

GenAI Misinformation, Trust, and News Consumption: Evidence from a Field Experiment

Filipe Campante
JHU & NBER

Ruben Durante
NUS/UPF

Felix Hagemeister
Süddeutsche Zeitung

Ananya Sen
CMU

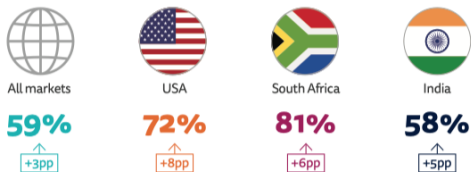
NBER POL Summer Institute
July 2026

Misinformation and trust

- Online misinformation viewed as a major societal problem
 - Reduces trust in news and in institutions ([Brennan 2024](#), [Zhuravskaya et al. 2020](#))
 - Increases polarization ([Lazer et al. 2018](#), [Tucker et al. 2018](#), [Jerit-Zhao 2020](#))
 - Potential negative effects on players in the platform ecosystem ([Ahmad et al. 2024](#))

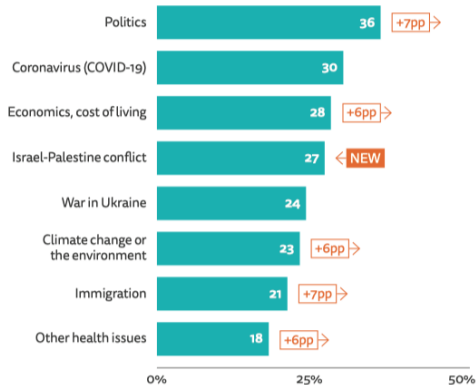
Misinformation and trust (cont.)

PROPORTION CONCERNED ABOUT WHAT IS REAL AND WHAT IS FAKE WHEN IT COMES TO NEWS ONLINE



Q_FAKE_NEWS_1. Please indicate your level of agreement with the following statement. 'Thinking about online news, I am concerned about what is real and what is fake on the internet.' Base: Total sample in all markets = 94,943, USA = 2023, South Africa = 2013, India = 2016.

PROPORTION THAT SAY THEY HAVE SEEN FALSE OR MISLEADING INFORMATION ABOUT EACH TOPIC IN THE LAST WEEK – ALL MARKETS



Q_FAKE_NEWS_2021a. Have you seen false or misleading information about any of the following topics, in the last week? Please select all that apply. Base: Total sample in all markets = 94,943.

AI enters the picture

- Generative AI (GenAI) makes it easier to produce visual and audio content virtually indistinguishable from authentic one (“deep fakes”)
- Can increase misinformation and further destroy trust in the media
(Vaccari-Chadwick 2020, Spitale et al. 2023, Toff and Simon, 2025)

AI enters the picture

- Generative AI (GenAI) makes it easier to produce visual and audio content virtually indistinguishable from authentic one (“deep fakes”)
- Can increase misinformation and further destroy trust in the media
(Vaccari-Chadwick 2020, Spitale et al. 2023, Toff and Simon, 2025)
- Columbia Journalism Review’s awareness campaign on visual misinformation:
 - 34 million AI-generated images daily, reach 80% of readers
 - 76% of U.S. news users failed to spot AI-generated images

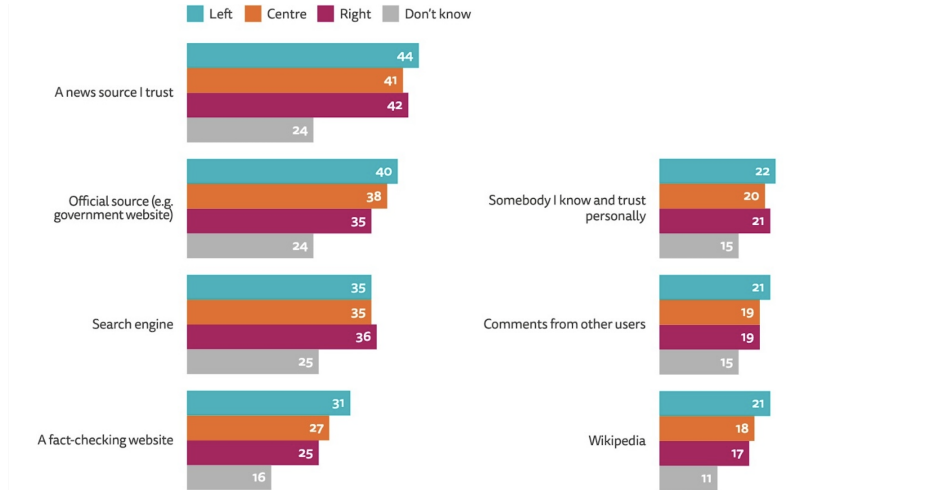
What implications for the news industry?

- If impossible to tell apart true from false, willingness to pay for news could fall
- High-quality news - being more expensive to produce - may no longer be viable

What implications for the news industry?

- If impossible to tell apart true from false, willingness to pay for news could fall
- High-quality news - being more expensive to produce - may no longer be viable
- But could **lower trust** in online information **increase the value of trustworthiness**?
- Could trusted news organizations **benefit** from this situation?
 - Especially if they can help readers distinguish fake content from real one?

Who would you turn to to check an online news that you suspect may be false?



Source: Reuters Institute Digital News Report 2025

What implications for the news industry?

- If impossible to tell apart true from false, willingness to pay for news could fall
- High-quality news - being more expensive to produce - may no longer be viable
- But could **lower trust** in online information **increase the value of trustworthiness**?
- Could trusted news organizations **benefit** from this situation?
 - Especially if they can help readers distinguish fake content from real one?
- Potential business opportunity (Opdahl et al 2023)

VIDEO



Visual Investigations

The Visual Investigations team combines traditional reporting with digital sleuthing and the forensic analysis of visual evidence to find truth, hold the powerful to account and deconstruct important news events. Sign up for [our e-mail newsletter](#).

[Send us a tip.](#) Make sure to include "Visual Investigations" in your message.

Israel-Gaza Conflict



How a Single Family Was Shot Dead on a Street in Gaza



How the Israeli Hostage Rescue Led to One of Gaza's Deadliest Days



Israeli Bulldozers Flatten Mile After Mile in the West Bank



2,000-Pound Bombs Likely Used in Al-Mawasi Strike



'We're Aware of the Location': Aid Groups in Gaza Coordinated With I.D.F. but Still Came Under Fire

[More in Israel-Gaza Conflict](#)

Russia-Ukraine War



Caught on Camera, Traced by Phone: The Russian Military Unit That Killed Dozens in Bucha



Their Final Moments: Victims of a Russian Atrocity in Bucha



'Putin is a Fool': Intercepted Calls Reveal Russian Army in Disarray



New Evidence Shows How Russian Soldiers Executed Men in Bucha



Under Fire, Out of Fuel: What Intercepted Russian Radio Chatter Reveals

[More in Russia-Ukraine War](#)

Award Winning Visual Investigations



How a U.S. Drone Strike Killed the Wrong Person



An Israeli Soldier Killed a Medic in Gaza. We Investigated the Fatal Shot.



One Building, Dozens Killed in Syria: How Bashar al-Assad Gassed His Own People



Killing Khawaboggi: How a Brutal Saudi Hit Job Unfolded



10 Minutes, 12 Gunfire Bursts, 30 Videos. Mapping the Las Vegas Massacre.

[More in Award Winning Visual Investigations](#)

What implications for the news industry?

- If impossible to tell apart true from false, willingness to pay for news could fall
- High-quality news - being more expensive to produce - may no longer be viable
- But could **lower trust** in online information **increase the value of trustworthiness**?
- Could trusted news organizations **benefit** from this situation?
 - Especially if they can help readers distinguish fake content from real one?
- Potential business opportunity (Opdahl et al 2023)
- **Broader point:** effect of AI on news consumption may be much more nuanced

This paper

Questions

1. Does highlighting the difficulty to distinguish between AI-generated and real content affect people's attitudes towards the news?
2. Can this make them *more* likely to engage with and subscribe to trusted news sources?

This paper

Questions

1. Does highlighting the difficulty to distinguish between AI-generated and real content affect people's attitudes towards the news?
2. Can this make them *more* likely to engage with and subscribe to trusted news sources?

Methodology

- Simple theoretical framework
- Field experiment with major German newspaper (Süddeutsche Zeitung, SZ)
- Treatment: quiz to tell apart AI-generated pictures from real ones
- Survey responses (trust, concern for misinformation) and behavioral outcomes (browsing behavior, subscription)

Theoretical framework in a nutshell

- **Core idea:**

- Misinformation reduces overall trust, including in the trustworthy outlet
- But it may increase demand for trustworthy outlets

Theoretical framework in a nutshell

- **Core idea:**
 - Misinformation reduces overall trust, including in the trustworthy outlet
 - But it may increase demand for trustworthy outlets
- Consumers seek accurate signals; the demand for an outlet is decreasing in the variance of its signal (σ_i)
- GenAI increases the variance of the signals of all outlets by the same amount (Δ) $\Rightarrow$ demand for quality outlets goes down

Theoretical framework in a nutshell

- **Core idea:**

- Misinformation reduces overall trust, including in the trustworthy outlet
- But it may increase demand for trustworthy outlets
- Consumers seek accurate signals; the demand for an outlet is decreasing in the variance of its signal (σ_i)
- GenAI increases the variance of the signals of all outlets by the same amount (Δ) $\Rightarrow$ demand for quality outlets goes down
- **Trustworthy** outlets can mitigate the effect of the shock ($\alpha\Delta$, with $0 < \alpha < 1$)
- As misinformation increases, demand for (sufficiently) trustworthy sources grows
- **Implication:** trustworthy sources become more valuable even though trust in their content falls, because the alternatives are degrading more rapidly

Our partner: Süddeutsche Zeitung

- Largest non-tabloid German newspaper
(260K daily copies + 285K online subscribers in 2024)
- Based in Munich
- Very respected news organization
(e.g., broke the “Panama Papers” and “Paradise Papers”)
- Online team regularly runs experiments with subscribers
(reached via email) and website visitors (pop-up ads)



Süddeutsche Zeitung Abo mit Prämie

Our partner: Süddeutsche Zeitung (cont.)



“The truth cannot be generated. Only investigated.”

Treatment and control

- **Treatment: AI awareness quiz**

- Show three pairs of pictures of which one is AI-generated and one is real
- Ask respondents to indicate which one is real and which one is not
- Goal: increase awareness of the difficulty to tell apart AI-generated from real content

Treatment and control

- **Treatment: AI awareness quiz**

- Show three pairs of pictures of which one is AI-generated and one is real
- Ask respondents to indicate which one is real and which one is not
- Goal: increase awareness of the difficulty to tell apart AI-generated from real content

- **Control: current affairs knowledge quiz**

- Show three pairs of pictures: real ones from treatment + original real pictures modified for AI quiz
- Ask respondents to indicate in what countries the pictures were taken
- Goal: mimic daily news consumption but have them engage with pictures and take quiz

Treatment and control

- **Treatment: AI awareness quiz**

- Show three pairs of pictures of which one is AI-generated and one is real
- Ask respondents to indicate which one is real and which one is not
- Goal: increase awareness of the difficulty to tell apart AI-generated from real content

- **Control: current affairs knowledge quiz**

- Show three pairs of pictures: real ones from treatment + original real pictures modified for AI quiz
- Ask respondents to indicate in what countries the pictures were taken
- Goal: mimic daily news consumption but have them engage with pictures and take quiz

- **Alternative control: non quiz-takers (as benchmark)**

Treatment: AI awareness



Which image was generated with AI?

- Left
- Right
- Both
- None

Treatment: AI awareness



Which image was generated with AI?

- Left
- Right
- Both
- None

Treatment: AI awareness (cont.)



Which image was generated with AI?

- Left
- Right
- Both
- None

Treatment: AI awareness (cont.)



Which image was generated with AI?

- Left
- **Right**
- Both
- None

Treatment: AI awareness (cont.)



Which image was generated with AI?

- Left
- Right
- Both
- None

Treatment: AI awareness (cont.)



Which image was generated with AI?

- Left
- Right
- Both
- None

Treatment: AI awareness



Which image was generated with AI?

- Left
- Right
- Both
- None

Control: current affairs



In which countries were these images taken in 2024?

- Germany and Spain
- Belgium and Spain
- France and Belgium
- Germany and Belgium

Treatment: AI awareness (cont.)



Which image was generated with AI?

- Left
- **Right**
- Both
- None

Control: current affairs (cont.)



In which countries were these images taken?

- **America and Germany**
- Canada and France
- America and Italy
- Canada and Germany

Treatment: AI awareness (cont.)



Which image was generated with AI?

- Left
- Right
- Both
- None

Control: current affairs (cont.)



In which countries were these images taken?

- France and Georgia
- France and Germany
- Germany and Spain
- Italy and Finland

Samples, outcomes, and performance

- Sample:

- Pop-up window (shown to 1% of visitors) and email (sent to subscribers) between Feb 25 and Mar 9, 2025
- 17,199 users took the quiz with 16,195 subscribers and 1,004 non-subscribers
- For a subset of $\approx 6,000$ subjects we can link survey outcomes with platforms data

- Outcomes:

- Survey measures (for all): concern with misinformation, trust in the media, WTP
- Engagement (for matched sample): browsing behavior, subscription decisions

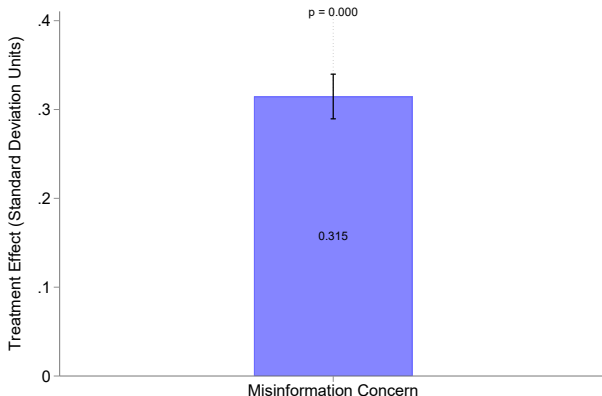
- Performance

- On average 28% of correct answers in T vs. 88.4% in C
- Similar time spent (4:20 vs. 3:54)
- Higher perceived difficulty in T (4.7/6) vs. C (1.52/6)

Balance checks

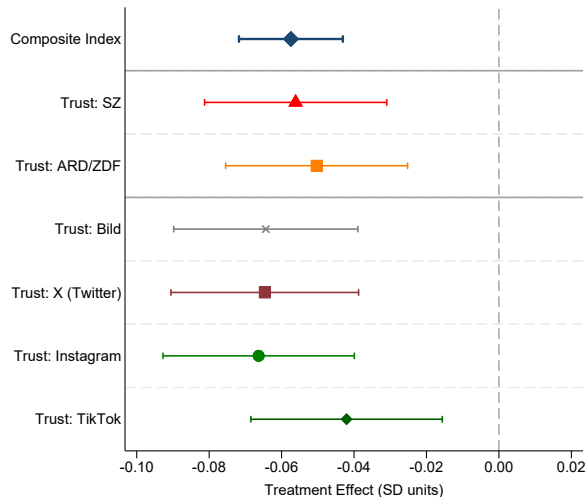
	Total Obs	Control			Treatment			Difference	
		Obs	Mean	s.d.	Obs	Mean	s.d.	T - C	p-value
Survey sample									
Female	17,199	8,674	0.46	0.50	8,525	0.45	0.50	-0.006	0.413
SZ subscriber	17,199	8,674	0.94	0.24	8,525	0.94	0.23	0.003	0.340
Tracked	17,199	8,674	0.35	0.48	8,525	0.34	0.47	-0.006	0.388
Early March	17,199	8,674	0.84	0.37	8,525	0.83	0.37	-0.007	0.229
Old (60+)	17,199	8,674	0.49	0.50	8,525	0.48	0.50	-0.011	0.150
Young (<40)	17,199	8,674	0.21	0.41	8,525	0.22	0.42	0.009	0.169
Tracked sample									
Female	5,915	3,010	0.40	0.49	2,905	0.40	0.49	0.001	0.929
Early March	5,915	3,010	0.88	0.33	2,905	0.87	0.33	-0.004	0.609
Old (60+)	5,915	3,010	0.52	0.50	2,905	0.53	0.50	0.008	0.537
Young (<40)	5,915	3,010	0.19	0.39	2,905	0.19	0.40	0.001	0.913
Pre-intervention Daily Visits	5,967	2,988	96.31	98.77	2,979	97.51	102.08	1.196	0.646
High Pre-intervention Affinity									
Politics	5,718	2,871	0.57	0.50	2,847	0.57	0.49	0.008	0.527
Sports	5,718	2,871	0.27	0.44	2,847	0.27	0.44	0.003	0.824
Music	5,718	2,871	0.18	0.38	2,847	0.18	0.38	0.001	0.909
Crime	5,718	2,871	0.39	0.49	2,847	0.40	0.49	0.012	0.351
Location									
Germany	5,585	2,860	0.80	0.40	2,725	0.79	0.41	-0.015	0.157
Bayern	5,585	2,860	0.15	0.36	2,725	0.14	0.35	-0.009	0.330
Nordrhein-Westfalen	5,585	2,860	0.13	0.34	2,725	0.13	0.34	-0.002	0.846
Schleswig-Holstein	5,585	2,860	0.11	0.31	2,725	0.10	0.30	-0.004	0.635
Rheinland-Pfalz	5,585	2,860	0.09	0.28	2,725	0.08	0.27	-0.006	0.420
Standard Monthly Subscription	5,915	3,010	0.36	0.48	2,905	0.36	0.48	-0.006	0.605

Results: survey-based outcomes



- Large increase in concern with misinformation (+0.31 st.dev).
- *“Nowadays, it is easy and deceptively realistic to create and spread disinformation. What do you think about that?”* 0 (not at all problematic) to 6 (extremely problematic)."

Results: survey-based outcomes



- Significant decrease in trust in all news outlets/platforms (-0.05 st. dev.) [SZ (2.91/3 in C), ARD/ZDF (2.8), Bild (0.82), IG (0.76), TikTok (0.26), X (0.25)]

▸ Regression Table

▸ Tracked Sample

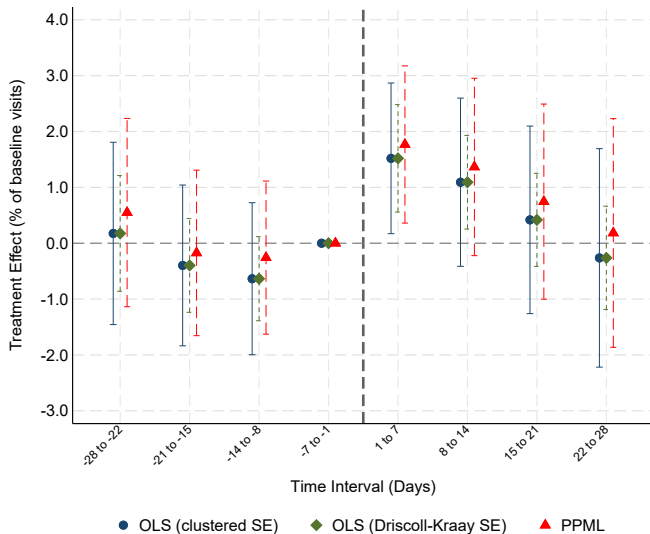
▸ Robustness: Controls

Daily visits to SZ pages (rolling-window diff-in-diff)

Dep. variable	(1)	(2)	(3)	(4)
	Number of daily visits			
Timespan	(-21;+3)	(-21;+5)	(-21;+10)	(-21;+14)
AI Treatment $\times$ Post	0.114** (0.052)	0.085** (0.042)	0.074** (0.035)	0.067* (0.034)
Post	0.192** (0.089)	0.202** (0.087)	0.208** (0.083)	0.215*** (0.082)
Sample mean	4.690 (5.534)	4.653 (5.493)	4.615 (5.457)	4.614 (5.463)
Individual FE	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes
Obs	143,208	155,142	184,977	208,845
Adjusted-R ²	0.768	0.768	0.766	0.765
F-stat	6.756	6.224	6.800	6.868

- Number of visits in 3 days after increased by 2.5% for T relative to C

Weekly visits to SZ pages: event study



Heterogeneity: Visits

Dep. variable	(1)	(2)	(3)	(4)
	Number of daily visits			
Subgroup	Low Politics	High Politics	Light Reader	Heavy Reader
AI Treatment $\times$ Post	0.1385** (0.0618)	0.0522 (0.0603)	0.0003 (0.0365)	0.2584** (0.1264)
Sample mean	4.030 (5.103)	5.514 (5.760)	0.347 (0.863)	11.480 (5.988)
Individual FE	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes
Observations	63,778	84,838	38,584	38,532
Adjusted-R ²	0.7738	0.7627	0.1512	0.5760
F statistic	2.914	5.441	0.075	2.459

- First stage: effect on misinformation concern ▶ 1
- (Pre-registered) Heterogeneity: perceived difficulty ▶ 2

Patterns are driven by hard news (politics + economy)

	(1)	(2)	(3)	(4)	(5)
Subgroup	Full	Low Politics	High Politics	Heavy Reader	Light Reader
AI Treatment $\times$ Post	0.3140 (0.1981)	0.6442** (0.2729)	0.1055 (0.2982)	1.0049* (0.5879)	-0.0499 (0.1309)
Sample mean	10.251 (20.286)	8.116 (18.400)	12.638 (21.929)	22.785 (29.296)	0.738 (4.075)
Individual FE	Yes	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes	Yes
Observations	155,142	63,778	84,838	38,532	38,584
Adjusted-R ²	0.5685	0.5791	0.5503	0.5473	0.1899
F statistic	3.594	2.786	4.630	2.605	0.450

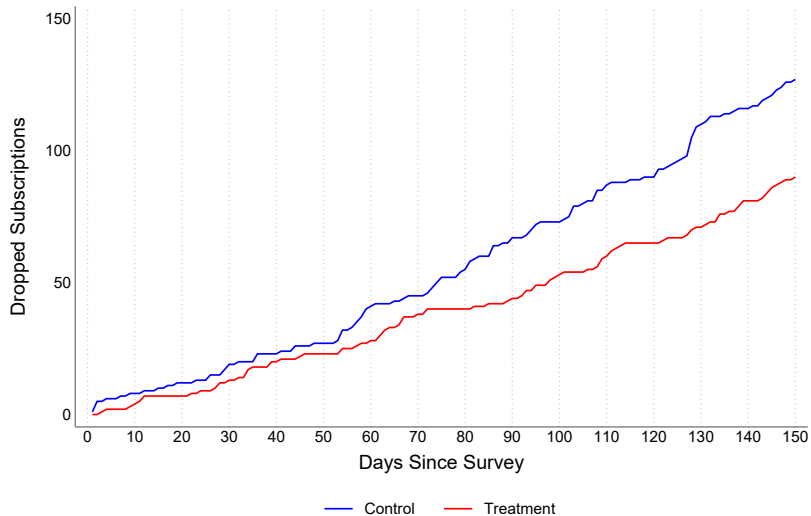
- No effect on visits for sports articles ▶ 1

Robustness

- Engagement

- Similar results excluding subjects who took surveys on SZ platform after treatment ▶ 1
- Effect driven by organic visits, not visits through social media ▶ 2
- Readers do not demand more AI-specific content nor does SZ offer more ▶ 3
- Robust to alternative control group (i.e., people who did not take the quiz) ▶ 4
- Robust to alternative models (i.e., Poisson regression) ▶ 5
- No difference between subjects in control group who found the quiz hard vs. easy

Post-intervention subscription status



Post-intervention subscription status

Dep. variable	(1)	(2)	(3)	(4)
	Subscription Status			
Months Post-Intervention	2 Months	3 Months	4 Months	5 Months
AI Treatment	0.0042 (0.0027)	0.0074** (0.0034)	0.0080** (0.0040)	0.0119** (0.0047)
Sample mean	0.989 (0.105)	0.982 (0.133)	0.975 (0.157)	0.965 (0.184)
Observations	6,158	6,158	6,158	6,158
R-squared	0.0004	0.0008	0.0007	0.0010
F-statistic	2.436	4.783	4.057	6.423

- Attrition rate 1/3 lower; retention rate after 5 months 1.1 pp higher
- Similar effect comparing treated with alternative control (non quiz-takers)

Subscription retention heterogeneity

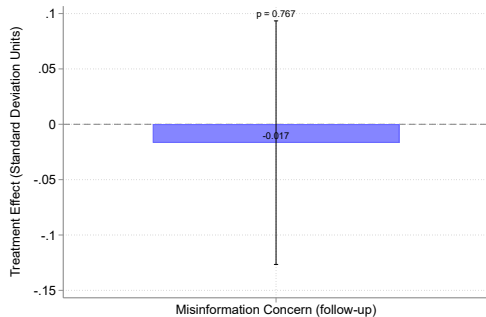
Dep. variable Subgroup	(1)	(2)	(3)	(4)
	Subscription status (5 Months)			
	Low Politics	High Politics	Heavy Reader	Light Reader
AI Treatment	0.0174** (0.0075)	0.0086 (0.0063)	0.0145* (0.0085)	0.0044 (0.0111)
Sample mean	0.964 (0.186)	0.967 (0.178)	0.972 (0.164)	0.952 (0.213)
Observations	2,433	3,234	1,477	1,472
R-squared	0.0022	0.0006	0.0019	0.0001
F statistic	5.409	1.877	2.882	0.157

Economic significance of results

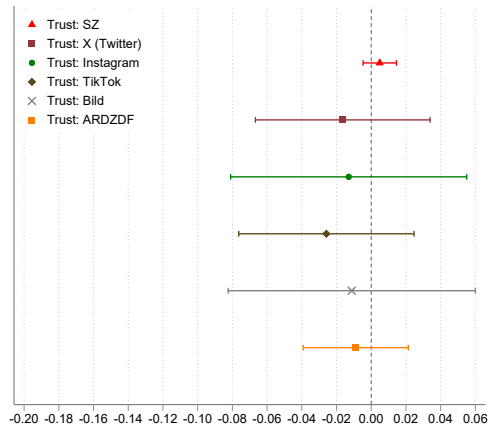
- Increased revenues from higher subscriber retention → €456,000 / year
- Additional ad revenue from higher engagement in the short run → 23,000 visits
- Additional ad revenue from extra traffic from retained subscribers
→ 3.3 million visits / year
- Large effects considering intervention was mild and that 80% of experiments in digital markets yield null results (Kohavi et al., 2013)

Long(er)-Run Outcomes

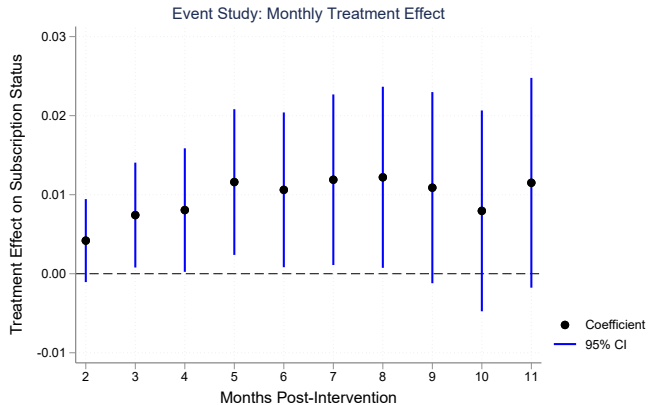
- Secondary Survey in Feb 2026 (Different sample, N=1,683)



► Regression Table



Cumulative effects: evolution of subscription retention

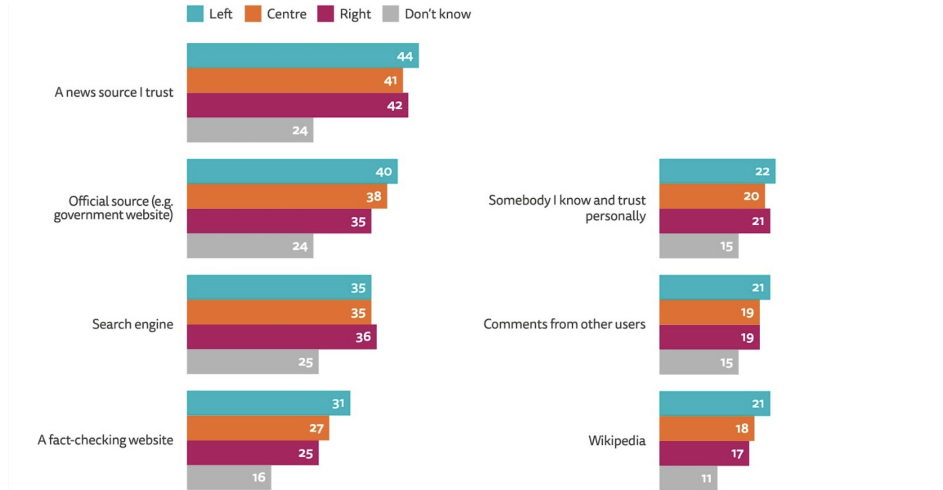


Note: Monthly treatment effects from Table 3. Each point shows the treatment effect for that specific month.

Takeaways

1. Increased AI awareness boosts concern about misinformation and reduces trust in news, including in the trusted source
2. Yet, it increases actual engagement with the trusted source
 - Higher engagement (short-run), higher retention (long-run)
3. Impact of AI on the news environment more nuanced than first thought:
 - AI misinformation need not lead to a collapse in trust
 - Potential business opportunity if news outlets can keep up with the challenge

Who would you turn to to check an online news that you suspect may be false?



Source: Reuters Institute Digital News Report 2025

Takeaways

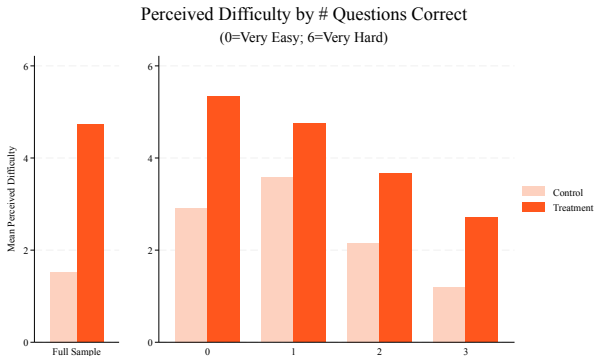
1. Increased AI awareness boosts concern about misinformation and reduces trust in news, including in the trusted source
2. Yet, it increases actual engagement with the trusted source
 - Higher engagement (short-run), higher retention (long-run)
3. Impact of AI on the news environment more nuanced than first thought:
 - AI misinformation need not lead to a collapse in trust
 - Potential business opportunity if news outlets can keep up with the challenge
4. Broader implications concerning the role of experts in the era of AI

THANK YOU!

Model extensions

- **Endogenous trustworthiness:** outlets choose α at convex cost $C(\alpha)$
 - Quality and trustworthiness are complements: higher-quality outlets invest more in mitigation
 - Rising misinformation generally increases investment in trustworthiness
- **Price-setting:** outlets optimally set prices given demand
 - Trustworthy outlets are able to charge higher prices *and* sell more
 - Less trustworthy outlets forced to charge lower prices
- **Upshot:** extensions reinforce the core mechanism — trustworthiness increases market power and, for sufficiently trustworthy outlets, an increase in misinformation ends up reinforcing said market power

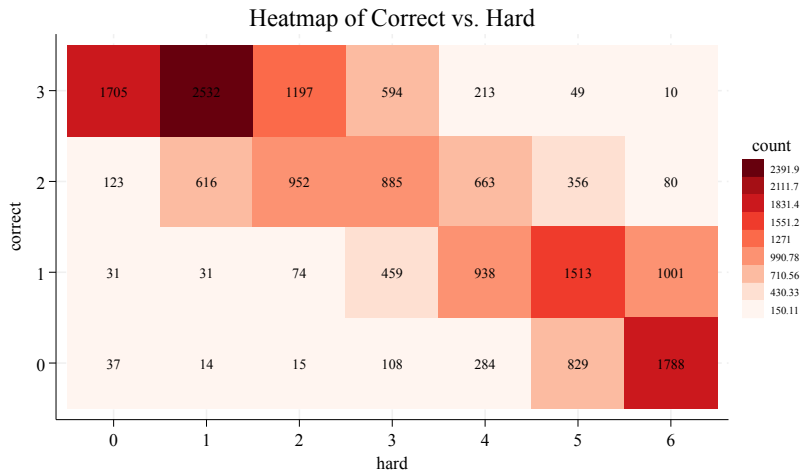
Did people find the quiz hard?



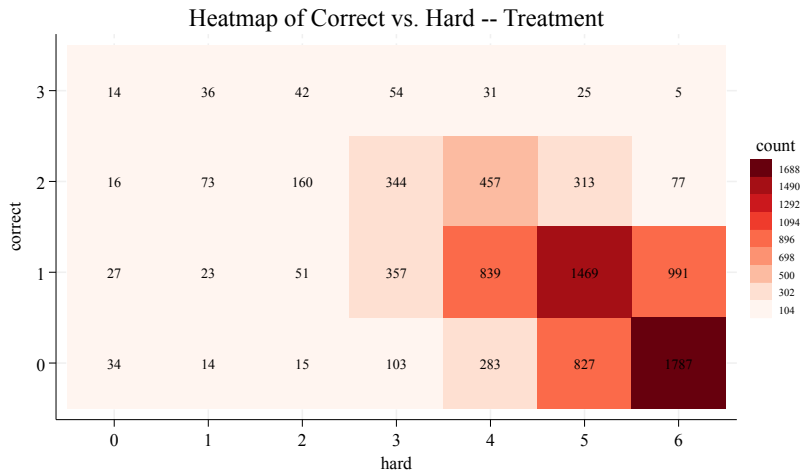
- Strong correlation between perceived difficulty and actual performance [▶ Heat maps](#)
- Similar time spent in T and C (4:20 vs. 3:54)

[▶ Back](#)

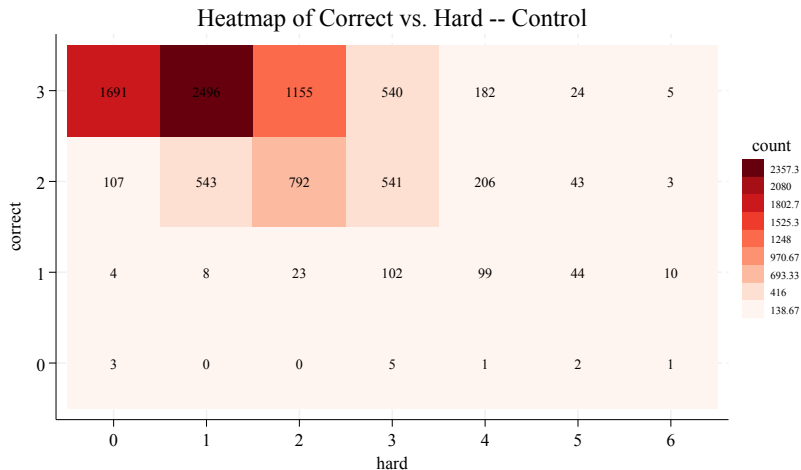
Heatmap - # Correct vs. Perceived Difficulty (All)



Heatmap - # Correct vs. Perceived Difficulty (Treatment)



Heatmap - # Correct vs. Perceived Difficulty (Control)



ATE on Full Sample

Dep. variable	(1) Misinfo Concern	(2) WTP	(3) Trust SZ	(4) Trust X	(5) Trust IG	(6) Trust TikTok	(7) Trust Bild	(8) Trust ARDZDF	(9) Index
AI Treatment	0.223*** (0.011)	-0.085 (0.184)	-0.019*** (0.005)	-0.031*** (0.008)	-0.047*** (0.011)	-0.020*** (0.008)	-0.049*** (0.012)	-0.023*** (0.007)	-0.0374*** (0.006)
Sample mean	5.605 (0.729)	21.002 (11.057)	2.898 (0.334)	0.236 (0.482)	0.740 (0.703)	0.247 (0.472)	0.794 (0.755)	2.789 (0.458)	1.357 (0.397)
Observations	17,284	14,461	17,113	16,096	15,498	15,496	16,718	17,153	17,294
Adjusted-R ²	0.023	-0.000	0.001	0.001	0.001	0.000	0.001	0.001	0.002
F-statistic	416.980	0.214	13.470	16.846	17.066	6.857	17.326	10.865	38.32

[▶ Back](#)

ATE on Tracked Sample

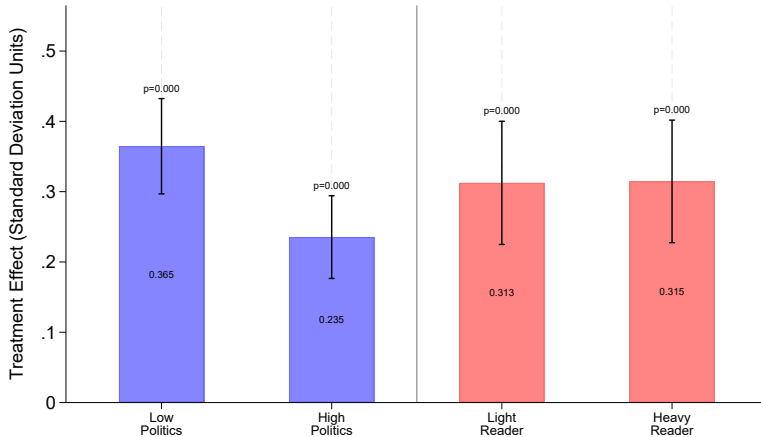
Dep. variable	(1) Misinfo Concern	(2) WTP	(3) Trust SZ	(4) Trust X	(5) Trust IG	(6) Trust TikTok	(7) Trust Bild	(8) Trust ARDZDF	(9) Index
AI Treatment	0.197*** (0.017)	0.198 (0.311)	-0.019*** (0.007)	-0.004 (0.012)	-0.034* (0.019)	-0.012 (0.013)	-0.045** (0.020)	-0.042*** (0.011)	-0.026*** (0.010)
Sample mean	5.624 (0.682)	22.749 (10.977)	2.938 (0.256)	0.221 (0.462)	0.774 (0.705)	0.252 (0.468)	0.778 (0.755)	2.823 (0.410)	1.372 (0.387)
Observations	5,923	4,980	5,888	5,527	5,329	5,316	5,736	5,902	5,936
Adjusted-R ²	0.021	-0.000	0.001	-0.000	0.000	-0.000	0.001	0.002	0.001
F-statistic	127.328	0.406	8.286	0.081	3.049	0.901	4.995	15.357	6.81

[▶ Back](#)

ATE with Controls

	(1)	(2)	(3)
Dep. variable	Misinfo Concern	Trust SZ	WTP
AI Treatment	0.224*** (0.011)	-0.020*** (0.005)	-0.089 (0.177)
Sample mean	5.605 (0.729)	2.898 (0.334)	21.002 (11.057)
Controls	Yes	Yes	Yes
Observations	17,284	17,113	14,461
R-squared	0.040	0.032	0.078
F-statistic	87.205	36.155	189.204

Heterogeneity: Misinformation Concern



Pre-registered heterogeneity: perceived difficulty

	Misinfo Concern		Visits (5d)		Subscription (150d)	
	Hard	Not Hard	Hard	Not Hard	Hard	Not Hard
AI Treatment	0.3070*** (0.0143)	0.0846*** (0.0182)			0.0161** (0.0069)	0.0072 (0.0073)
AI Treatment $\times$ Post			0.1213** (0.0604)	0.0221 (0.0626)		
Sample mean	5.668 (0.648)	5.533 (0.803)	4.717 (5.485)	4.665 (5.563)	0.964 (0.185)	0.964 (0.187)
Individual FE			Yes	Yes		
Date FE			Yes	Yes		
Observations	9,318	7,859	81,770	66,898	3,258	2,655
R-squared	0.0541	0.0026	0.7729	0.7836	0.0018	0.0003

- Perceived difficulty: *"How easy were the questions for you? (0-6 scale)"*
- Caveat: quiz difficulty not randomly assigned

Heterogeneity: age and gender

Dep. variable	(1) Concern	(2) Trust SZ	(3) WTP	(4) Concern	(5) Trust SZ	(6) WTP
Treatment	0.228*** (0.016)	-0.014** (0.007)	0.033 (0.232)	0.251*** (0.016)	-0.021*** (0.007)	0.059 (0.254)
Treatment $\times$ Old (60+)	-0.004 (0.022)	-0.008 (0.010)	-0.151 (0.364)			
Treatment $\times$ Female				-0.052** (0.022)	0.006 (0.010)	-0.316 (0.368)
Old (60+)	-0.012 (0.017)	-0.015** (0.007)	2.582*** (0.252)			
Female				0.160*** (0.017)	0.010 (0.007)	-0.009 (0.256)
Obs	17,118	17,011	14,475	17,118	17,011	14,475
Mean dep. var.	5.606 (0.728)	2.899 (0.333)	21.001 (11.056)	5.606 (0.728)	2.899 (0.333)	21.001 (11.056)
Adjusted-R ²	0.024	0.001	0.013	0.033	0.001	-0.000
F statistic	143.081	9.097	63.742	191.815	6.367	0.583

Daily visits to SZ: robustness

Dep. variable	(1)	(2)	(3)
	N. of visits		
Sample	Baseline	T + No quiz	Baseline
Post	0.202** (0.087)	0.047 (0.105)	0.207*** (0.053)
Post $\times$ AI Treatment	0.085** (0.042)	0.113*** (0.043)	0.018** (0.009)
AI Treatment			0.012 (0.027)
Obs	155,142	142,470	155,142
Mean dep. var.	4.653 (5.493)	4.017 (5.169)	4.65 (5.493)
Adjusted-R ²	0.768	0.770	.
F statistic	6.224	3.882	.
Individual FE	Yes	Yes	No
Date FE	Yes	Yes	Yes
Method	OLS	OLS	Poisson

Daily visits to SZ: Split Window

Dep. variable	(1)	(2)	(3)	(4)
	Number of daily visits			
Post Window	(1;5)	(6;10)	(11;15)	(16;20)
AI Treatment $\times$ Post	0.0847** (0.0418)	0.0627 (0.0424)	0.0509 (0.0462)	0.0326 (0.0475)
Post	0.2023** (0.0873)	-0.2632 (0.4014)		
Sample mean	4.653 (5.493)	4.586 (5.437)	4.621 (5.479)	4.558 (5.404)
Individual FE	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes
Observations	155,142	155,142	155,142	155,003
R-squared	0.768	0.764	0.763	0.761
F-statistic	6.224	1.259	1.213	0.470

Heterogeneity: Engagement (Sports)

	(1)	(2)	(3)	(4)	(5)
Subgroup	Full	Low Politics	High Politics	Heavy Reader	Light Reader
AI Treatment $\times$ Post	-0.0275 (0.0572)	-0.0769 (0.0858)	0.0054 (0.0820)	0.0021 (0.1775)	-0.0674 (0.0418)
Sample mean	1.206 (6.007)	0.971 (5.707)	1.476 (6.425)	2.791 (9.070)	0.052 (1.160)
Individual FE	Yes	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes	Yes
Obs	155,142	63,778	84,838	38,532	38,584
Adjusted-R ²	0.4618	0.4817	0.4471	0.4849	0.1998
F statistic	1.337	0.672	3.280	0.502	1.305

Alternate Explanation: Entertainment Effects

Dep. variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Number of daily visits							
Sample	Drop Visits to Surveys				Drop People with Other Surveys			
Timespan	(-21;+3)	(-21;+5)	(-21;+10)	(-21;+14)	(-21;+3)	(-21;+5)	(-21;+10)	(-21;+14)
AI Treatment $\times$ Post	0.113** (0.0517)	0.0848** (0.0418)	0.0745** (0.0354)	0.0673* (0.0344)	0.116** (0.0517)	0.0860** (0.0418)	0.0742** (0.0354)	0.0664* (0.0344)
Post	0.195** (0.0888)	0.204** (0.0875)	0.210** (0.0834)	0.219*** (0.0822)	0.189** (0.0887)	0.200** (0.0873)	0.206** (0.0833)	0.214*** (0.0821)
Sample mean	4.690 (5.534)	4.653 (5.493)	4.615 (5.457)	4.614 (5.463)	4.690 (5.534)	4.653 (5.493)	4.615 (5.457)	4.614 (5.463)
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	143,133	155,062	184,887	208,745	142,992	154,908	184,698	208,530
R-squared	0.778	0.777	0.774	0.772	0.778	0.777	0.774	0.772

Alternate Explanation: Social Media Buzz

Dep. variable	(1)	(2)	(3)
	Number of daily visits		
Source	Overall	Organic	Social
AI Treatment $\times$ Post	0.0847** (0.0418)	0.0847** (0.0418)	-0.0001 (0.0003)
Post	0.2023** (0.0873)	0.2035** (0.0873)	-0.0009 (0.0014)
Sample mean	4.653 (5.493)	4.653 (5.493)	0.001 (0.031)
Individual FE	Yes	Yes	Yes
Date FE	Yes	Yes	Yes
Observations	155,142	155,116	155,116
R-squared	0.768	0.768	0.037
F-statistic	6.224	6.274	0.331

AI Article Demand and Supply

Dep. variable Window	(1) AI articles clicked (daily) [−21, +5]	(2) [−21, +21]
AI Treatment × Post	-0.0043 (0.0087)	0.0007 (0.0070)
Sample mean	0.087 (0.904)	0.071 (0.806)
Individual FE	Yes	Yes
Date FE	Yes	Yes
Obs	155,142	250,360
Adjusted-R ²	0.0837	0.0662
F statistic	1.706	0.861

Dep. variable Window	(1) AI articles published (daily) Feb 10–Mar 9	(2) Jan 1–Apr 9	(3) Jan 1–Jun 9
Sample mean	1.036 (1.138)	0.929 (1.272)	0.919 (1.176)
Week FE	Yes	Yes	Yes
Obs	28	98	160
Adjusted-R ²	0.5774	0.3350	0.2754
F statistic	0.277	0.299	0.300

[▶ Back](#)

Subscription status: alternate control

Dep. variable	(1)	(2)	(3)	(4)
	Subscription Status			
Months Post-Intervention	2 Months	3 Months	4 Months	5 Months
AI Treatment (vs. Alternate Control)	0.0132* (0.00785)	0.0206** (0.00946)	0.0270** (0.0111)	0.0249** (0.0120)
Constant	0.980*** (0.00259)	0.968*** (0.00318)	0.955*** (0.00393)	0.950*** (0.00460)
Controls	Yes	Yes	Yes	Yes
Observations	3,284	3,284	3,284	3,284
R-squared	0.009	0.009	0.007	0.007

- Alternate control: highly engaged subscribers who didn't take the quiz
- Not statistically different from control group

Control group analysis: perceived difficulty

Dep. Variable	(1) Visit Count	(2) Subscription Status	(3) Misinfo Concern	(4) Trust SZ
Hard $\times$ Post	-0.0510 (0.0588)			
Hard		-0.00644 (0.00720)	-0.0148 (0.0174)	-0.00651 (0.00706)
Post	0.270** (0.135)			
Constant	4.582*** (0.0256)	0.964*** (0.00458)	5.501*** (0.0115)	2.911*** (0.00467)
Observations	75,608	3,005	8,658	8,586
R-squared	0.774	0.000	0.000	0.000

- No difference in behavior for who found the quiz hard vs. those who didn't
- Caveat: Descriptive analysis with limited causal validity

Long-run ATE (Feb 2026 sample)

Dep. variable	(1) Misinfo Concern	(2) Trust SZ	(3) Trust X	(4) Trust IG	(5) Trust TikTok	(6) Trust Bild	(7) Trust ARDZDF
AI Treatment	-0.016 (0.046)	0.005 (0.005)	-0.018 (0.026)	-0.011 (0.035)	-0.023 (0.026)	-0.007 (0.036)	-0.009 (0.016)
Sample mean	5.441 (0.816)	2.997 (0.087)	0.220 (0.421)	0.532 (0.513)	0.210 (0.414)	0.538 (0.568)	2.976 (0.244)
Observations	1,261	1,190	1,065	866	1,030	972	995
Adjusted-R ²	-0.001	-0.000	-0.000	-0.001	-0.000	-0.001	-0.001
F-statistic	0.115	1.000	0.487	0.093	0.820	0.038	0.346

[▶ Back](#)

Relevant literature

1. Misinformation and the role of fact-checking
Vosoughi et al. (2018); Hengel et al. (2025); Henry et al. (2022); Guriev et al. (2025); Nyhan et al. (2020); Chopra et al. (2022)
2. Responses to one's inability to detect misinformation
Assenza et al. (2024) and Harris et al. (2024)
3. Role of technology in “content” creation
Longoni et al. (2022); Toff and Simon (2025); Groh et al. (2024); Vaccari and Chadwick (2020)
4. Strategies to combat low-quality information
Costello et al. (2024); Ahmad et al. (2024); Allen et al. (2021)