INDEPENDENT THINKING ABOUT ARTIFICIAL INTELLIGENCE

BY ADRIAN DUNKLEY

AI BOSS

THE DAILY AI FLUENCY NEWSPAPER

MONDAY 9 NOVEMBER 2026 ISSUE 008 AIBOSSHQ.COM

AI SECURITY

When a document tries to steer an AI summary

A 2023 experiment showed how instructions embedded in a page could change what a search assistant left out. Try an altered document before relying on an automated summary.



The Inserted Strip
AI-generated artwork · AI Boss Ltd

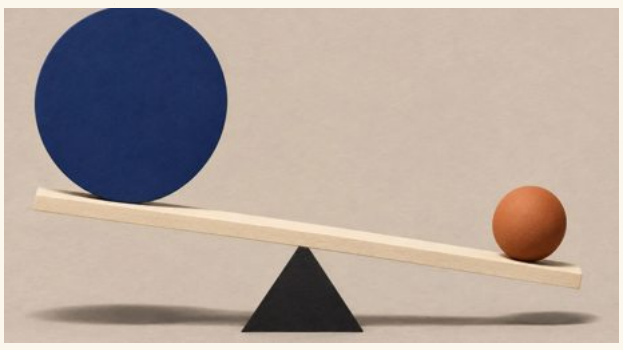
If a test summary leaves out a failed lift inspection, pause the use of automated summaries to close jobs.

Give retrieved passages a separate label from application instructions and repeat the test after changes to the model or document pipeline. Have staff check the contractor’s original report before closing the job.

Continue reading at aibosshq.com

HUMAN JUDGEMENT

Check whether an AI label changes what staff approve



The Wrong Weight
AI-generated artwork · AI Boss Ltd

In a chest X-ray experiment, radiologists gave human-written advice lower ratings when it carried an AI label. The label made no clear difference to diagnostic accuracy.

Keep the advice and its timing fixed. Compare what people approve when the stated source changes.

Continue reading at aibosshq.com



William Warby / Unsplash · Illustrative photograph
Unsplash Licence

AI AT WORK

Time the coding task beyond the first patch

Log the minutes staff work on a change separately from the time it takes to clear review and tests.

Continue reading at aibosshq.com



Erik Mclean / Unsplash · Illustrative photograph
Unsplash Licence

DIGITAL SAFETY

Call a known number before sending emergency money

If an organisation is named, use its contact details from a statement you already hold or an official site.

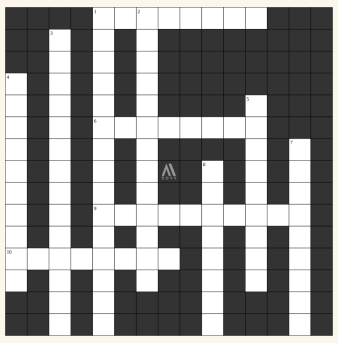
Continue reading at aibosshq.com

THE CROSSWORD DESK

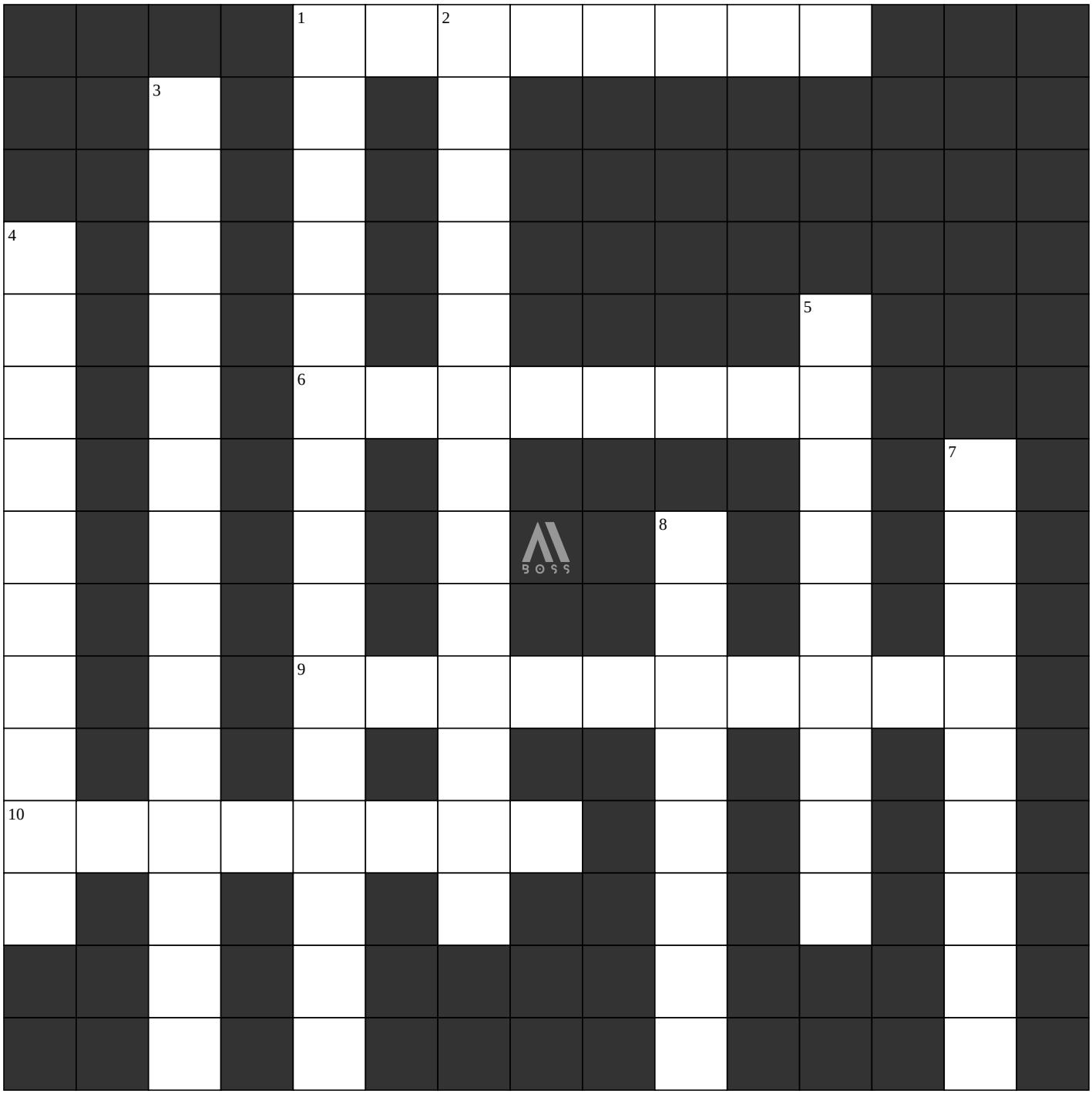
How well do you know your AI?

Practise AI vocabulary and judgement through workplace clues. Play the crossword or fill the grid with Quiz Mode at **aibosshq.com/games/daily/2026-11-09/**.

The printable puzzle follows on page 2. Its solution is on page 3.



Article sources and illustration credits appear with the full essays.



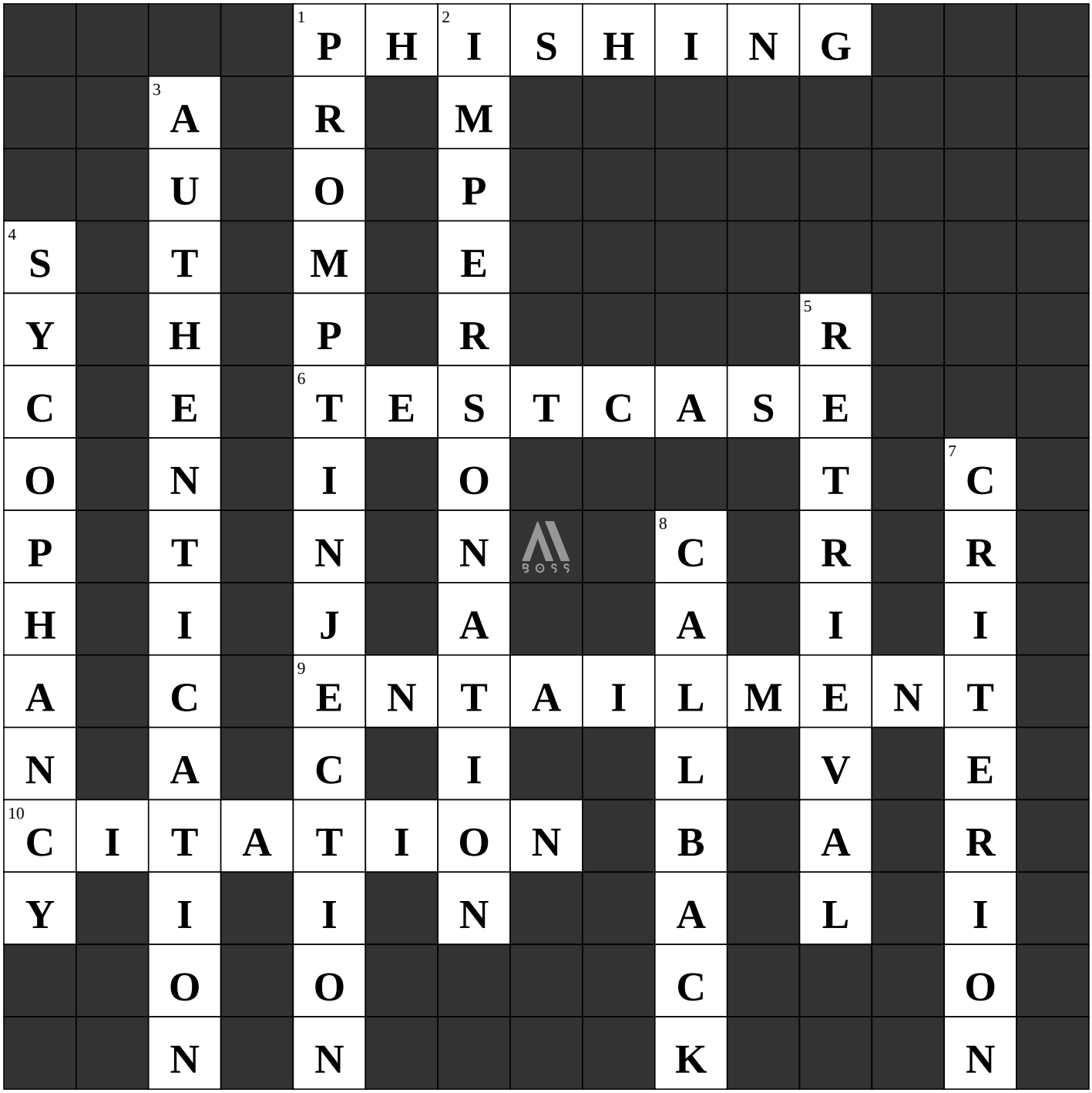
ACROSS

1. Fake messages seeking sensitive information. (8)
6. Test input paired with an expected outcome. (8)
9. Relation where one claim follows from another. (10)
10. Reference that points readers to a source. (8)

DOWN

1. Planted instruction that tries to redirect AI. (15)
2. Pretending to be another person or organisation. (13)
3. Checking a claimed identity at sign-in. (14)
4. Undue agreement with a user's stated view. (10)
5. Finding and fetching existing information. (9)
7. Standard used to judge a result. (9)
8. Return phone call used to confirm a request. (8)

Solution


1 P H 2 I S H I N G
3 A R M
U O P
4 S T M E
Y H P R 5 R
C E 6 T E S T C A S E
O N I O T 7 C
P T N N 8 C R R
H I J A A I I
A C 9 E N T A I L M E N T
N A C I L V E
10 C I T A T I O N B A R
Y I I N A L I
O O C O
N N K N