

AGENCY 150 | WHITE PAPER

Text or Video? A Science-Based Case for Text-Based AI Interviewing of Employees and Consumers

A review of the evidence on participation, candor, trust, visual-cue value, and cost

July 2026

Lamar Pierce & Natasha Swift

Agency 150, Inc.

Executive summary

Organizations deploying AI-moderated interviews of employees, job applicants, or consumers face a design choice between text-based conversations (chat, messaging, conversational survey agents) and video-based ones (synchronous or asynchronous recorded video, often with automated facial and vocal analysis). Video is intuitively appealing: a face carries more information than a line of text, and richer media feel more “human.” This review weighs that intuition against the empirical record and reaches a clear conclusion for interview and research applications: the visual channel's incremental value is small and unreliable, while its costs to respondent participation, candor, and trust are high and well documented. For the purpose of collecting honest, analyzable qualitative data at scale, text-based AI interviewing is the stronger default.

The argument rests on five evidence pillars, each grounded in peer-reviewed research:

1. Participation and completion. In a controlled field study, a conversational (text) AI agent achieved a completion rate of 54% versus 24% for an equivalent web survey, more than double [Xiao et al. 2020]. Heavier-technology and richer channels consistently recruit and retain fewer respondents [Ritter et al. 2004; Shih & Fan 2008; Oates et al. 2022].
2. Candor. Removing a human (and a camera) from the loop reduces social-desirability pressure and increases disclosure of sensitive information [Tourangeau & Yan 2007; Gnambs & Kaspar 2015; Kreuter et al. 2008; Pickard et al. 2016].
3. Trust and fairness. Video and especially AI-analyzed video interviews reliably provoke more negative reactions, lower perceived fairness, and greater privacy and “creepiness” concerns than less intrusive modes [Langer et al. 2017; Langer et al. 2019; Brenner et al. 2016; Acikgoz et al. 2020].
4. The (weak) value of visual cues. Humans detect deception from behavior at only ~54% accuracy, barely above chance, and do *better* from audio than from visual cues [Bond & DePaulo 2006; DePaulo et al. 2003; Vrij et al. 2019]. Automated inference of emotion or character from faces lacks a sound scientific basis [Barrett et al. 2019], and a leading vendor withdrew facial analysis after concluding it added negligible predictive value at significant bias risk [HireVue 2021].

5. Cost. Video multiplies the cost of capture, bandwidth, transcription, captioning, and storage, and imposes a measurable cognitive burden on respondents [Bailey 2008; Bailenson 2021; Fauville et al. 2021; Oates et al. 2022].

Crucially, the strongest single test that directly pitted modes against one another found no honesty or data-equivalence advantage for video and located the real differences in logistics and cost [Oates et al. 2022]. We treat that finding not as a caveat to explain away but as the empirical core of the case.

1. Introduction

Agency 150's problem is one of integration: drawing reliable, total-population insight from a collection of qualitative studies whose questions, sub-populations, and instruments differ. The instrument-level choices made when data are *collected* determine how cleanly they can later be *pooled*. An interview mode that suppresses participation introduces selection bias; one that dampens candor introduces measurement bias; one that produces heterogeneous audio-video artifacts raises downstream analysis costs and lowers fidelity. Mode is therefore not a cosmetic decision; it propagates into every cross-study and cross-client synthesis the agency attempts.

This paper reviews the scientific evidence bearing on a specific mode question that recurs as AI interviewing tools mature: should interviews be conducted as text or as video? We restrict the scope to interviews and qualitative data collection (employees, applicants, consumers, research participants), not to entertainment or relationship-building contexts where the goals differ.

Our thesis is that, based on current evidence, the value added by the video channel is overwhelmed by drops in respondent participation, honesty, and trust, and further penalized by cost. We deliberately foreground the study that most challenges a naïve version of our thesis, a randomized comparison that found little mode effect on honesty [Oates et al. 2022], because, read carefully, it strengthens rather than weakens the practical conclusion: if video buys no honesty gain but costs substantially more and recruits worse, its case collapses.

Figure 1 previews the empirical core.

The empirical core: text wins participation; visual behaviour reveals little

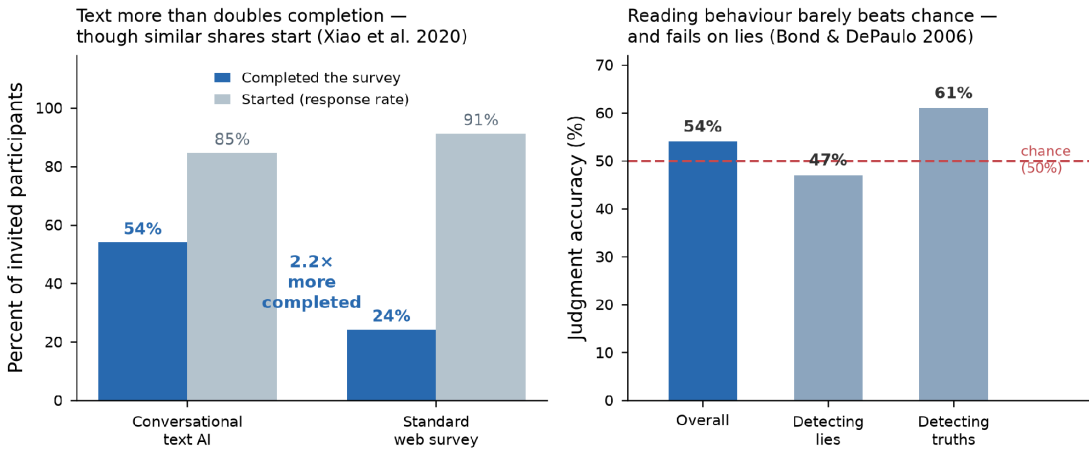


Figure 1. The empirical core in two anchors. Left: in a controlled field study, a conversational text AI more than doubled completion versus an equivalent web survey (54% vs 24%), even though similar shares of each group started (Xiao et al. 2020). Right: humans judge behavior and detect deception at only 54% overall, and just 47% for actual lies, below the 50% chance line, showing the visual/behavioral channel carries little diagnostic signal (Bond & DePaulo 2006).

2. Background and theoretical framing

2.1 What we mean by text- vs. video-based AI interviews

By text-based AI interview, we mean an interaction in which an AI agent poses questions and interprets free-text responses in a chat or messaging interface, including conversational survey agents that probe and follow up [Xiao et al. 2020]. By video-based AI interview, we mean recorded video responses (typically asynchronous), often paired with automated analysis of facial expression, gaze, and voice, the paradigm popularized by hiring platforms such as HireVue [HireVue 2021; Langer et al. 2019].

2.2 Two competing predictions from media theory

Media Richness Theory predicts that richer channels, those carrying more cues (facial expression, tone, immediacy), improve communication [Daft & Lengel 1986]. Taken naively, this favors video. But the theory's own logic is conditional: richness helps most for *equivocal, ambiguous* tasks, and can be counterproductive when it introduces evaluation apprehension or cognitive overload. Interview candor is precisely a case where more cues can *hurt*: a visible interviewer (or camera) raises the salience of being judged.

The Computers Are Social Actors (CASA) paradigm supplies the countervailing prediction [Nass et al. 1994]. People automatically apply human social scripts, politeness, reciprocity, and disclosure to computers, even knowing they are machines. A text agent therefore does not need a face to elicit genuine, socially engaged responses. What it lacks in richness, it may repay in reduced social threat: no human gaze to perform for, and no camera capturing one's image.

The synthesis these theories jointly support is that richness is not free. Each added cue channel also adds self-consciousness, surveillance, and processing cost. Whether richness nets positive depends entirely on the task, and for honest disclosure, the balance tips toward leaner media.

3. Participation and completion

The first and least contestable disadvantage of video is that fewer people will do it, and fewer of those who start will finish.

3.1 Text conversation raises completion

In the most directly relevant controlled study, Xiao et al. [2020] randomly assigned roughly 600 participants to complete either a standard Qualtrics web survey or a conversational survey conducted by an AI chatbot, using mostly open-ended questions. The chatbot's completion rate was 54%, versus 24.2% for the web survey, a 2.2× difference ($z = -12.16$, $p < 0.01$), and it drove significantly higher engagement and better response quality, judged on informativeness, relevance, specificity, and clarity. (The chatbot's initial *response* rate was marginally lower, 84.6% vs 91.2%, attributed to unfamiliarity, but of those who engaged, far more saw it through.) The conversational, text-based format did not merely match the survey; it retained more than twice as many respondents to completion.

3.2 Richer channels recruit worse

The general pattern in survey methodology is that each increment of technological or logistical burden depresses participation. A randomized comparison found an Internet questionnaire drew a 45% response rate versus 58% for a mailed version, a 13-point gap ($p < 0.01$), despite the Internet arm's apparent convenience [Ritter et al. 2004]. A meta-analysis of 39 studies similarly found web response rates about 10 points below mail [Shih & Fan 2008], and online response rates have continued to decline across the board [Wu et al. 2022]. Video is the extreme of this gradient: it demands a working camera, adequate bandwidth, a presentable environment, scheduling (for synchronous modes), and willingness to be recorded.

3.3 Direct mode comparison: video is the hardest to field

The cleanest test comes from Oates et al. [2022], who conducted 154 interviews on sensitive topics across seven randomly assigned conditions (video, audio, chat, non-anonymous chat, email, survey, scheduled survey). Their headline logistical finding was unambiguous: audio and video modes had substantially lower eligibility and completion rates and higher time and monetary costs than the leaner text modes. For an agency that must assemble adequately powered samples across many sub-populations, a mode that systematically loses eligible respondents is a mode that systematically biases the resulting picture of the total population.

4. Candor and honest disclosure

The second pillar concerns whether respondents tell the truth. Here the evidence is nuanced, and we present it as such.

4.1 The self-administration effect

A large body of survey-methodology research shows that self-administered modes, those without a human interviewer, elicit more honest reporting of sensitive or stigmatized behaviors and fewer socially desirable answers. Tourangeau and Yan's [2007] authoritative review concluded that the presence of an interviewer reliably increases misreporting on sensitive questions, and that self-administration mitigates it. Gnambs and Kaspar's [2015] meta-analysis of 460 effect sizes found that more computerized, less interviewer-mediated modes increased disclosure of sensitive behaviors. Kreuter et al. [2008] showed experimentally that web collection (no human voice) yielded higher disclosure and less social-desirability bias than telephone.

The mechanism matters for the text-vs-video question. The mechanism driven by *removing* the self-administration effect is *removing a judging human presence*. A video interview, especially one with a visible interlocutor or a camera capturing the respondent's face, reintroduces exactly the social-evaluative cues that suppress candor. Text, by contrast, preserves the low-threat, self-administered character that the literature associates with honesty.

4.2 People disclose to machines

Consistent with CASA, people will disclose freely to a computer, and sometimes *more* than to a human. Pickard et al. [2016] found participants revealed more sensitive information to a computer agent than to a human interviewer. Ho et al. [2018] showed that the emotional and relational benefits of self-disclosure were equivalent whether people believed they were confiding in a human or a chatbot; disclosing to a bot "worked." A text AI interviewer is thus a legitimate, and in some respects advantaged, recipient of candid disclosure.

4.3 The honest counter-evidence

The most important qualification comes from Oates et al. [2022], whose randomized design directly compared self-reported honesty and measured disclosure across modes on sensitive questions. They found little evidence of differences in self-reported honesty or social-desirability bias across modes, and no differences in disclosure, self-reported or measured. In other words, video did not make respondents *more* honest than text, but neither did text clearly beat video on candor. Participants were *aware* of mode and reported it shaped how much they were willing to share, but the aggregate honesty outcome did not move.

We take two lessons from this. First, intellectual honesty requires acknowledging that the candor advantage of text over video is better supported in the broad survey-mode literature (which contrasts human-interviewer vs. self-administered) than in the narrower set of head-to-head text-vs-video interview experiments, where it can attenuate. Second, and decisively for the practical argument, if video confers no candor benefit, then the participation, trust, and cost penalties documented in the other pillars have nothing to offset them. The counter-evidence removes video's best potential justification rather than rescuing it.

5. Trust, privacy, and perceived fairness

The third pillar is how respondents *feel* about the mode, which shapes not only whether they participate but also how guarded they are in their responses.

5.1 Video interviews provoke more negative reactions

The applicant-reactions literature is consistent and unfavorable to video. Langer et al. [2017] found that asynchronous (automated) video interviews produced more negative reactions, lower fairness perceptions, and higher “creepiness” and privacy concerns than more interactive formats. Brenner et al. [2016] reported that applicants experienced asynchronous video interviews as impersonal and as offering fewer opportunities to perform, lowering fairness perceptions relative to conventional interviews. The absence of a responsive human interlocutor, combined with awareness of being recorded, makes the format feel simultaneously exposed and unheard.

5.2 AI analysis of video amplifies distrust

When automation is layered onto video, reactions worsen. Langer et al. [2019] found that highly automated (AI) video interviews were seen as less fair and elicited lower acceptance and more discomfort than human-mediated interviews. Acikgoz et al. [2020] similarly found AI-driven selection perceived as less procedurally just, with less opportunity to perform, less interpersonal warmth. Knowing an algorithm is

scoring one's face and voice adds a distinct layer of surveillance concern that text interfaces do not trigger to the same degree.

5.3 Why trust matters for data quality, not just optics

Trust is not a public-relations footnote; it is an input to data quality. A respondent who feels surveilled or unfairly judged manages their self-presentation more tightly, the very social-desirability dynamic that Section 4 identified as the enemy of candor. Low trust and low candor are two views of the same mechanism. For a consumer-insights or employee-listening program, a mode that respondents distrust yields guarded, performed answers that are precisely the wrong raw material for honest total-population insight.

6. The purported value of visual cues, critically examined

The entire case *for* video rests on one premise: that seeing a respondent's face and body adds diagnostic value, that one can read honesty, emotion, or character from visual cues. The scientific evidence does not support this premise to the degree intuition assumes.

6.1 Nonverbal deception cues are faint and unreliable

DePaulo et al.'s [2003] meta-analysis of 158 putative cues to deception concluded that behavioral cues are, in aggregate, "faint and unreliable." Bond and DePaulo's [2006] meta-analysis of 206 documents and 24,483 judges found human deception-detection accuracy averages just 54%, barely above the 50% coin-flip, with a marked "veracity effect" (61% accuracy on truths but only 47% on lies). Critically for the mode question, people are *more* accurate when judging audible cues than visible ones: the visual channel is the least diagnostic. Vrij et al.'s [2019] authoritative review reaches the same verdict: no reliable nonverbal cue to deception exists, and observers should not rely on body language; verbal content outperforms it.

The implication is direct. If the visual channel is the weakest basis for judging veracity, weaker even than audio, and far weaker than the content of what is said, then a text interview, which captures verbal content in full, is not missing the important signal. It is discarding mostly noise.

6.2 Automated facial and emotion analysis lacks a scientific basis

The premise fares no better when the "reader" is an algorithm. Barrett et al.'s [2019] systematic review of more than a thousand studies concluded that emotion cannot be reliably inferred from facial configurations: a scowl is not diagnostic of anger, nor a smile of happiness, across people and contexts. Automated facial-expression

analysis therefore rests on a construct, the universal, readable emotional face, that the science does not support.

6.3 The industry's own verdict

The most telling evidence is revealed preference. HireVue, the highest-profile vendor of AI video interviewing, discontinued its facial-analysis feature in 2021, with reporting indicating the company concluded visual/nonverbal data contributed negligibly to predictions (on the order of a fraction of a percent in most cases) and was not worth the mounting bias and legal risk [HireVue 2021]. A privacy complaint to the U.S. Federal Trade Commission had alleged the tool produced biased, unprovable results. When the most committed commercial proponent of reading faces abandons reading faces, the “value-added” premise is in serious trouble.

6.4 Visual analysis actively introduces bias risk

Beyond adding little, the visual channel adds *harm*. AI video-analysis tools have been documented to disadvantage non-native speakers, people with disabilities, and other protected groups, generating discrimination exposure that text interfaces largely avoid [HireVue 2021; Barrett et al. 2019]. For an agency operating across many client organizations and demographic sub-populations, systematically biased instrumentation is not a peripheral risk; it corrupts the cross-population comparisons that are the whole point.

6.5 A fair steelman

To steelman the visual channel: “thin-slice” research shows that brief samples of nonverbal behavior *can* predict some interpersonal outcomes above chance [Ambady & Rosenthal 1992]. This is real, but narrow. Thin-slice accuracy concerns *impression and trait judgments* (e.g., perceived warmth or competence), not the *honesty of substantive content*, and it describes what human observers infer, not what current automated systems reliably extract. The steelman shows that faces carry *some* signal; it does not show that interviews need that signal, that machines can read it, or that it is worth the participation, trust, and cost penalties documented throughout this review.

Figure 2 consolidates the mode comparison across all six dimensions.

Text-based vs. video-based AI interviews across six decision dimensions

Dimension	Text-based AI Interview	Video-based AI Interview	Winner
Respondent participation & completion	Highest — chatbot completion 54% vs 24% for web survey	Lowest; higher recruitment friction & dropout	✓ Text
Candor / honest disclosure	Equal or higher; less social-desirability pressure (no face)	No measured gain over text (Oates 2022); possible chilling	= Tie
Trust, fairness & privacy comfort	Highest; low 'creepiness', self-paced, anonymous-feeling	Lowest; surveillance concern, algorithmic bias, low fairness	✓ Text
Value of visual cues (evidence)	N/A — relies on verbal content, the more diagnostic channel	Weak: 54% deception accuracy; facial-emotion inference invalid	✓ Text*
Cost to field, process & store	Lowest; text is analysis-ready, no transcription/storage	Highest; capture, bandwidth, transcription, captioning, storage	✓ Text
Analytic integration across studies	Easiest; uniform text corpus, directly poolable	Hardest; heterogeneous A/V, costly to normalize	✓ Text

*Visual cues carry weak, unreliable diagnostic value for honesty/emotion; text preferred on net. Sources: Xiao et al. 2020; Oates et al. 2022; Bond & DePaulo 2006; Barrett et al. 2019.

Figure 2. Text-based vs. video-based AI interviews across six decision dimensions. Text leads on participation, trust, cost, and analytic integration; the two tie on candor in the strongest head-to-head test; and video's putative advantage, visual cues, is empirically weak and bias-prone.

7. Cost structures

The fourth pillar is cost, which compounds the case. Video is more expensive at every stage of the data lifecycle and imposes hidden costs on the respondent.

7.1 Direct production and processing costs

Text arrives analysis-ready. Video and audio must be captured, transmitted (bandwidth), stored (at far larger file sizes), and, for qualitative analysis, transcribed and often captioned for accessibility. Transcription alone is a major line item: transcribing one hour of recorded interview takes on the order of 3–10 hours of skilled labor, and professional transcription commonly runs roughly \$90–150 per recorded hour [Bailey 2008]. Automated speech-to-text lowers but does not eliminate this, since accuracy on real-world audio still requires human correction. All of these steps are absent when the interview is conducted as text.

7.2 The direct-comparison cost finding

This is not merely theoretical. Oates et al. [2022] measured this: audio and video modes incurred higher time and monetary costs than text modes, along with lower completion rates, a double penalty of paying more to collect less.

7.3 The respondent's cognitive cost

Video also taxes respondents in ways text does not. Bailenson's [2021] "nonverbal overload" framework identifies specific mechanisms by which videoconferencing

fatigues people: mirror anxiety from constant self-view, restricted physical mobility, and the elevated cognitive load of producing and monitoring nonverbal cues. Fauville et al. [2021] developed and validated a measure of this fatigue, showing that camera-on and self-view are principal drivers. Riedl et al. [2023] demonstrated that videoconference fatigue is detectable neurophysiologically (EEG/ECG), not merely self-reported. A fatigued, self-conscious respondent is a worse source of thoughtful, candid data, so the cognitive cost of video feeds back into the candor and quality penalties of the earlier pillars.

8. Synthesis: weighing the ledger

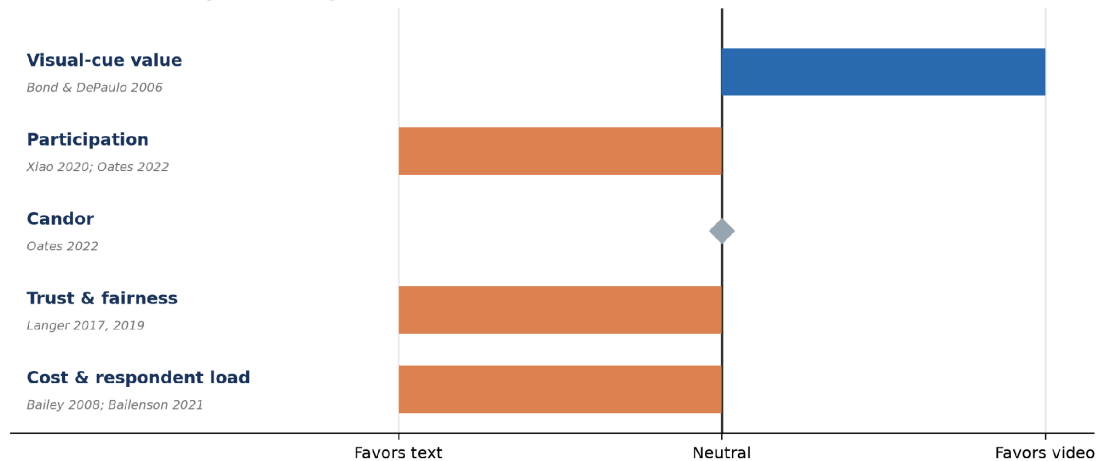
Assembling the pillars produces a lopsided ledger.

On the benefit side, video offers the visual channel. The evidence values that channel at close to zero for an interview: visual cues are the *least* diagnostic basis for judging honesty [Bond & DePaulo 2006; Vrij et al. 2019]; automated reading of faces for emotion or character lacks scientific support [Barrett et al. 2019]; and the leading vendor abandoned it as adding negligible value at high risk [HireVue 2021]. The one genuine positive, thin-slice impression signal [Ambady & Rosenthal 1992] is narrow, human-not-machine, and irrelevant to substantive candor.

On the cost side, video reduces participation and completion [Xiao et al. 2020; Ritter et al. 2004; Oates et al. 2022], reintroduces the social-evaluative presence that suppresses candor [Tourangeau & Yan 2007; Gnambs & Kaspar 2015], lowers trust and perceived fairness while raising surveillance and bias concerns [Langer et al. 2017; Langer et al. 2019; Acikgoz et al. 2020], multiplies direct production cost [Bailey 2008; Oates et al. 2022], and imposes a measurable cognitive burden on respondents [Bailenson 2021; Fauville et al. 2021; Riedl et al. 2023].

The key structural point is that these are not independent line items to be netted arithmetically; they compound. Lower participation biases the sample; lower candor biases the measurement; lower trust deepens both; higher cost caps the achievable sample size; and none of it is redeemed by the visual channel, because that channel carries little usable signal. Figure 3 renders the ledger: a single, weak advantage for video on the right, set against four dimensions on which the peer-reviewed evidence favors text or finds no difference.

**The evidence ledger: one weak advantage for video,
four documented penalties against it**



Each row shows the direction of the peer-reviewed evidence — which mode it favors — not a measured effect size. Bars are not summed.

Figure 3. The evidence ledger. Each row shows the direction of the peer-reviewed evidence points, which mode it favors, across the five dimensions of this review. Video enjoys one advantage (visual cues), and that advantage is itself weak and unreliable; on participation, trust, and cost the evidence favors text, and on candor it finds no difference. Bars indicate direction, not measured effect sizes, and are not summed.

This is the science-based case in one sentence: for interviewing employees and consumers, video adds little value and is undermined by measurable drops in participation, honesty, and trust, before the higher cost is even counted.

9. Limitations and research gaps

Several honest limitations bound this review. First, much of the strongest candor evidence contrasts *human-interviewer* with *self-administered* modes rather than *text-AI* with *video-AI* directly; we infer the text-vs-video implication from the shared mechanism (removal of a judging presence), but direct head-to-head experiments remain sparse, and the best one found the candor difference small [Oates et al. 2022]. Second, several trust and reactions studies come from the *personnel-selection* context, where stakes are high; consumer-research stakes differ, and reactions may be milder, though the direction of effect is consistent. Third, the visual-cue literature we rely on concerns human and current-generation automated perception; we cannot rule out that future systems extract more usable signal from video, though Barrett et al.'s [2019] critique targets the underlying construct, not merely present technology, which makes dramatic improvement unlikely. Fourth, the evidence ledger in Figure 3 summarizes the *direction* of evidence for each dimension, not measured effect sizes; its bars are not summed, and it should be read as a decision aid, not a measurement. Priority research gaps include large-sample,

pre-registered text-AI-vs-video-AI interview experiments measuring participation, verified disclosure, trust, and cost jointly on the same respondents, in consumer as well as hiring contexts.

10. Conclusion

Video interviewing appeals because a face seems to carry more truth than text. The science says otherwise. The visual channel is the *least* reliable basis for judging honesty, automated facial analysis rests on a discredited construct, and even the industry's leading proponent abandoned it as adding almost nothing. Against that thin, uncertain benefit stand large, compounding, well-documented costs: fewer people participate and finish; the reintroduced sense of being watched dampens candor; trust and perceived fairness fall; and the price of collecting, processing, and storing video rises, while the respondent bears an added cognitive burden. For interviewing employees and consumers, and especially for an agency whose value lies in integrating honest qualitative data across many studies and clients, the conclusion is clear and evidence-based: text-based AI interviewing should be the default, and the value added by video is overwhelmed by the drop in participation, honesty, and trust.

References

- Acikgoz, Y., Davison, K. H., Compagnone, M., & Laske, M. (2020). Justice perceptions of artificial intelligence in selection. *International Journal of Selection and Assessment*, 28(4), 399–416. <https://doi.org/10.1111/ijsa.12306>
- Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin*, 111(2), 256–274. <https://doi.org/10.1037/0033-2909.111.2.256>
- Bailenson, J. N. (2021). Nonverbal overload: A theoretical argument for the causes of Zoom fatigue. *Technology, Mind, and Behavior*, 2(1). <https://doi.org/10.1037/tmb0000030>
- Bailey, J. (2008). First steps in qualitative data analysis: Transcribing. *Family Practice*, 25(2), 127–131. <https://doi.org/10.1093/fampra/cmn003>
- Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological Science in the Public Interest*, 20(1), 1–68. <https://doi.org/10.1177/1529100619832930>

- Bond, C. F., Jr., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review*, 10(3), 214–234. https://doi.org/10.1207/s15327957pspr1003_2
- Brenner, F. S., Ortner, T. M., & Fay, D. (2016). Asynchronous video interviewing as a new technology in personnel selection: The applicant's point of view. *Frontiers in Psychology*, 7, 863. <https://doi.org/10.3389/fpsyg.2016.00863>
- Daft, R. L., & Lengel, R. H. (1986). Organizational information requirements, media richness, and structural design. *Management Science*, 32(5), 554–571. <https://doi.org/10.1287/mnsc.32.5.554>
- DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, 129(1), 74–118. <https://doi.org/10.1037/0033-2909.129.1.74>
- Fauville, G., Luo, M., Queiroz, A. C. M., Bailenson, J. N., & Hancock, J. (2021). Zoom Exhaustion & Fatigue Scale. *Computers in Human Behavior Reports*, 4, 100119. <https://doi.org/10.1016/j.chbr.2021.100119>
- Gnambs, T., & Kaspar, K. (2015). Disclosure of sensitive behaviors across self-administered survey modes: A meta-analysis. *Behavior Research Methods*, 47(4), 1237–1259. <https://doi.org/10.3758/s13428-014-0533-4>
- Ho, A., Hancock, J., & Miner, A. S. (2018). Psychological, relational, and emotional effects of self-disclosure after conversations with a chatbot. *Journal of Communication*, 68(4), 712–733. <https://doi.org/10.1093/joc/jqy026>
- Kreuter, F., Presser, S., & Tourangeau, R. (2008). Social desirability bias in CATI, IVR, and web surveys: The effects of mode and question sensitivity. *Public Opinion Quarterly*, 72(5), 847–865. <https://doi.org/10.1093/poq/nfn063>
- Langer, M., König, C. J., & Krause, K. (2017). Examining digital interviews for personnel selection: Applicant reactions and interviewer ratings. *International Journal of Selection and Assessment*, 25(4), 371–382. <https://doi.org/10.1111/ijsa.12191>
- Langer, M., König, C. J., & Papathanasiou, M. (2019). Highly automated job interviews: Acceptance under the influence of stakes. *International Journal of Selection and Assessment*, 27(3), 217–234. <https://doi.org/10.1111/ijsa.12246>
- Maurer, R. (2021). HireVue discontinues facial analysis screening. *Society for Human Resource Management (SHRM)*; see also Electronic Privacy Information Center, *In re HireVue* (FTC complaint, 2019). <https://www.shrm.org/topics-tools/news/talent-acquisition/hirevue-discontinues-facial-analysis-screening>

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '94)*, 72–78. <https://doi.org/10.1145/191666.191703>

Oates, M., Crichton, K., Cranor, L., Budwig, S., Weston, E. J. L., Bernagozzi, B. M., & Pagaduan, J. (2022). Audio, video, chat, email, or survey: How much does online interview mode matter? *PLOS ONE*, 17(2), e0263876. <https://doi.org/10.1371/journal.pone.0263876>

Pickard, M. D., Roster, C. A., & Chen, Y. (2016). Revealing sensitive information in personal interviews: Is self-disclosure easier with humans or avatars? *Computers in Human Behavior*, 65, 23–30. <https://doi.org/10.1016/j.chb.2016.08.004>

Riedl, R., Kostoglou, K., Wriessnegger, S. C., & Müller-Putz, G. R. (2023). Videoconference fatigue from a neurophysiological perspective: Experimental evidence based on electroencephalography (EEG) and electrocardiography (ECG). *Scientific Reports*, 13, 18371. <https://doi.org/10.1038/s41598-023-45374-y>

Ritter, P., Lorig, K., Laurent, D., & Matthews, K. (2004). Internet versus mailed questionnaires: A randomized comparison. *Journal of Medical Internet Research*, 6(3), e29. <https://doi.org/10.2196/jmir.6.3.e29>

Shih, T.-H., & Fan, X. (2008). Comparing response rates from web and mail surveys: A meta-analysis. *Field Methods*, 20(3), 249–271. <https://doi.org/10.1177/1525822X08317085>

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, 133(5), 859–883. <https://doi.org/10.1037/0033-2909.133.5.859>

Vrij, A., Hartwig, M., & Granhag, P. A. (2019). Reading lies: Nonverbal communication and deception. *Annual Review of Psychology*, 70, 295–317. <https://doi.org/10.1146/annurev-psych-010418-103135>

Wu, M.-J., Zhao, K., & Fils-Aime, F. (2022). Response rates of online surveys in published research: A meta-analysis. *Computers in Human Behavior Reports*, 7, 100206. <https://doi.org/10.1016/j.chbr.2022.100206>

Xiao, Z., Zhou, M. X., Liao, Q. V., Mark, G., Chi, C., Chen, W., & Yang, H. (2020). Tell me about yourself: Using an AI-powered chatbot to conduct conversational surveys with open-ended questions. *ACM Transactions on Computer-Human Interaction*, 27(3), Article 15. <https://doi.org/10.1145/3381804>