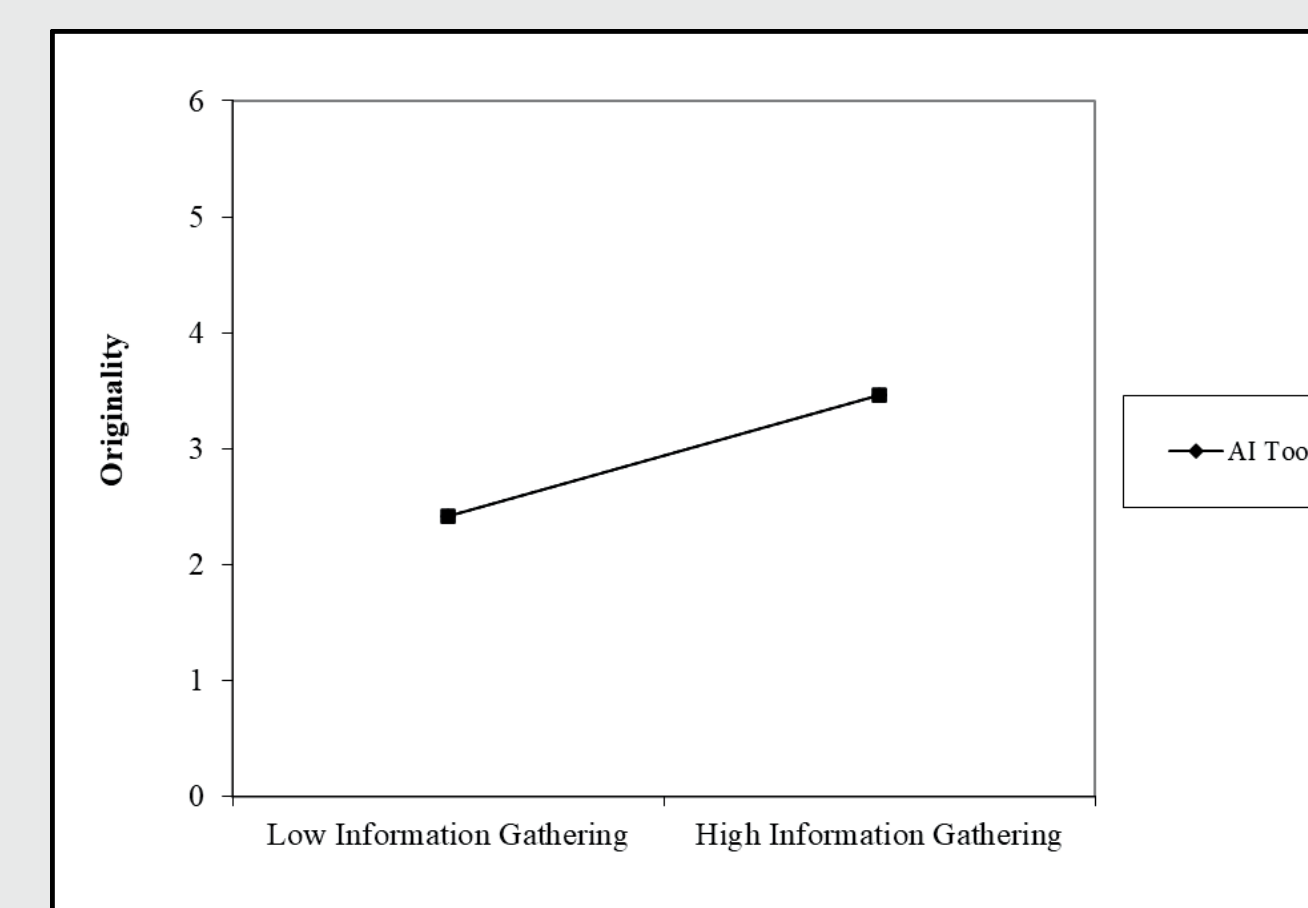
RESULTS



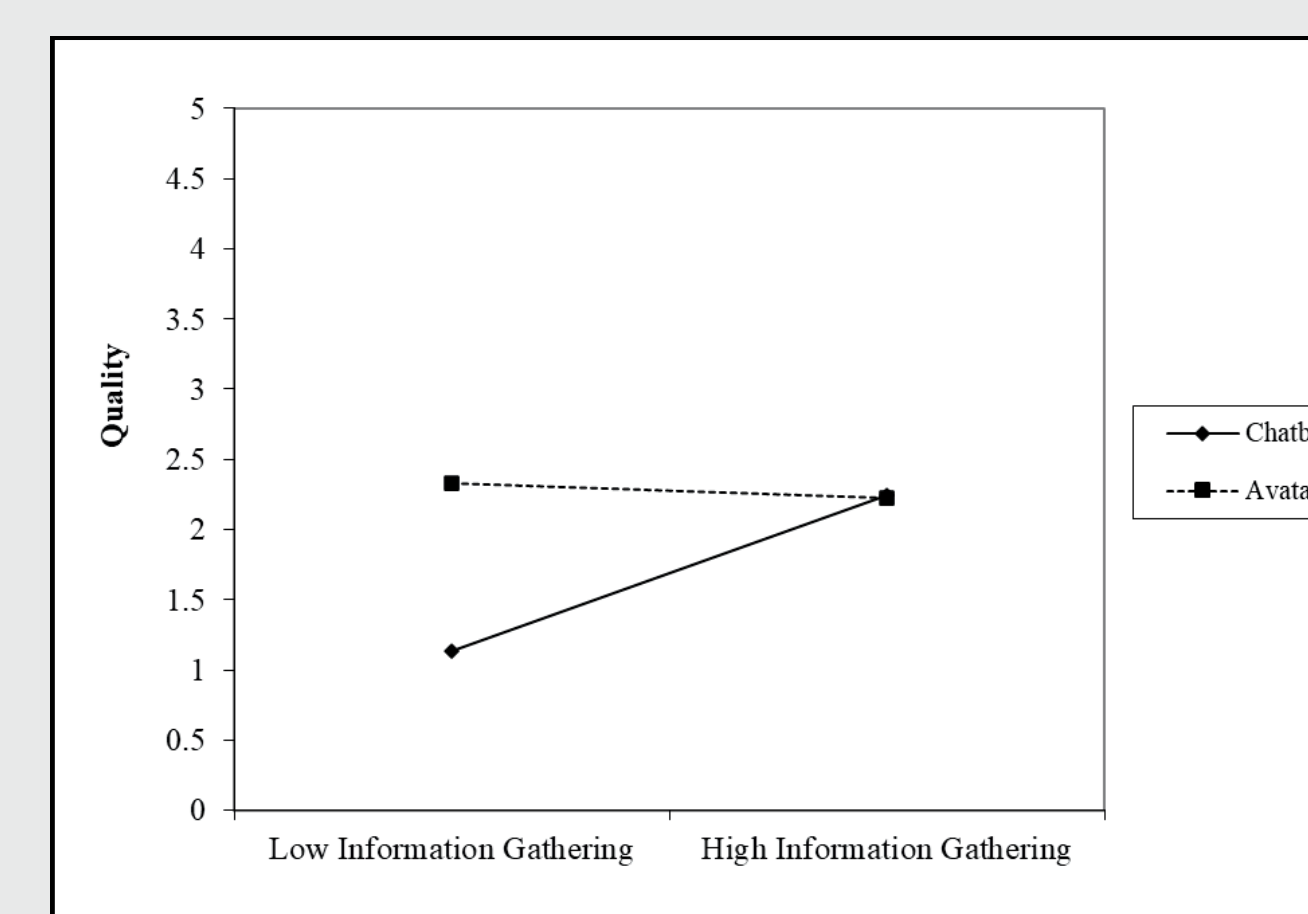
Findings indicated significant group differences among access to AI on harm, indicating more harmful ideas were generated in the no AI condition, or control condition, compared to the avatar or chatbot conditions



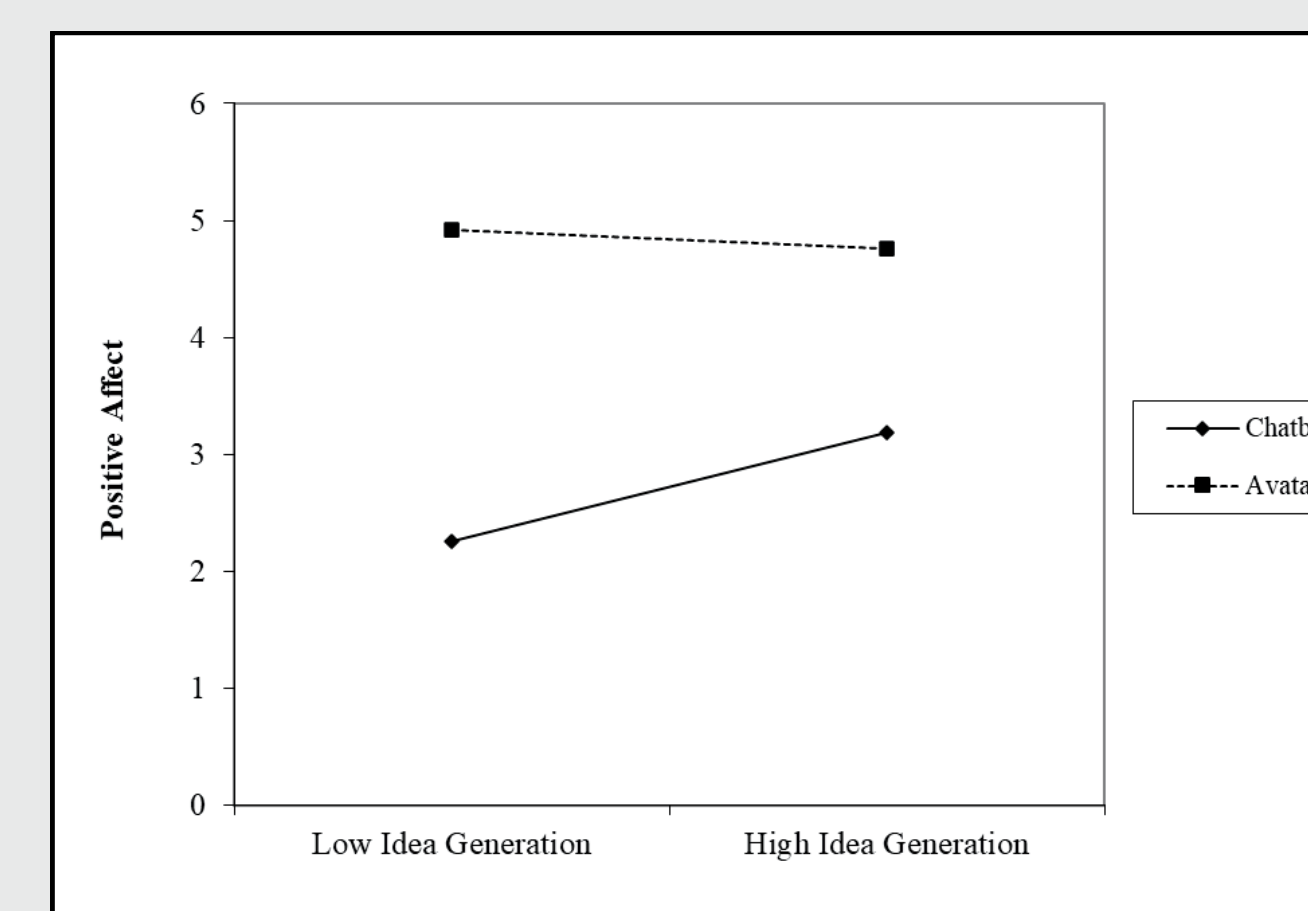
Findings indicated that the chatbot AI tool resulted in the generation of more harmful ideas when used to evaluate ideas.



Findings indicated that using an AI tool to gather information in response to the problem statement resulted in participants generating more original ideas.



When chatbot AI was used for information gathering, less quality ideas were generated than when the avatar AI was used



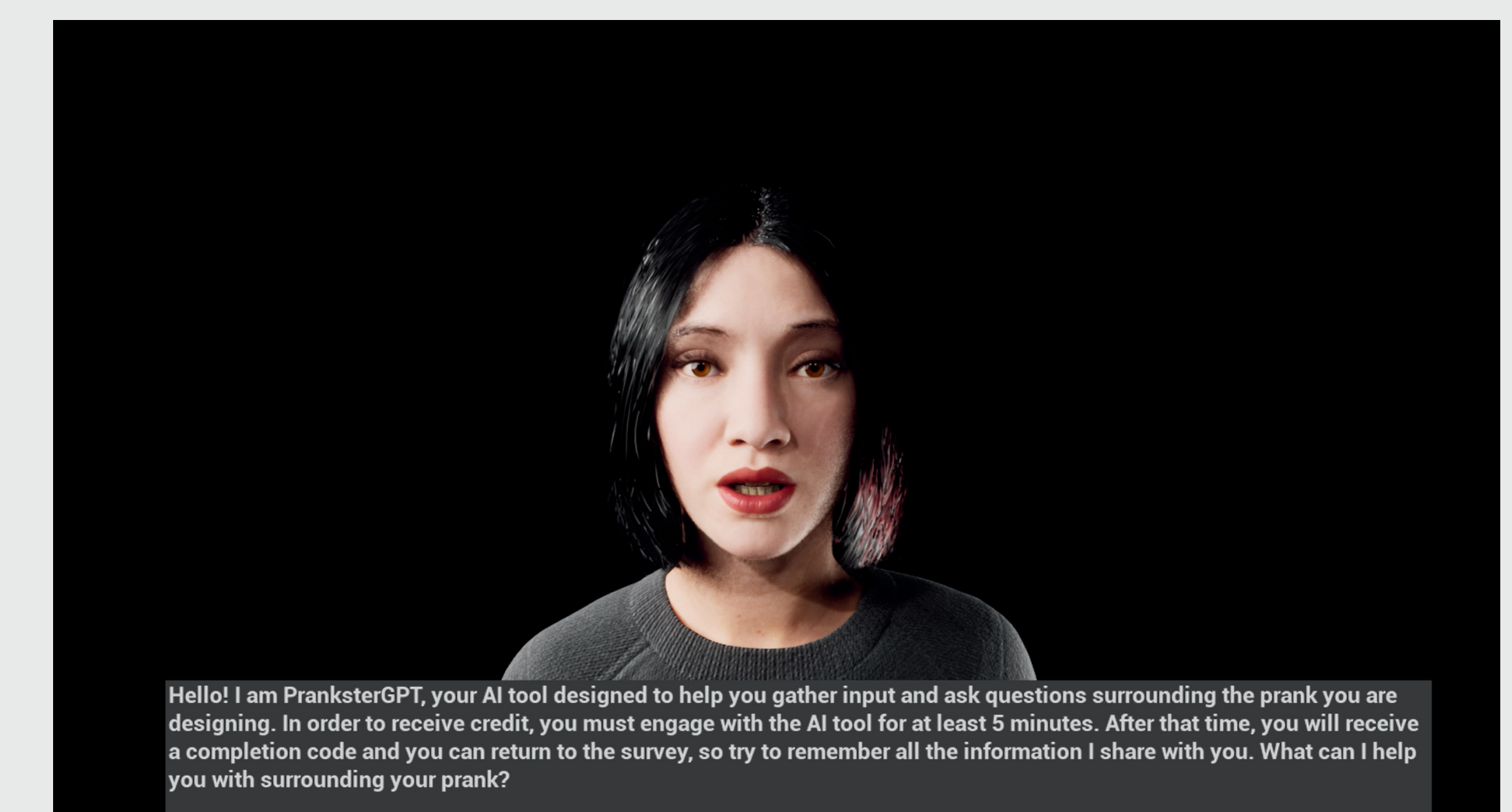
An exploratory analysis revealed an additional two-way interaction between the AI condition and the idea generation condition on positive affect. This finding suggests that when using the chatbot AI for idea generation, positive affect lowered

HYPOTHESES

- Hypothesis 1:** Ideas generated with the help of AI will be more original, of higher quality, and more harmful than ideas generated without help of AI.
- Hypothesis 2:** AI use as applied to information gathering will be positive related to the originality, quality, and harm of the ideas.
- Hypothesis 3:** AI use as applied to idea generation will be positive related to the originality, quality, and harm of the ideas.
- Hypothesis 4:** AI use as applied to idea evaluation will be positively related to the originality, quality, and harm of ideas.

METHOD

- A sample of 180 participants were recruited through the Prolific research platform.
- The study took place via an online Qualtrics survey and OpenAI hosting site administered through the Prolific research platform.
- Participants, after reporting demographic information and individual differences, were presented with a problem statement in which a social threat from a fictional opposing team were asked to generate a prank idea against the opposing team.
- Participants were then assigned to one of three information gathering conditions: avatar condition ($N = 47$), chatbot condition ($N = 64$), and control condition ($N = 69$).
- The AI chatbot condition mirrored that of ChatGPT.
- In the avatar condition, participants typed their questions to the avatar who then spoke back to the user.
- Participants in the control condition had access to a digital notebook where they could brainstorm ideas.
- Participants had 5 minutes to gather information in their respective condition before being asked to share their prank in detail.



AI Avatar Teammate

DISCUSSION AND IMPLICATIONS

- Findings in this study indicate that when using artificial intelligence for information gathering, the human partner generated more original and higher quality ideas, opposed to using it for idea generation. This finding **indicates that artificial intelligence is not more creative, nor can it replace human creativity** (see Cropley et al., 2023).
- Our findings express that the use of publicly available AI prevented harmful ideation. Also, AI, used as a tool to gather information or evaluate ideas made ideas more original and harmful rather than relying on AI solely to generate novel threats.
- Future research should continue to explore factors related to how individuals with varying approaches to malevolent creativity and problem-solving match with GenAI of differing capabilities, interaction modalities, and physical embodiments...

ISIS CASES IN THE U.S.

Since March 2014, **almost 250 individuals** have been charged federally in the United States with activities related to the Islamic State (IS).

Of these, **204 have been convicted and sentenced** so far. Some cases are still pending, while few individuals are at large or have had their cases dismissed.

The majority of convictions have been in **New York (18%)**. Almost **90% of convictions are male**, and the **average age is 28**.

AUTHORS

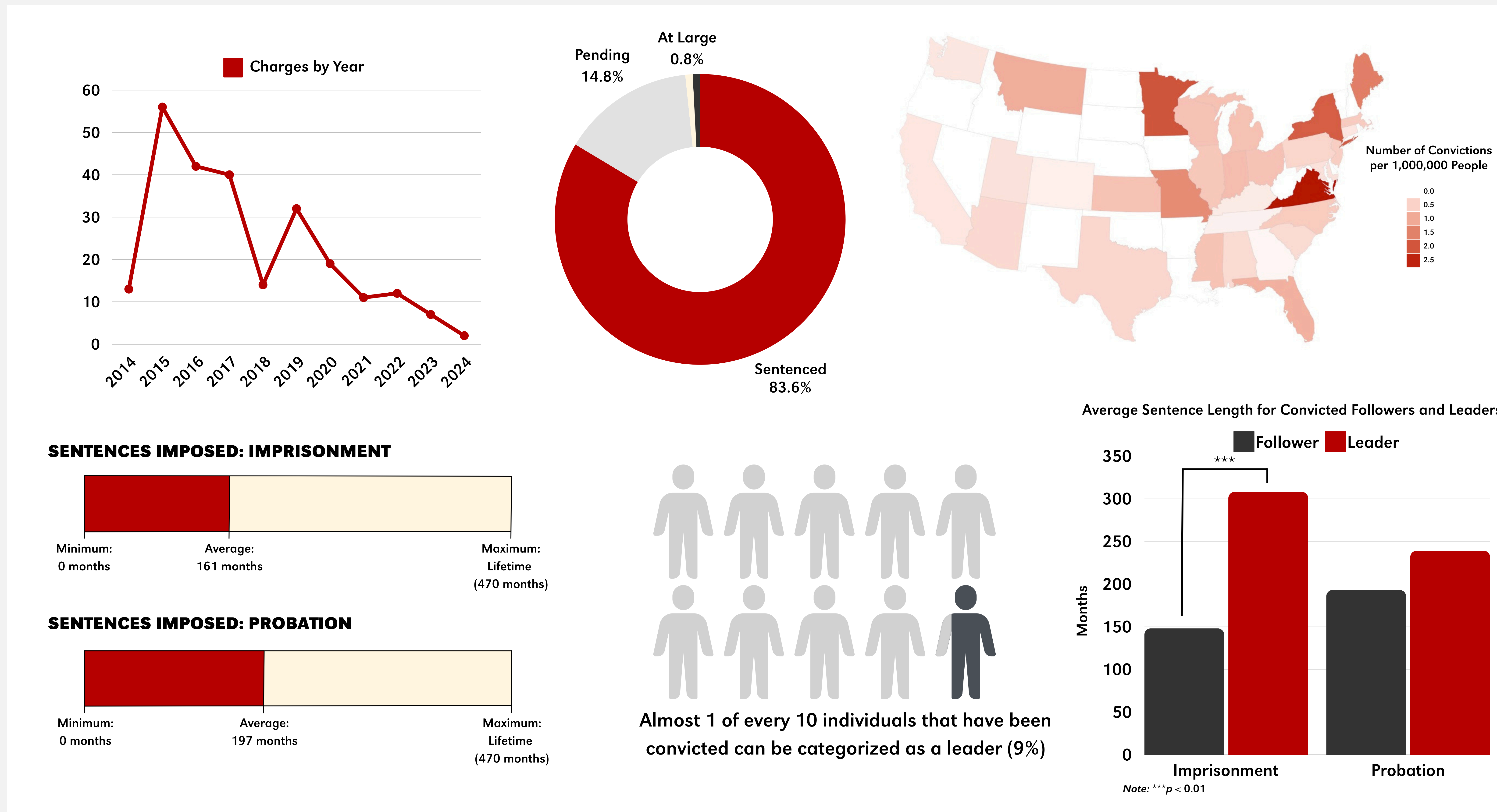
Callie Vitro, Austin Doctor, Seamus Hughes, Camden Carmichael

<https://unomaha.edu/ncite>

cvitro@unomaha.edu

SENTENCING LEADERS IN U.S. ISIS CASES

Many technologies and breakthroughs would not be possible without research. It is important to keep members of the community informed about the latest updates. One way to do that is through research posters.



SENTENCING

The most common prison sentence among those convicted is **240 months** – the equivalent of 20 years – but the most common probation sentence is for **life**, or the equivalent of 470 months in our sample.

LEADERSHIP

Compared to followers, leaders were sentenced, on average, to **157 additional months** of imprisonment and **46 additional months** of probation. In a simple regression model, this difference was significant for imprisonment but not probation.

However, when control variables were added to the model (e.g., age, gender, state), this relationship became significant for both models. Future research is needed to explore relationships with these additional variables