How Humans and AI Detect Lies: What the Research Shows

Name Date Class

Reading

1 Judging the **veracity** of a statement matters in courtrooms, border interviews, workplaces, and online. For centuries, people have relied on **intuition** and interview technique when assessing another person's credibility. More recently, software that analyzes language has been offered as an alternative. Comparing the two is harder than it sounds, because most studies define "better" as accuracy on staged tasks in which some participants were told to lie.

2 Human judgment usually begins with behavior. In a study carried out in 75 countries, the most widespread **stereotype** was that liars avert their gaze. Yet a meta-analysis of 158 possible cues by Bella DePaulo and colleagues found that many behaviors showed no discernible link, or only a weak one, to deception. The cues that did emerge were modest: liars were somewhat less forthcoming and more tense. Accuracy is correspondingly low. In a 2006 analysis of 206 documents and 24,483 judges, Charles Bond and DePaulo found an average of 54 percent correct. Judges identified 61 percent of truths but only 47 percent of lies, which suggests that people are **susceptible** to a **pervasive** truth bias: the tendency to assume honesty. Judges were also more accurate with audible lies than with visible ones.

3 Better human methods exist, but they demand discipline. Interviewers can ask unexpected questions or make the task harder, which raises the cognitive load on liars. In nine studies published in 2023, Bruno Verschuere and colleagues found that people judging statements by intuition performed at chance, whereas people told to rely on one **heuristic**, the level of detail, especially detail that investigators can **corroborate**, could **discriminate** lies from truths. The polygraph measures physiology rather than judgment. A 2003 review by the US National Research Council concluded that specific-incident tests separate liars from truth-tellers far better than chance but far from perfectly, and it could not confirm that they survive a **countermeasure**, such as a trained attempt to beat the test.

4 Automated approaches mostly analyze language, because words can be counted and compared. A meta-analysis of 44 studies by Valerie Hauch and colleagues found that software could detect several patterns: relative to truth-tellers, liars used fewer sensory and perceptual words, expressed more negative emotion, and distanced themselves more from events. Machine-learning classifiers are trained on thousands of statements, some honest and some written by participants asked to **fabricate** an account. Recent work fine-tunes large language models in the same way. These systems do not detect lies as such. They learn statistical differences between the honest and invented texts they have seen.

5 In head-to-head tests, algorithms often **outperform** people. Bennett Kleinberg and Verschuere collected 1,640 true and false statements about people's intentions, and a model trained on them reached 69 percent accuracy. When human judges could overrule the model, accuracy fell to chance, so in that study human input did **impair** the machine's performance. The authors suggest a **plausible** explanation: truth bias. A team led by Alicia von Schenk built a classifier that scored 66.86 percent, against 46.47 percent for human judges, in an experiment with 2,040 participants. Only 31.76 percent asked for its advice, but those who did tended to follow its accusations. In Italy, Riccardo Loconte and colleagues fine-tuned a language model that reached roughly 71 to 83 percent on the kind of statement it was trained on.

6 The results come with an important **caveat**. In the language-model study, accuracy fell to between about 49 and 56 percent on a type of statement the model had not seen, so its skill did not **generalize**. Most datasets contain low-stakes, invented stories, unlike real lies, which are often embedded in true accounts. Base rates matter too: when few people lie, even an accurate test produces many false accusations. For a hypothetical screening of 10,000 employees including 10 spies, the National Research Council calculated that a polygraph set to catch eight spies would fail about 1,600 innocent people. A 2025 preprint by Emilio Barkett and colleagues found that several advanced language models judged truthful statements well but deceptive ones poorly. The EU AI Act classes polygraph-type tools in migration and border control as high-risk, with the obligations applying from December 2027 under a 2026 amendment, and it has banned emotion inference in workplaces and schools since February 2025.

7 So which is better? In the laboratory comparisons reviewed here, algorithms beat intuitive human judgment on narrow, text-based tasks, scoring roughly 67 to 83 percent on familiar kinds of statement where human judges scored near or below chance. Whether they do so with real suspects, real stakes, and unfamiliar liars has not been shown in these studies, and humans with a simple rule can do better than chance. The evidence therefore supports a limited claim: machines are better in specific settings, and no method has been shown to be reliable enough to justify accusing someone on its own.

Vocabulary Matching

Match each word from the reading with its meaning. Write the letter on the line.

1	**veracity**	_____	**A**	understanding reached without conscious reasoning
2	**intuition**	_____	**B**	seeming reasonable and likely to be true
3	**stereotype**	_____	**C**	a warning that limits or qualifies a claim
4	**susceptible**	_____	**D**	an action taken to defeat a system
5	**pervasive**	_____	**E**	tell the difference between two things
6	**heuristic**	_____	**F**	apply a finding to other cases or settings
7	**corroborate**	_____	**G**	spreading widely through a place or group
8	**discriminate**	_____	**H**	a fixed and oversimplified idea about a group
9	**countermeasure**	_____	**I**	support a claim with independent evidence
10	**fabricate**	_____	**J**	do better than something or someone else
11	**outperform**	_____	**K**	weaken or damage the quality of something
12	**impair**	_____	**L**	easily affected or influenced by something
13	**plausible**	_____	**M**	a simple rule that helps you decide quickly
14	**caveat**	_____	**N**	the quality of being true or accurate
15	**generalize**	_____	**O**	invent something false and present it as true

Fill in the Blank

Choose a word from the box to complete each sentence. Use each word once. The words are already in the correct form.

caveat	**corroborated**	**discriminate**	**fabricate**	**generalize**	**heuristic**	**impair**	**intuition**	**outperform**
pervasive	**plausible**	**stereotype**	**susceptible**	**veracity**				

1 Journalists check the _____ of a claim before publishing it.

2 She had no evidence, only a strong _____ that the deal was a mistake.

3 The film relied on a _____ about teenagers.

4 Tired drivers are _____ to mistakes.

5 Corruption was _____ in the city government.

6 A simple _____ , such as choosing the cheapest option, saves time.

7 Two independent witnesses _____ her story.

8 Trained tasters can _____ between very similar wines.

9 He was fired for trying to _____ evidence.

10 The new model can _____ the old one on every test.

11 Lack of sleep can _____ your judgment.

12 The excuse sounded _____ , so nobody questioned it.

13 The study is interesting, but one _____ is its small sample.

14 It is risky to _____ from a sample of ten people.

Comprehension Questions

Answer in your notebook or discuss with a partner. Questions 10 to 12 are IELTS-style.

1 Why does the reading say that comparing human and AI lie detection is harder than it sounds?

2 Summarize what the research says about behavioral cues to deception and about human accuracy. How does the truth bias affect the results?

3 What advantages do the methods in the third paragraph offer over intuition, and what limits does the reading note for the polygraph?

4 How do the automated systems in the reading work, and why does the reading say they do not detect lies as such?

5 Compare the accuracy figures reported for humans and for algorithms in the fifth paragraph. What do the results suggest about combining human judgment with a machine?

6 Why does the reading treat the algorithms' higher accuracy with caution? Give at least three reasons from the sixth paragraph.

7 Explain why base rates matter when an accurate test is used to identify liars in a population where few people lie.

8 What do the EU AI Act provisions mentioned in the reading suggest about how regulators view lie-detection technology?

9 Evaluate the writer's final answer to the question of which is better. Which parts of the conclusion are firmly supported, and which are not?

IELTS-Style Questions

10 **True, False, or Not Given:** "*Trained police officers detect lies more accurately than ordinary people.*" Explain your answer using evidence from the reading.

11 **Yes, No, or Not Given:** "*The writer believes that algorithms have been shown to detect lies better than humans in all settings.*" Explain your answer using evidence from the reading.

12 **Author's purpose:** Why does the writer report both the algorithms' high accuracy and their failure on unseen statements? What effect does this have on the reader's conclusion?

True or False

Circle T (true) or F (false) according to the reading. Write a correction for each false statement on the dotted line.

1 **T / F** Most studies define better as accuracy on staged tasks in which some participants were told to lie.

2 **T / F** DePaulo and colleagues found strong links to deception for most of the 158 cues.

3 **T / F** Bond and DePaulo found that judges identified 61 percent of truths and 47 percent of lies.

4 **T / F** Verschuere and colleagues found that intuition worked better than relying on one cue.

5 **T / F** The National Research Council could not confirm that polygraph results survive a countermeasure.

6 **T / F** The reading says AI systems detect lies as such.

7 **T / F** The Kleinberg and Verschuere model reached 69 percent accuracy.

8 **T / F** Most participants in the von Schenk experiment asked for the classifier's advice.

9 **T / F** The EU AI Act has banned emotion inference in workplaces and schools since February 2025.

10 **T / F** The reading concludes that machines are reliably better than humans with real suspects.

Discussion Questions

Talk with a partner or in a small group. The questions move from personal to more abstract.

1 How good do you think you are at telling when someone is lying? What do you look for, and how sure are you that those signs are reliable?

2 The most widespread stereotype about liars is that they avert their gaze, yet the research finds weak links. Why might a false belief like this stay popular?

3 People are more accurate with audible lies than visible ones, according to the reading. What might explain that?

4 The truth bias makes people assume honesty. In everyday life, is this bias a weakness or a useful social habit?

5 Verschuere and colleagues found that relying on the level of detail helped people. Would you trust a simple rule like this in a real situation? Why or why not?

6 The polygraph is still used in some places. Should governments or employers use tests that are far from perfect? Under what conditions, if any?

7 Algorithms in the studies scored higher than humans, but mostly on staged, low-stakes tasks. How much should these results influence real decisions?

8 When human judges could overrule the model, accuracy fell to chance. What does this suggest about the role of human oversight in automated decisions?

9 Only about a third of participants asked for the classifier's advice, and those who did tended to follow its accusations. What might this suggest about how people use AI?

10 If very few people lie in a given setting, even an accurate test produces many false accusations. How should a society weigh the cost of a false accusation against the benefit of catching a liar?

11 Should employers or schools be allowed to use tools that infer emotion or honesty? What are the arguments for and against the EU's approach?

12 Liars could learn to fool a language-based detector. How might that change how the technology is used, and who might benefit?

13 Is a lie detector that is right two times out of three useful? What would you need to know before using it?

14 Which do you trust more to judge honesty in a high-stakes case: a trained person, a machine, or both together? What evidence would change your mind?

Language Focus: Comparatives with Degree Modifiers

Comparative forms (better, more accurate, less forthcoming, fewer words) show the direction of a difference. A modifier before the comparative (far, much, considerably, somewhat, slightly) shows its size, and "than" names what is being compared.

In the reading:

- "…specific-incident tests separate liars from truth-tellers **far better than chance**…"
- "…liars were **somewhat less forthcoming and more tense**."
- "Judges were also **more accurate with audible lies than with visible ones**."

Practice: Combine into one sentence with a comparative and a modifier that shows the size of the difference: *"The model scored 69 percent. The human judges scored 50 percent."*

Activities and speed challenge
englishbridges.com/lessons/human-ai-lie-detection

Notes