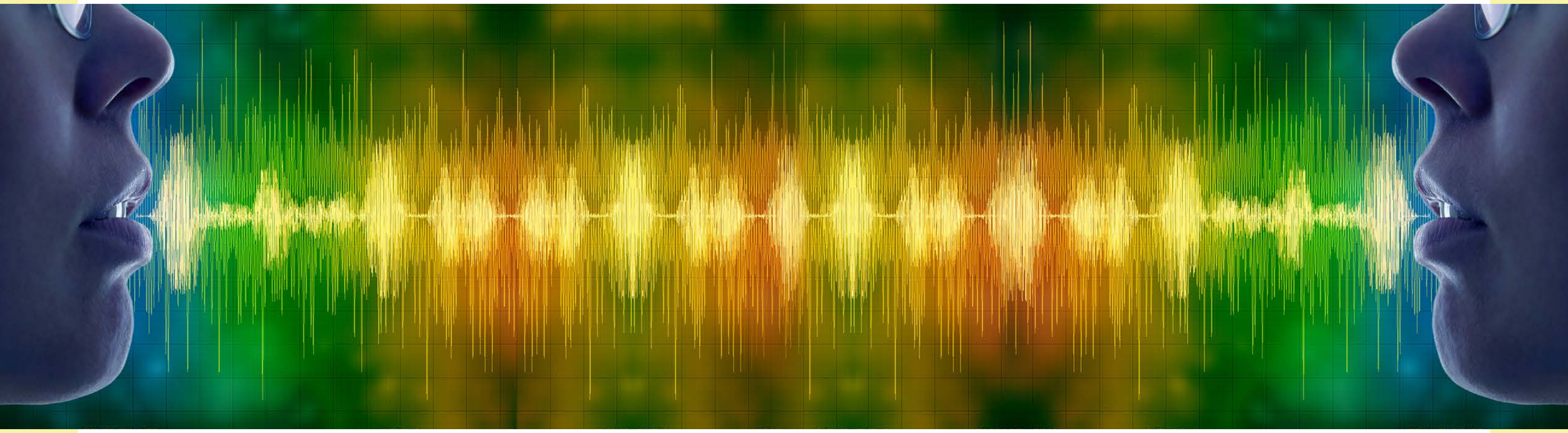
RESEARCH BRIEF

DECEPTION

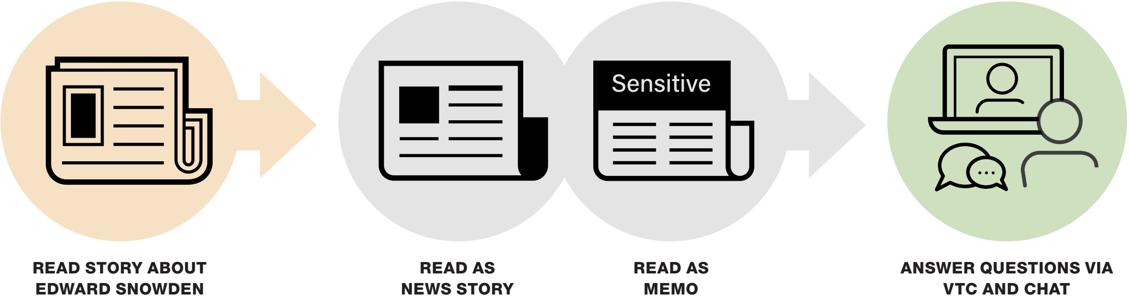


DETECTION

A group of RAND Corporation researchers found that machine-learning (ML) models can identify signs of deception during national security background check interviews. The most accurate approach for detecting deception is an ML model that counts the number of times that interviewees use common words.

How the experiment worked

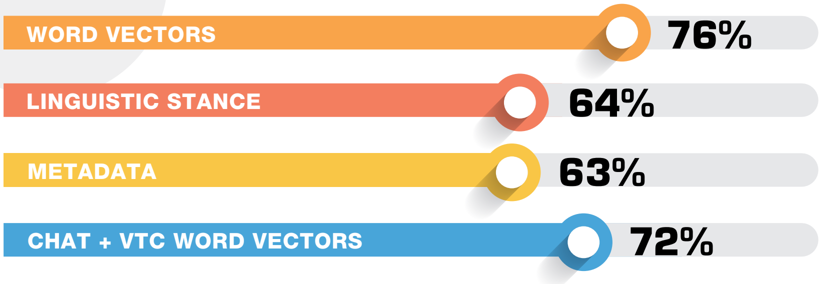
- The 103 participants read a story about Edward Snowden leaking classified information from the National Security Agency in 2013.
- Participants were randomly assigned to read the same story, but the story was presented as either a news report or a memo with markings indicating it contained sensitive information.
- Participants were assigned to one of two groups for interviews: One group was told to lie about what they had read, and the other was told to tell the truth.
- Former law enforcement officers interviewed participants via video teleconference (VTC) and text-based chat in randomized order.
- RAND researchers used the interview and chat transcripts to train several ML models to see whether the models could distinguish the liars from the truth-tellers.


READ STORY ABOUT EDWARD SNOWDEN
READ AS NEWS STORY
Sensitive
READ AS MEMO
ANSWER QUESTIONS VIA VTC AND CHAT

Four models were tested to see how accurately they detected deception

- **Word vectors:** This model counted the number of times the interviewees used common words.
- **Linguistic stance:** This model tested the use of stance vectors—sorting words into categories.
- **Metadata:** This model examined other meta-level features of speech, such as speaking cadence, average unique word length, and the fraction of words with more than six letters.
- **Chat and VTC word vectors:** This model looked at word vectors in the transcripts of interviews done via text-based chat and VTC.

How accurate were the models overall?



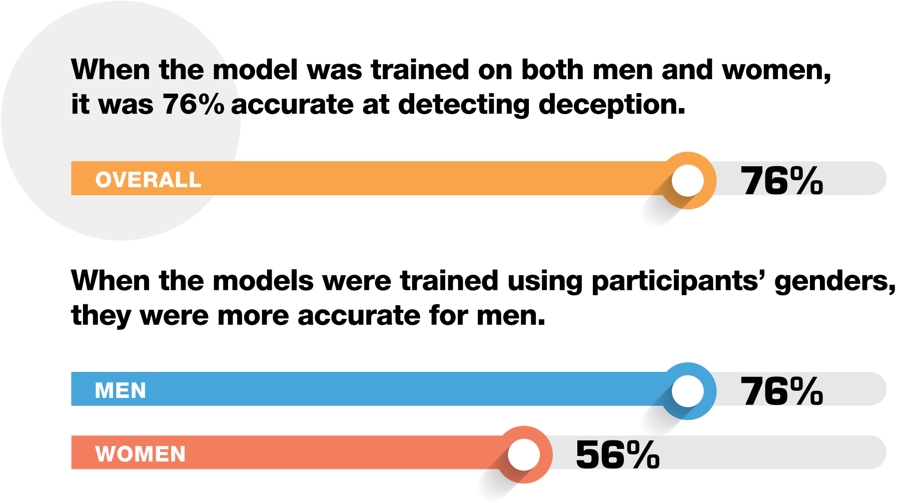
Three Big Takeaways

1. It is not only what one says but how one says it. Word frequency, speaking cadence, word choice, and other linguistic signals of potential lies.
2. ML models can detect signals of deceit in how people express themselves—even in text-based chats without the presence of a human interviewer.
3. The models are tools that can complement existing interview techniques; they cannot completely replace such techniques.

Bias can make the models less accurate for certain groups

All but one model detected deceit more accurately among men than women, which suggests that there was gender bias in the model outputs. If gender bias exists, there could be other biases, such as race and ethnicity. The following graphs show one example of the difference in accuracy by gender—in this case, for the word vector model.

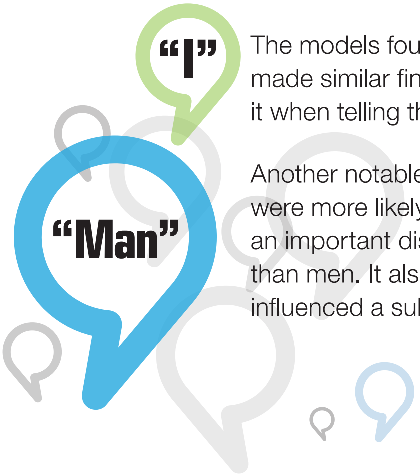
When the model was trained on both men and women, it was 76 percent accurate at detecting deception. When the models were trained using participants' genders, they were more accurate for men.



When the model was trained on both men and women, it was 76% accurate at detecting deception.

When the models were trained using participants' genders, they were more accurate for men.

Other Differences Between Men and Women



The models found that men and women used different words to deceive (other researchers have made similar findings). Men were less likely to use the word *I* when lying and more likely to use it when telling the truth.

Another notable speech pattern—although it had no link to lying or truth-telling—was that women were more likely than men to use the word *man* (for example, to describe Edward Snowden). This is an important distinction because it suggests that the subject's gender was more salient for women than men. It also raises the question of whether the subject's or interviewer's gender could have influenced a subject's responses to interview questions.

Implications for National Security

- Men conduct many of the background investigations for security clearances, and at least one-quarter of security clearance applicants are women. It is important to understand how the interviewer's gender could affect modeling results.
- Using ML tools inappropriately could lead to inequities in acceptance and rejection rates of applicants for security clearances.
- Because of potential biases in ML model outputs—and human beings—it is important to maintain a system of checks and balances that includes both people and machines.

Findings and Recommendations

Linguistic Tools Work

Finding: There are linguistic signals when people try to deceive, and ML tools can detect some of these signals for interviews conducted via VTC and chat.

Recommendation: The federal government should test ML modeling of interview data that uses word vectors to identify attempts at deception.

Interviewing Alternatives

Finding: Accuracy rates for detecting deception were similar for interviews conducted over VTC and chat.

Recommendation: The federal government should test VTC and chat-based interviews as alternatives to in-person security clearance interviews for certain cases.

Recommendation: The federal government should test chat-based questionnaires without a human interviewer present to augment existing in-person interviews. Investigators could use the data from chat *and* a face-to-face (or virtual) interview to identify concerns that merit further investigation.

Transcription Workflow

Finding: This exploratory study developed a workflow for transcribing and analyzing interviews that were audio recorded: (1) Amazon Web Services Transcribe used ML to automatically transcribe each interview recording, (2) a member of the research team listened to the audio and corrected the transcripts, and (3) the models analyzed the corrected transcripts.

Recommendation: The federal government should use ML tools to augment existing investigation processes by conducting additional analysis on pilot data, but the government should not replace existing techniques with these tools until they are sufficiently validated.

Accuracy Varies by Gender

Finding: Most models were more accurate for men than women. This finding suggests that other sources of biases could also exist.

Recommendation: The federal government should validate any ML models that it uses for security clearance investigations to limit bias on the basis of ascribed characteristics (e.g., race, gender, age) of interviewees.

Recommendation: The government should have a human-in-the-loop to continuously calibrate any ML models used to detect deception during the security clearance investigation process.

This research brief describes research conducted within the Forces and Resources Center of the RAND National Security Research Division (NSRD) and documented in *Looking for Lies: An Exploratory Analysis for Automated Detection of Deception*, by Marek N. Posard, Christian Johnson, Julia L. Melin, Emily Ellinger, and Hilary Reininger, RR-A873-1, 2022 (available at www.rand.org/t/RRA873-1). To view this brief online, visit www.rand.org/t/RBA873-1. The RAND Corporation is a research organization that develops solutions to public policy challenges to help make communities throughout the world safer and more secure, healthier and more prosperous. RAND is nonprofit, nonpartisan, and committed to the public interest. RAND's publications do not necessarily reflect the opinions of its research clients and sponsors. **RAND**® is a registered trademark.