

Claim Inflation in Science Journalism

A Case Study of Five Futurism Articles: Levels, Patterns, Explanatory Models

Author: Thomas Schüller (ORCID 0009-0003-9799-6747) · Research project Syntagma Forschung · Vienna

Contact: forschung@tschueller.com

Identifier: FORSCH-PROV0476 (provisional; decided 20 Oct 2026, the earlier numbers 0465 and 0471 were assigned twice)

Version: 0.5 · 14 Oct 2026 · Preprint draft, passive edition (read, verify)

License: CC BY 4.0

Earlier versions: v0.1 (withdrawn), v0.2, v0.3 and v0.4 (superseded). See the revision note.

Revision note

This version fully replaces v0.1 to v0.4. v0.1 and v0.2 contained errors that affected the conclusions; v0.3 contained open points that v0.4 and v0.5 close against primary sources or downgrade in source status (rows 9–15). A paper about claim inflation states its corrections openly.

#	Error in earlier version	Version	Correction in v0.3
1	Unsupported sources (Shirley 2016, Zhang et al. 2024, a DOI for the Alzheimer case, citation URLs)	v0.1	removed; reference list re-verified from scratch
2	The 23andMe audit row examined a reconstructed headline that had never been read	v0.1	replaced by the actual article
3	23andMe rated "correct", Regeneron presented as buyer of the biobank. After the auction the buyer was TTAM Research Institute (USD 305 m, court order late June 2025)	v0.2	case re-rated (§3.5)
4	Sumner 2014 mixed up with the sample sizes of Sumner 2016 (534/582)	v0.2	462 press releases, 668 news items
5	GLP-1 article said to hide countermeasures; Novo/Ypsomed statements	v0.1/v0.2	the article presents the water-based synthesis as a solution; manufacturer statements removed from the assessment
6	Harvard case: "peer-review caveat missing", "n=1,200, 37.4% across all responses"	v0.2	caveat appears in the third paragraph; numbers differ between versions (§3.3)
7	Engagement thesis struck out as refuted	v0.2	overstated; unsupported at press-release→news level, supported at click level (§5)
8	Ordering of cases by verification cost called "monotone"	v0.2	no longer monotone with corrected findings (§4, §6)
9	Harvard figure 43% carried as "version difference unresolved"	v0.3	resolved: 43% appears in the abstract of v1/v2, the results section gives 37.4% in all versions (excluding Flourish); Futurism took 43% and "five of six apps" from Psychology Today (§3.3)
10	Arginine: approval in Japan only via secondary sources	v0.3	approval for urea-cycle disorders supported by the prescribing information and the study text; dose (HED 939.9 mg/kg/day) and limitations quoted from the study text (§3.1)
11	23andMe: "about 15 m records"; final round "06/2025"	v0.3	about 13 m customers per AP; final round on 13 Jun 2025; the 8-K of 19 May 2025 states the court condition explicitly (§3.5)
12	Literature table §5 without reference quantities	v0.3	reference quantities specified: Robertson (headline of average length, relative change), Schat (RR 1.45 = uptake into the news; RR 16.1 = carry-over of exaggeration), Adams (aligned headlines and claims)
13	Reddit: wording "sensitive topics" not compared	v0.3	Futurism turns it into "controversial topics" (§3.4)
14	Fujii: dose (HED), sex of the animals and limitations carried as study text	v0.4	source status downgraded: these statements come from text excerpts and secondary sources, not from an opened full text; two check runs disagree on source status. Verbatim verified are the abstract, the highlights and the legend of figure 1 (§3.1)
15	23andMe: approval order "not consulted"	v0.4	Memorandum Opinion of 27 Jun 2025 read verbatim (freely available on the court site); the sale order (Doc. 910) remains unopened (§3.5)

Abstract (EN)

Five articles from Futurism.com (09/2025–02/2026; health, AI, genetics) are checked against their primary sources. Coding uses a four-level scale from "no deviation" to "evidence jump in the core claim". Result: headline condensation (level 1) occurs in three cases, text overstatement (level 2) in one case, and an evidence jump in the core verb (level 3) in one case. The two stronger cases follow different patterns: in the preclinical Alzheimer article the verb exceeds the evidence; in the 23andMe article a provisional event is presented as concluded on the day of the announcement and is contradicted by the later course of events.

The default assumption that exaggeration follows engagement incentives is not supported at the press-release→news step (Sumner et al. 2014, 2016; Adams et al. 2019: no higher uptake of exaggerated press releases), but is supported at the click level (Robertson et al. 2023; Schat et al. 2018). Two further hypotheses are proposed on equal footing, verification cost/sanction and time pressure, together with a preregistrable phase-2 design that tests both against the engagement hypothesis. With n = 5 and one coder the study is illustrative.

Keywords: science communication, claim inflation, press release, headline, epistemic vigilance, credence goods, p/δ asymmetry

1. Research questions

- **Q1** At which level of an article (headline, text, core claim) does claim inflation occur, and in which patterns?
- **Q2** Does the engagement explanation hold for the observed cases?
- **Q3** Which alternative, testable explanations are available?

The responsibility chain author → press office → editorial desk is largely settled empirically (§7) and is only reported here.

2. Material and method

2.1 Cases

#	Article (headline)	Author	Published	Reading depth	Primary source
1	Scientists Intrigued by Old Drug That Reverses Signs of Alzheimer's in Mice	F. Landymore	29 Nov 2025	full text	Fujii et al. 2025, <i>Neurochem. Int.</i> 191:106082 (abstract, highlights and legend of figure 1 read; methods and discussion only via text excerpts and secondary sources; full text not opened)
2	GLP-1s Are an Environmental Catastrophe	V. Tangermann	8 Feb 2026	full text	<i>Nature Sustainability</i> (Univ. of Melbourne); university accompanying text read
3	Harvard Research Finds That AI Is Emotionally Manipulating You to Keep You Talking	V. Tangermann	24 Sep 2025	full text	arXiv 2508.19258 (v1, v2, v3; abstracts and results section) and HBS Working Paper 26-005
4	Reddit's AI Suggests That People Suffering Chronic Pain Try Opioids	R. Witte	19 Oct 2025	full text	404 Media (E. Maiberg), 16 Oct 2025 (original paywalled; teaser, Engadget and syndicated copy read)
5	Bankrupt 23andMe Just Sold Off All Your DNA Data	M. Harrison Dupré	19 May 2025	full text	Form 8-K of 23andMe, 19 May 2025 (Regeneron) and 13 Jun 2025 (TTAM); Memorandum Opinion of 27 Jun 2025 verbatim; sale order (Doc. 910) not opened

Selection was not random (topic list of the author). Every headline and text was read; in v0.1 this was not the case for one article (revision note, row 2).

2.2 Coding (one coder, no reliability test)

Level	Label	Criterion
0	no deviation	headline and text match the primary source and the facts
1	headline condensation	headline or teaser stronger than the text; the text contains counter-context or a qualification
2	text overstatement	a statement in the text goes beyond what was established at the time

Also coded: **verification cost** for readers (low/medium/high; author's estimate) and **timeliness gap** (days between primary event and article).

3. Case findings

3.1 Alzheimer / arginine: level 3

Primary study. Fujii K, Takeuchi T, Fujino Y, Tanaka N, Fujino N, Takeda A, Minakawa EN, Nagai Y: *Oral administration of arginine suppresses A β pathology in animal models of Alzheimer's disease*. *Neurochem Int* 2025;191:106082. DOI 10.1016/j.neuint.2025.106082, PMID 41175945. Models: A β 42 aggregation in vitro, Drosophila, App^{NL-G-F} knock-in mouse (male). The ScienceDirect identifier (PII) is S019701862500155X.

Press release (Kindai University, EurekAlert, 21 Nov 2025): the title names "animal models" and uses the verb "curbs"; the subheading frames arginine as a "therapeutic approach". A dosage caveat and the need for clinical trials appear in the text. Further passages: "compelling proof of concept" and a quote from the study lead that arginine "could be rapidly translated to clinical trials". The dosage caveat reads: the protocol was "optimized for research purposes and does not correspond to commercially available formulations"; the study's own remark that the dose is about twice the approved maximum for humans is missing. The press release already contains translational rhetoric, but not the verb "reverses". The translational rhetoric is already in the study itself: the highlights call arginine "a safe and cost-effective repositioning drug candidate for AD" (verified verbatim). The verb "reverses" appears in none of the three stages.

Futurism. Headline with "Reverses Signs"; the abstract says "suppresses", the press release "curbs". The lead puts the therapy expectation in the first sentence, without the animal reference. Deviations from the study (text excerpts) and from secondary reports:

Futurism statement	Finding in study text and secondary sources
Arginine "already prescribed for high blood pressure"	Study text per excerpt: "urea cycle deficits in Japan"; prescribing information (KEGG/JAPIC) lists congenital urea-cycle disorders and lysinuric protein intolerance; no Japanese hypertension indication found
Mice received "small doses" in "drinking water and food"	6% arginine in drinking water from five weeks (text excerpt, media reports); human-equivalent dose 939.9 mg/kg/day, "approximately twice" the approved maximum (text excerpt; Lifespan.io: 940 mg/kg/day); food only for the flies
Arginine could "break up" and "flush" plaques	treatment from five weeks of age, i.e. before plaque formation: prevention design (study text); at six months fewer plaques in cortex and hippocampus, at nine months per secondary sources only a non-significant trend toward smaller plaque area in the hippocampus, with fewer dense plaques at both time points
Y-maze described with "open arms"	arm entries and total distance measured in a Y-maze; "open arms" belongs to the elevated plus maze (mix-up)
Nagai as "study coauthor"	Nagai is senior author (last position)

The caveat on the amyloid hypothesis is in the last paragraph; the article contains no independent expert voice, and "Scientists Intrigued" rests on quotes from the study lead.

Context of the amyloid hypothesis. Lecanemab (Clarity AD, *NEJM* 2023;388:9–21) lowered CDR-SB by 0.45 points versus placebo after 18 months (95% CI –0.67 to –0.23); donanemab (TRAILBLAZER-ALZ 2, *JAMA* 2023;330:512–527) by 0.67 points. Critical analyses regard both as statistically significant and of questionable clinical relevance. Amyloid reduction is thus pharmacologically achievable and linked to a small clinical effect; amyloid reduction in a mouse model is not a "reversal of signs of Alzheimer's".

Limitations per text excerpt (not confirmed against the full text): male mice (PubMed indexing "Male" and a media report support this; the methods text was not confirmed in a second check run); App^{NL-G-F} model with familial Arctic mutation (E22G) that most sporadic cases do not share; no tau tangles and no apparent neuron loss in the model.

Strength of evidence: low (animal models, prevention design, dose about twice the approved maximum, no human data). **Verification cost:** high. **Timeliness gap:** 8 days (press release→article).

Open: sample sizes per group, numerical in-vivo effect sizes with dispersion and p-values (plaque load, Aβ42, behavior) and the full wording of the limitations. Per Crossref the article is open access (CC BY 4.0, since 3 Nov 2025) and freely readable on ScienceDirect but blocked for automated retrieval; no PMC version or preprint exists. The figures must be read off the PDF. Part of the behavioral and nine-month statements comes from secondary sources (news-medical.net, ScienceBlog).

3.2 GLP-1 / peptide manufacturing: level 1

The article draws on the University of Melbourne *Nature Sustainability* paper and on an accompanying text by study lead John Wade. Headline "Catastrophe", subheading "dirty secret". The text describes solid-phase peptide synthesis and its solvent demand (dimethylformamide), cites more than GBP 123 million of annual solvent waste for semaglutide (figure according to Wade), and presents the research group's water-based alternative as a solution in the third paragraph. It closes with the caveat of whether the process can be scaled industrially.

The headline is thus stronger than the text; the text contains the solution and the scale-up caveat. The word "disaster" already appears in the title of the university's accompanying publication or its reproduction on phys.org (02/2026): the sharpening comes from the source; Futurism replaces "disaster" with "catastrophe".

Correction to v0.1/v0.2: The suppression of countermeasures claimed there does not apply to the article. The statements on Novo Nordisk and Ypsomed come from earlier research, are not re-verified here, and are removed from the assessment.

Verification cost: medium. **Timeliness gap:** 5 days (phys.org 3 Feb 2026 → 8 Feb 2026).

3.3 Harvard / AI companion apps: level 1

In the text the article frames the study correctly: companion apps were studied, not assistants like ChatGPT; the study is not yet peer-reviewed (third paragraph, linking arXiv 2508.19258); five of six apps showed manipulative farewells, one (Flourish) did not; a second study with 3,300 adults found up to 14-fold higher engagement after the goodbye; some participants felt put off, so the tactic could backfire.

Deviations: the headline generalizes to "AI" and to "You". The description (meta teaser) speaks of "almost half of the most popular apps", the text of five of six apps and 43% of interactions: the teaser confuses the share of interactions with the share of apps.

State of the numbers (checked 14 Oct 2026 against versions v1–v3). The working paper examines six apps (PolyBuzz, Character.ai, Talkie, Chai, Replika, Flourish) with 200 farewells each, 1,200 in total. The farewells are scripted: GPT-4o generates the user messages, followed by a randomly drawn farewell formula; what is measured are the apps' replies. The abstract nevertheless speaks of "real farewells". Platform values: PolyBuzz 59.0% (118/200), Talkie 57.0% (114/200), Replika 31.0% (62/200), Character.ai 26.5% (53/200), Chai 13.5% (27/200), Flourish 0%. The 37.4% in the results section is the mean of the five apps **excluding** Flourish; across all six it would be 31.2%.

Version	Abstract	Results section
v1 (15 Aug 2025)	43% (PDF version)	37.4%
v2	43%	not retrieved separately
v3 (7 Oct 2025)	37%	37.4%

The discrepancy thus lies in the source itself (abstract v1/v2 versus results section), not with Futurism. A derivation of the 43% from the study data was not found. Futurism states "43 percent of the interactions" and "five out of six" and adds "As spotted by Psychology Today"; the Psychology Today piece by M. Wei of 22 Sep 2025 gives the same figures. What is attributable to Futurism is the teaser ("almost half of the most popular AI companion apps"), which swaps share of replies and share of apps, and the omission of the GPT-4o simulation. The article's "five times longer" was not verified. The working paper carries the note "draft form ... for purposes of comment and discussion only"; "not peer-reviewed" in the article is covered. v0.2 stated "37.4% across all 1,200 responses"; that was imprecise. Note on the source record: the arXiv abstract page of v1 shows 37%, the v1 PDF 43%; the arXiv version history was not checked. The HBS page for WP 26-005 ("Revised October 2025") gives different key figures (3,458 adults, up to 16-fold).

Verification cost: low to medium (preprint freely accessible). **Timeliness gap:** about four weeks (arXiv 08/2025 → 24 Sep 2025).

3.4 Reddit Answers / opioids: level 1

Futurism (R. Witte, 19 Oct 2025) relies on 404 Media (E. Maiberg, 16 Oct 2025, updated 17 Oct 2025) and a post in Reddit's moderator forum (r/ModSupport; the trigger was a thread in r/FamilyMedicine). The text reports: the AI first recommended kratom; only after a user asked whether there was a medical rationale for heroin did it describe heroin as occasionally used in pain treatment but controversial and strictly regulated, and it quoted a user post describing heroin as life-saving but linked to dependence. Reddit then removed "Related Answers" on "sensitive topics" from display.

The headline ("Suggests ... Try Opioids") condenses: heroin came only on follow-up and with a dependence warning; the unprompted recommendation concerned kratom. The text discloses this sequence. Two sharpenings in the text: Reddit's "sensitive topics" becomes "controversial topics" at Futurism, and the phrase "DIY heroin use might be an effective strategy" hides that the bot's answer was more careful ("controversial and subject to strict regulations"). The assessment of the answer as inappropriate is supported by the incident; the two phrasings are borderline cases between levels 1 and 2 and stay at level 1 here because the sequence is disclosed. Engadget also reports a questionable answer on fever in newborns; the Futurism article does not mention it. The 404 Media original is paywalled; Reddit's quote is documented via Futurism and Engadget.

Verification cost: low (public forum post, cross-check possible). **Timeliness gap:** 3 days (404 Media 16 Oct → 19 Oct 2025).

3.5 23andMe / sale of the biobank: level 2 (*wrongly "correct" in v0.2*)

On 19 May 2025 23andMe and Regeneron announced an agreement to buy nearly all assets of 23andMe for USD 256 m in cash, including the biobank (about 13 m customers per AP). The auction had ended on 16 May 2025, the purchase agreement is dated 17 May 2025. The 8-K of 19 May 2025 carries the subheading "Transaction Subject to Court Approval"; it names approval by the bankruptcy court (Eastern District of Missouri) and clearance under the Hart-Scott-Rodino Act as conditions, with a hearing set for 17 Jun 2025. The article appeared the same day. Headline: "Just Sold Off All Your DNA Data"; kicker: "There's nothing you can do about it."; meta description: 23andMe had sold the biobank to Regeneron; opening: "Your DNA results are in — and they're 100 percent the property of a pharmaceutical manufacturer." The text describes Regeneron as buyer that "announced today" it would buy, and then presents the data as already held by the pharmaceutical company. It mentions neither the condition of court approval nor the possibility of overbids.

Course of events: the agreement was an opening bid in the bankruptcy proceedings. TTAM Research Institute (nonprofit, controlled by co-founder Anne Wojcicki) outbid it with USD 305 m. The auction was reopened on 4 Jun 2025 (secondary source); the final round between TTAM and Regeneron took place on 13 Jun 2025 (23andMe 8-K of 13 Jun 2025: TTAM winning bidder, Regeneron backup bidder, again subject to court approval). Regeneron did not bid further. The bankruptcy court (Judge Brian Walsh) approved the sale to TTAM on 27 Jun 2025 (23andMe 8-K; AP, 1 Jul 2025). The buyer of the biobank was not Regeneron. The retrieved article carries no update note; the "modified_time" field equals the first publication (published_time 19 May 2025, 4:17 pm US Eastern). That does not prove the text stayed unchanged, but no correction note was found. The Memorandum Opinion of Judge Brian C. Walsh of 27 Jun 2025 (freely available on the court site) was checked verbatim: the court grants the sale; the customer count is about 13 million, about 1.9 million customers had deleted their accounts since the filing; the proceeds of USD 305 m (of which USD 302.5 m in cash) are nearly six times the opening auction bid of USD 52 m; the privacy policy permitted the sale of personal data in a sale of assets subject to an opt-out right, and TTAM continues deletion rights and opt-out "in perpetuity". The actual sale order (Doc. 910, 177 pages, via the Kroll portal) was not opened; its conditions (deletion right, external audits) are documented only through a law-firm analysis. Walsh is also quoted secondarily calling the sale of genetic data "a scary proposition".

The article reports a fact that was correct at the time of publication (agreement, price, scope) but presents it as concluded. The situation was not final on the day of publication, and the report's own primary source (the 8-K) said so explicitly. This is a text overstatement (level 2), not a headline condensation.

The outlet's earlier coverage was more precise: on the 2023 data breach Futurism ran "Hackers Selling Stolen Customer DNA Data From 23AndMe" (12 Oct 2023), and the agent lay with the attackers.

Verification cost: low (8-K and press releases freely accessible). **Timeliness gap:** 0 days.

4. Overall picture

Case	Domain	Level	Pattern	Verification cost	Timeliness gap
Alzheimer/arginine	preclinical	3	evidence jump in the verb; detail deviations	high	8 days
23andMe sale	corporate event	2	provisional presented as concluded	low	0 days
GLP-1 manufacturing	impact assessment	1	sharpening taken over from the source	medium	5 days
Harvard/AI	published study	1	generalization in headline; teaser mix-up	low-medium	ca. 28 days
Reddit/AI	documented incident	1	condensation of the sequence	low	3 days

Three observations:

1. Headline condensation (level ≥ 1) occurs in **all five cases**; in four cases the text itself contains the qualification or counter-context. In this sample condensation is the rule, overreach of the text or the core claim the exception.
2. The two cases ≥ 2 follow **different patterns**: an evidence jump (Alzheimer: high verification cost) and an overstatement of finality (23andMe: low verification cost, timeliness gap zero).
3. The monotone relation between verification cost and inflation asserted in v0.2 does **not** hold: the case with the lowest verification cost (23andMe) is at level 2, the case with medium cost (GLP-1) at level 1, as are the two with low cost. A pattern shows only for the evidence jump (level 3 at the highest verification cost).

5. Engagement thesis: state of the evidence

Finding	Source	Effect on the thesis
462 UK press releases, 668 news items: exaggeration in the press release goes with exaggeration in the news (odds ratios 6.5 [CI 3.5-12] / 20 [7.6-51] / 56 [15-211] for advice / causal claim / animal→human); little evidence that exaggeration raises uptake	Sumner et al. 2014, <i>BMJ</i> 349:g7015	against (press-release→news level)
534 journal press releases, 582 news items: no evidence that exaggeration increases coverage or that caveats reduce it	Sumner et al. 2016, <i>PLOS ONE</i> 11:e0168217	against
Randomized trial, 312 press releases, 2,257 news items: press releases with aligned headlines and claims lost no uptake (intention-to-treat 55% vs. 55%); for caveats 53% vs. 56%	Adams et al. 2019, <i>BMC Medicine</i> 17:91	against
Replication for causal and animal→human claims, not for advice; no higher uptake	Bratton et al. 2019, <i>Wellcome Open Res.</i> 4:148	against
129 Dutch press releases, 185 news items: relative chance that an exaggerated press release is taken up into the news 1.45 (CI 1.02-2.04); relative chance of exaggeration in the news given an exaggerated press release 16.1	Schat et al. 2018, <i>Ned Tijdschr Geneesk</i> 162:D1936	supports (narrowly)
About 105,000 Upworthy headline variants (22,743 tests, about 5.7 m clicks): for a headline of average length each additional negative word raises the click rate by 2.3% (relative)	Robertson et al. 2023, <i>Nat. Hum. Behav.</i> 7:812-822	supports (click level, not science news)
Neither caveats in the text nor attenuating wording in headlines changed the judged newsworthiness	Bott et al. 2019, <i>J. Exp. Psychol. Appl.</i> 25:517-542	against, insofar as attention is meant

Summary: at the press-release→news level there is no evidence of a reach gain from exaggeration. At the level of headline wording, click research shows advantages for negativity. Both are compatible: condensation in headlines can bring clicks without exaggeration in press releases bringing coverage. The engagement thesis is thus not refuted (v0.2 went too far) but narrowed to the headline level. Headline condensation in 5 of 5 cases is compatible with it.

6. Explanatory models for cases beyond the headline

6.1 Verification cost and sanction (H1)

Science news can be modeled as a credence good (Darby & Karni 1973; Dulleck & Kerschbamer 2006): even after consumption the reader can judge quality only at high cost. Epistemic vigilance (Sperber et al. 2010) takes effort; where checking is expensive it declines. Quality uncertainty favors poor quality (Akerlof 1970). Prediction: overreach of the core claim rises with verification cost and falls with the probability of sanction (expert public, correction practice).

Support in the sample: the evidence jump (level 3) at the highest verification cost. Against: 23andMe (level 2 at the lowest verification cost). Against the sanction component speaks the replication record: Bratton et al. (2020) found a drop in causal overstatement in press releases from 28% to 13% after Sumner 2014 (correlational); an automated large-scale analysis (Yu et al. 2020, 16,431 pairs) did not replicate the sharp decline. Against the vigilance assumption, Michaelian argues (speaker honesty instead of hearer vigilance).

None of the studies found operationalized verification cost as an independent variable. H1 is theoretically anchored and empirically untested.

6.2 Time pressure (H4, new)

The only case with a timeliness gap of zero is also the case of finality overstatement. A report on the day of an announcement has little time to ask what is provisional about the announcement. Prediction: overstatement of finality (provisional as final) falls with the distance to the primary event. $n = 1$; H4 is a hypothesis from a single case.

6.3 Source adoption (H5)

In the GLP-1 case the sharpening comes from the university's accompanying publication; in the Alzheimer case the press release contains the translational rhetoric. This matches the coupling in Sumner et al. (OR up to 56): the editorial desk transfers what the institution formulates. H5 explains the origin of the sharpening, not its amplification (the verb "reverses" appears in none of the examined sources).

7. Chain of responsibility

The coupling between press release and news item is strong (odds ratios 6.5–56). Responsibility is spread over author, press office, editorial desk and platform; in such chains each station can point to the previous one (*problem of many hands*, Thompson 1980, *APSR* 74:905–916). The editorial desk retains residual responsibility for the headline: the Alzheimer case shows a verb amplification that appears in no earlier stage.

8. Link to FORSCH-0083 (ρ/δ)

Quantity	Bearer in the case
δ (disposal over wording)	author, press office, editorial desk
ρ (risk of misorientation)	readership (therapy expectation, wrong data-protection assumption)
Feedback	weak: with cheaply checkable statements counter-pressure is possible, with expensive ones it is not

H1 and H4 are two ways in which the feedback fails to occur (expensive checking; no time to check).

9. Limitations

- $n = 5$, one outlet, no random selection, one coder without reliability test.
- The assignment of verification cost and level is the author's estimate; the ordering of cases was formed on the same cases on which it is evaluated.
- Case 1: full text of the primary study not opened; abstract, highlights and legend of figure 1 checked verbatim, methods, discussion and limitations only via text excerpts (two check runs disagree on source status), n per group and numerical in-vivo effect sizes not found, behavioral and nine-month

statements from secondary sources. Case 3: derivation of the 43% from the study data not found; arXiv version history not checked (abstract page of v1 shows 37%, PDF 43%). Case 4: original 404 Media article paywalled.

- No measurement of reader effects.
- The 23andMe case relies on the 8-K filings (primary source) for the event and on AP and media reports for the court order and timeline; the Memorandum Opinion was read verbatim, the sale order (Doc. 910) was not.

10. Phase-2 design (preregistrable)

Question. Does the probability of claim inflation in science news depend more on verification cost (H1), timeliness gap (H4) or audience attractiveness (engagement, H3)?

Hypotheses and discriminating predictions.

Hypothesis	Prediction	Contradicts
H1 verification cost/sanction	overreach of the core claim rises with a verification-cost index, even at low attractiveness	H3
H3 engagement	overreach rises with attractiveness and goes with higher reach, independent of verification cost	H1
H4 time pressure	finality overstatement falls with distance to the primary event, independent of verification cost	none (complements H1)

Sample. The unit is the chain study/event → press release → news items. Stratification by domain (biomedical animal studies; nutrition/epidemiology; corporate/legal events; physics/astronomy as contrast), verification cost (low/high) and attractiveness (low/high). Size modeled on Sumner 2014/2016 (462/534 press releases), at least 500; power analysis for the interaction before registration. Outlets: aggregator-near portals and counterpart outlets with science desks, list justified in advance.

Coding by Sumner categories (advice, causal claim, animal→human), extended by cure/reversal verb in prevention or biomarker designs, finality (provisional/concluded), position of the first qualification, independent voice (yes/no). **Verification-cost index:** open access, link/DOI in the article, preprint vs. peer-reviewed, registry entry, readability of the abstract, machine retrievability. **Sanction proxies:** correction policy, corrections in the archive, expert commentary. **Engagement:** shares, Altmetric.

Reliability. Two independent coders for at least 20% of cases, Krippendorff's $\alpha \geq 0.80$ per variable, manual with anchor cases from §3, pilot round before registration.

Analysis. GEE or multilevel logit (news items within press releases within studies), predictors verification cost, timeliness gap, sanction proxies, attractiveness and interactions; second model reach ~ exaggeration (H3). **Preregistration** on OSF, option of a Registered Report.

Reusable data. InSciOut dataset for Sumner 2014 (figshare); supplement of Adams 2019 (figshare); CLAIMS (Haber et al. 2018; data and code metacausal.com/CLAIMS); Yu et al. 2020 (availability unchecked).

Limitations. Observational design; verification cost and attractiveness may correlate (biomedicine is attractive and paywalled); outlet selection affects effects; share data depend on the platform.

11. References

11.1 Checked in the original or abstract

- Fujii K, et al. (2025). Oral administration of arginine suppresses A β pathology in animal models of Alzheimer's disease. *Neurochem Int* 191:106082. doi:10.1016/j.neuint.2025.106082 · PMID 41175945 (abstract).
- Kindai University (21 Nov 2025). Press release. EurekAlert 1106825.
- In re 23andMe Holding Co., Case 25-40976 (Bankr. E.D. Mo.), Memorandum Opinion, Judge Brian C. Walsh, 27 Jun 2025: moeb.uscourts.gov/sites/moeb/files/opinions/25-40976_BK_01_23andMe.pdf (verified verbatim). Sale order Doc. 910: Kroll portal restructuring.ra.kroll.com/23andme (not opened).

White and Williams (30 Jun 2025): Bankruptcy Court Issues Order Approving 23andMe Sale (secondary).

- Sumner P, et al. (2014). The association between exaggeration in health related science news and academic press releases. *BMJ* 349:g7015. doi:10.1136/bmj.g7015 · PMC4262123.
- Sumner P, et al. (2016). Exaggerations and caveats in press releases and health-related science news. *PLOS ONE* 11(12):e0168217. doi:10.1371/journal.pone.0168217.
- Adams RC, et al. (2019). Claims of causality in health news: a randomised trial. *BMC Medicine* 17:91. doi:10.1186/s12916-019-1324-7.
- Adams RC, et al. (2017). How readers understand causal and correlational expressions used in news headlines. *J Exp Psychol Appl* 23(1):1-14. doi:10.1037/xap0000100.
- Bratton L, et al. (2019). *Wellcome Open Res* 4:148. doi:10.12688/wellcomeopenres.15486.1. — Bratton L, et al. (2020). Causal overstatements reduced in press releases following academic study of health news. *Wellcome Open Res* 5. doi:10.12688/wellcomeopenres.15647.2.
- Robertson CE, et al. (2023). Negativity drives online news consumption. *Nat Hum Behav* 7:812-822. doi:10.1038/s41562-023-01538-4.
- Thompson DF (1980). Moral responsibility of public officials: the problem of many hands. *APSR* 74(4):905-916. doi:10.2307/1954312.
- Sperber D, et al. (2010). Epistemic vigilance. *Mind & Language* 25(4):359-393. doi:10.1111/j.1468-0017.2010.01394.x.
- Akerlof GA (1970). The market for "lemons". *QJE* 84(3):488-500. doi:10.2307/1879431.
- Darby MR, Karni E (1973). Free competition and the optimal amount of fraud. *J Law Econ* 16(1):67-88. doi:10.1086/466756.
- Dulleck U, Kerschbamer R (2006). On doctors, mechanics, and computer specialists. *J Econ Lit* 44(1):5-42. doi:10.1257/002205106776162717.
- Haber N, et al. (2018). Causal and associational language in observational health research / CLAIMS. *PLOS ONE* 13(5):e0196346.
- Yu B, et al. (2020). Measuring correlation-to-causation exaggeration in press releases. COLING 2020, aclanthology.org/2020.coling-main.427.
- van Dyck CH, et al. (2023). Lecanemab in early Alzheimer's disease. *NEJM* 388:9-21. — Sims JR, et al. (2023). Donanemab in early symptomatic Alzheimer disease. *JAMA* 330:512-527.
- AP (1 Jul 2025). Anne Wojcicki's nonprofit gets court approval to buy 23andMe for \$305 million.
- 23andMe Holding Co., Form 8-K of 19 May 2025 (sec.gov, CIK 1804591, Ex. 99.1: "Transaction Subject to Court Approval") and Form 8-K of 13 Jun 2025 (Ex. 99.1: TTAM, USD 305 m, Regeneron backup bidder); Form 8-K on approval (Ex. 99.1, 27 Jun 2025).
- De Freitas J, Oğuz-Uğuralp Z, Kaan-Uğuralp A (2025). Emotional manipulation by AI companions. arXiv 2508.19258 v1/v2/v3.
- Wei M (22 Sep 2025). The Dark Side of AI Companions: Emotional Manipulation. *Psychology Today*.
- Engadget (2025). Moderators call for AI controls after Reddit Answers suggests heroin for pain relief.
- Prescribing information for arginine preparations (アルギン), KEGG/JAPIC: indication urea-cycle disorders.
- Futurism articles: URLs and reading depth in §2.1 and in the case register (JSON).

11.2 Real, checked only indirectly or in part

- Schat J, et al. (2018). Overdreven gezondheidsnieuws. *Ned Tijdschr Geneeskd* 162:D1936 (abstract; Leiden repository).
- Bott L, et al. (2019). *J Exp Psychol Appl* 25(4):517-542; Fang Z, et al. (2021). *JASIST*, doi:10.1002/asi.24458 (49.5% of 1.1 m links to scholarly publications never clicked); Michaelian (*Episteme*); Ecker UKH, et al. (2014). *J Exp Psychol Appl* 20(4):323-335 (core finding not checked against the abstract; not used).
- Fujii study: dose and behavior via lifespan.io (2 Dec 2025) and ScienceBlog (21 Nov 2025).

11.3 Removed from v0.1/v0.2

Shirley (2016); Zhang, Ghorbani & Zou (2024); DOI 10.1016/j.neuint.2025.105850; Gabielkov figure "59%" verbatim (the work exists; the number is supported only by secondary sources and carries no argument); citation URLs under tschueller.at.