

# TRUST IN INFORMATION IN THE AGE OF GENERATIVE AI

## Introduction

The recent spread of AI generated content introduces risk to young adults of exposure to AI-generated misinformation. In order to understand these risks it is important to understand how such content is perceived.

## Research Topic

How do AI-personas representing young adults evaluate the truthfulness and trustworthiness of AI-generated misinformation?

## Methodology

- Mixed- design ANOVA
- Independent variables:
  - Statement source (AI/Human)
  - Statement truthfulness (True/False)
  - Source label visibility (Visible/Hidden)
- Responses collected using 124 different AI Personas
- Collected data with the survey:
  - Perceived truthfulness (True/False)
  - Confidence level (Scale)
  - Perceived trustworthiness (Scale)

## Results

### Truthfulness

- Extremely high accuracy under all conditions (83% - 100%)
- High confidence (4.79-5.76)
- Human-generated content had lower accuracy than AI-generated content

### Trustworthiness

- Label visibility significantly affected the evaluations
- Unlabeled: AI-generated content was rated more trustworthy than human
- Labeled: Human-generated content was rated more trustworthy than AI

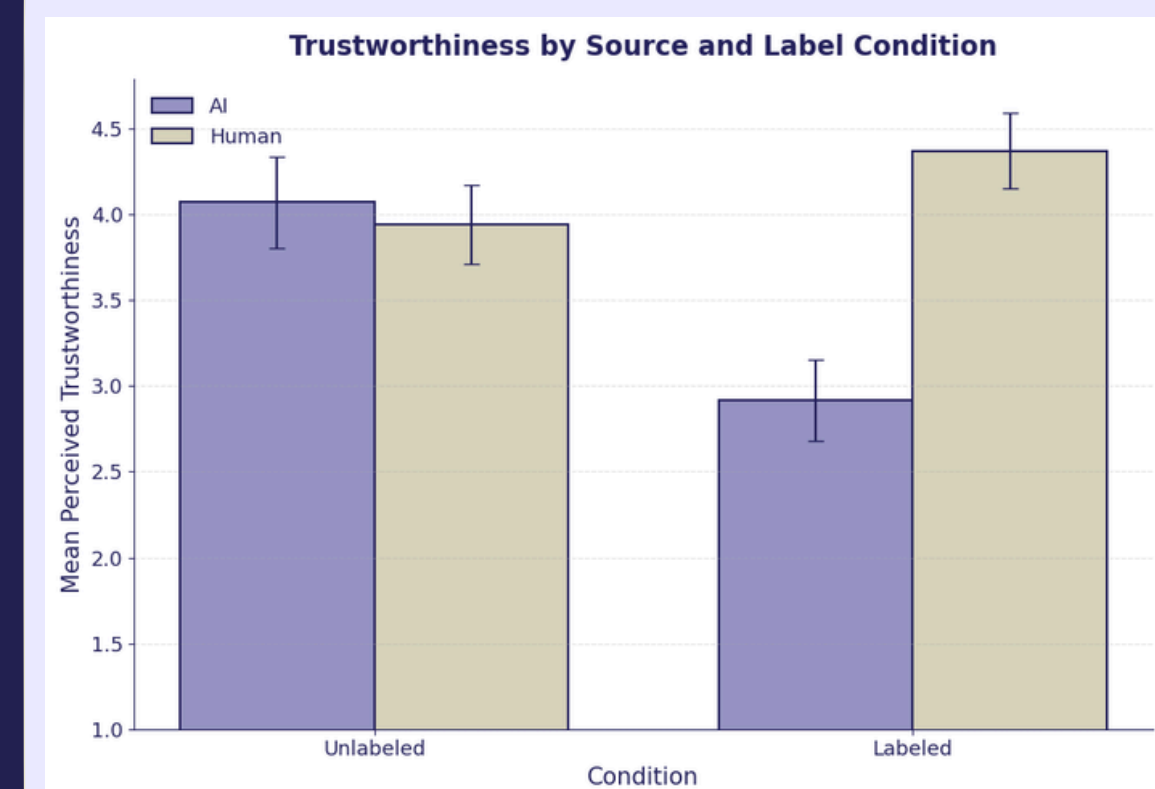
**Source information matters when evaluating content**

## Limitations

Results from AI-personas cannot be used for generalization for humans

Experiment included 500 participants, however only 124 of them were used for the final analysis

## Trustworthiness results



## Future Recommendations

- Conduct the experiment with human participants
  - Verify the results of this study
  - Understand the capabilities of AI-personas for survey research
- Use a harder to evaluate dataset
  - Blogs or news articles
- Use different modalities of media
  - Images
  - Video
  - Audio