

From Visual Cues to Audience Response

How multimodal AI models emotion

For research buyers and creative teams evaluating emotional readings.

CONTENTS

1. Audience emotion is measurable, and it follows patterns
2. The cues that move an audience are on the screen and in the soundtrack
3. Who is watching changes what they feel
4. Context and memory make the audience specific
5. A small panel's moment-by-moment response predicts a large audience
6. Built to be checked
7. Client history adds a calibration layer
8. Where to use it today
9. Conclusion

Executive summary

Three decades of research show that audience emotion is not a mystery. Film clips reliably produce specific, measurable feelings in viewers. Those feelings follow from cues on the screen and in the soundtrack: the face, the cut, the music. They also depend on who is watching. Personality, mood and circumstance change what a person takes from the same scene.

FishDog builds both halves of that finding into one system. A multimodal model watches the material itself, in short windows for short work and in eight-minute segments for a feature. A synthetic audience recruited from a digital twin of the US population then responds to it. Each viewer carries a real demographic profile, a personality, a zip code, current news and a memory of what they watched.

The output is an emotional reading of a specific audience, moment by moment, with any disagreement visible and traceable.

Neuroscience studies found that a small group's moment-by-moment response predicts how a much larger audience reacts: viewer synchrony tracked the ratings and Twitter response to TV episodes and Super Bowl ads (Dmochowski et al., 2014), and brain responses to trailers predicted box office beyond what viewers said they liked (Boksem & Smidts, 2015; Boksem et al., 2025).

FishDog produces the same kind of moment-by-moment trace within a day, where a recruited screening takes weeks.

Every run is recorded so it can be checked: the exact cut, the model configuration and the panel. A buyer can test the reading against their own human data.

1. Audience emotion is measurable, and it follows patterns

Researchers have used film to produce emotion in the lab since the 1990s, because film works.

Gross and Levenson showed in 1995 that short clips reliably produce discrete emotions such as amusement, sadness and fear, and that viewers report them consistently (Gross & Levenson, 1995). Later databases extended this to dozens of clips and thousands of ratings (Schaefer et al., 2010). Cowen and Keltner had people rate 2,185 short videos and found 27 distinct categories of emotion, connected by smooth gradients rather than hard borders (Cowen & Keltner, 2017).

Engagement is measurable at the level of the brain. When people watch the same film, their brain activity moves in step. Hasson and colleagues found that the more tightly a film controls attention, the more closely viewers' brains synchronise (Hasson et al., 2008). Well-made film produces a shared response, and that shared response can be measured second by second.

An emotional reading still has to say what it measures. Consider a scene in which a character smiles while leaving an empty house. The character may be portraying relief. The lighting and music may suggest melancholy. One viewer may feel hope; another may feel loss. Psychologists separate the emotion a viewer *perceives* in a scene from the emotion they *feel*, and the two often diverge (Gabrielsson, 2001).

Exhibit 1. Four meanings of an emotional reading.

READING	MEANING
Portrayed emotion	What a character appears to express.
Intended mood	The emotional effect the creative treatment appears designed to produce.
Viewer-felt response	What a person reports experiencing.
Modeled audience response	A prediction of viewer-felt response for a defined audience, given the material and the audience information supplied.

FishDog reports the fourth, kept separate from the first two.

In humans, felt response is rarely one thing. People can feel happy and sad at the same time, and the mix is common in film (Larsen et al., 2001). Sad films are among the most enjoyed (Oliver, 1993). A low pleasantness score does not automatically mean a scene should go.

A useful reading therefore reports the distribution of responses, including mixed and minority ones, against a specific question: which audience, which cut, which measure, which decision.

2. The cues that move an audience are on the screen and in the soundtrack

Emotional response is not random. It tracks what the filmmaker puts in front of the viewer.

- **Editing.** In the Kuleshov effect, the same neutral face reads as hungry, grieving or desiring depending on the shot cut next to it. Brain-imaging work confirmed that context shots change the emotion viewers attribute to the face (Mobbs et al., 2006), and a 2016 replication found the effect holds (Barratt et al., 2016).
- **Music.** The same ambiguous scene is understood differently depending on the soundtrack, and viewers remember different details as a result (Boltz, 2001).
- **Pace.** Shot lengths in Hollywood film have evolved toward rhythms that match the natural fluctuation of human attention (Cutting et al., 2010).
- **Signal.** Motion, cut rate and sound energy can be computed from the video itself and combined into an arousal curve that follows the viewer's emotional intensity across a film

(Hanjalic & Xu, 2005).

Because these cues are visible and audible, a model that receives the material can read them. Machine learning has spent a decade connecting images to language and to human-reported feeling.

CLIP learned visual concepts from 400 million image-text pairs (Radford et al., 2021). ArtEmis collected people's emotional reactions to artworks, with explanations, and trained models to produce both (Achlioptas et al., 2021). LIRIS-ACCEDE and COGNIMUSE recorded viewers' felt valence and arousal across film excerpts and Hollywood movies, making induced emotion a measurable target (Baveye et al., 2015; Zlatintsi et al., 2017). Valence is how pleasant an experience feels; arousal is how activating it is.

Current multimodal models already read emotion from video in a way that resembles people. When Gemini and GPT models rated the kind of video clips Cowen and Keltner used, their ratings reproduced the structure of human emotion at the category level, with weaker agreement clip by clip (Asanuma et al., 2025).

Vision-language models predicting how people feel about still images correlated with human ratings above 0.75 (Nowicki et al., 2026).

How FishDog watches

For the moment-by-moment trace, every synthetic viewer watches the video itself, with its sound, one checkpoint window at a time: as short as one second on a short clip, 30 seconds on a feature. Each viewer carries a one-line memory of its own previous reaction and nothing from any other viewer, so two viewers can attend to different things. For survey questions on anything up to 20 minutes, a model first writes neutral notes on what the video shows, and each viewer answers in persona from those notes.

A feature film is handled the way a human focus group handles it: from memory. Before any viewer answers, a model watches the film end to end in eight-minute segments and is barred from describing anything it cannot see in the segment in front of it. The segment notes are combined into a structured record of acts, beats, characters, themes, pacing and likely attention dips. Every claim in that record cites the segment it came from. The ending is then re-watched and checked against the record; where they disagree, the footage wins.

Segmenting is deliberate: multimodal models lose accuracy as video gets longer (Fu et al., 2025), and eight minutes keeps each pass inside the range where they read reliably. Nobody re-watches a two-hour film to answer a survey. FishDog's memory of the film is written down and checkable.

3. Who is watching changes what they feel

The same scene lands differently on different people, and the differences follow patterns. Entertainment preferences across music, film, television and books fall into a few stable patterns that relate to personality and demographics (Rentfrow et al., 2011). People choose and respond to entertainment according to their current mood (Zillmann, 1988). Whether a viewer enjoys a sad film depends on who they are (Oliver, 1993).

An average across a national sample hides those differences. A thriller enthusiast and an occasional viewer recognise the same suspense cue and respond to it differently. Audience research that matters for a creative decision has to model the intended audience.

How FishDog builds the audience

FishDog constructs a synthetic population before any brief arrives.

The frame comes from US Census and American Community Survey microdata, calibrated to about 1 percent average error against selected census marginals. That frame is then enriched from more than 40 additional sources covering role, location, local conditions and personality. Every person in the population carries a zip code, a household, an occupation and an income, and a five-factor (OCEAN) personality. The personality scores are synthetic: the system generates correlated trait scores, adjusts them for demographic and occupational characteristics, and calibrates the resulting distribution against published Big Five research for US adults (Soto & John, 2017). They are not observed personality measurements of the individual people. Each response is conditioned on that profile.

The population is connected to the world. News is collected from more than 350 sources and summarised hourly, then matched to each person using their location, occupation, interests and broader profile. Stories are ranked by geographic plausibility, fit with the person's likely sources and topics, story importance and diversity. Political context, values and cultural traits contribute to the matching indirectly. Each person's personalised news is updated daily, and an industry briefing keyed to their occupation is refreshed weekly. Each person also gets local weather.

A study then recruits its audience from that population against the brief, rather than generating respondents to order. The methodology separates attributes grounded directly in population data from modeled enrichments, and says which is which (Gales & Duess, 2026).

The research on simulating people with language models explains why this design matters:

- **Grounded simulation works.** Argyle and colleagues showed that a language model conditioned on the backstories of real survey respondents reproduces the response patterns of those subgroups, which they called algorithmic fidelity (Argyle et al., 2023). Park and colleagues found that agents grounded in rich self-reports predict individuals' survey answers far better than agents given demographics alone (Park et al., 2026). Giving models detailed

life backstories brought simulated responses up to 18 percent closer to real survey distributions (Moon et al., 2024).

- **Simulation predicts real results.** Model-simulated respondents predicted the results of 70 preregistered survey experiments with a correlation of 0.85, including studies published after the model's training cutoff (Hewitt et al., 2024). In marketing, model-generated brand perception maps closely matched maps built from human surveys (Li et al., 2024).
- **Thin profiles fail.** Demographic labels on their own explain little of how people answer subjective questions (Hu & Collier, 2024). Models prompted with identity labels alone flatten groups into stereotypes and understate the variety within them (Wang et al., 2025). Pew's September 2026 study found large errors when a model answered surveys from demographic and attitudinal profiles (Chapekis et al., 2026).

The pattern across these studies is consistent. Simulation succeeds when it is grounded in real population data and rich individual profiles, and fails when it leans on a label. FishDog's architecture is built on the first and designed against the second: a population from microdata, recruitment rather than creation, rich profiles, and reports that keep the disagreement.

4. Context and memory make the audience specific

People do not watch in a vacuum. Mood, circumstance and what they already know shape the response.

Context

Each FishDog profile receives news selected for that person and updated daily, a weekly industry briefing for their occupation, local weather and a persistent life context. A viewer in Phoenix and a viewer in Buffalo do not walk into the same week. A household on a tighter entertainment budget can like a trailer and still be less likely to buy a ticket. Context is supplied per person, not as one bundle of headlines handed to everyone.

Context should change a reading only when it would change a person's response. The research supports that discipline. A study found that weather shifted people's reported life satisfaction (Schwarz & Clore, 1983).

Memory

FishDog viewers remember what they watched and what they said. A panel can be asked an hour later what stayed with them, which is how recall works in a human screening.

Prior knowledge changes experience. One study found that spoilers made stories more enjoyable (Leavitt & Christenfeld, 2011); a larger replication found people enjoyed unspoiled stories more (Johnson & Rosenbaum, 2015). Either way, a viewer who knows what happens is in a different experiment. FishDog therefore treats first exposure and repeat exposure as different study conditions, chosen by the client before the run.

5. A small panel's moment-by-moment response predicts a large audience

The strongest scientific support for a moment-by-moment emotional reading comes from neuroscience studies of audiences.

- **TV and advertising.** Dmochowski and colleagues recorded the brain activity of a small group watching TV episodes and Super Bowl ads. How closely the viewers' responses moved together moment to moment predicted the ratings and Twitter activity of thousands of viewers, and predicted them better than the participants' own ratings did (Dmochowski et al., 2014).
- **Movie trailers.** Boksem and Smidts found that brain responses to trailers predicted both individual preferences and the films' box office performance, beyond what stated preferences predicted (Boksem & Smidts, 2015). A 2025 follow-up across five datasets and 161 trailers confirmed that neural response predicts box office beyond movie characteristics and liking, with one measure, gamma-band activity, carrying the effect; synchrony between viewers did not predict box office in that data (Boksem et al., 2025).
- **Public messaging.** Falk and colleagues found that brain responses in a small group predicted which anti-smoking campaign drove the most calls to a quitline across the population, where the group's own ratings got the order wrong (Falk et al., 2012).
- **Ad response.** Facial expressions recorded while people watched online ads predicted whether they kept watching (Teixeira et al., 2012), and large-scale facial coding predicted ad liking and purchase intent (McDuff et al., 2015).
- **Story shape.** The emotional arcs of 1,300 novels cluster into six basic shapes (Reagan et al., 2016). In film, the "man in a hole" arc, a fall followed by a rise, earned the highest box office regardless of genre and budget (Del Vecchio et al., 2021).

Two lessons follow. The first is that the shape of a response over time carries signal that a single end-of-screening score misses. The second is that this signal can be read from a small panel and projected to a population.

These studies needed a lab, an EEG cap or an fMRI scanner, and weeks of recruitment. FishDog produces the same kind of output, a moment-by-moment engagement and emotion trace across a defined audience, in a single run that completes within a day. The neuroscience

literature also hands buyers a ready benchmark: a human time course for the same material, against which the FishDog trace can be scored.

6. Built to be checked

A reading is only as useful as a buyer's ability to check it. FishDog's measurement system is built around that.

- **Every run has an identity.** The instrument is fingerprinted: a hash covers the model, its settings, the prompt contract and the response schema. Change any of them and the fingerprint changes. A panel is identified by a hash of its exact members. Two runs are compared only when their instrument, sampling and panel match.
- **Every viewer is independent.** No viewer sees another viewer's answers. The moment-by-moment trace and the survey are separate measurements of the same material, so when they agree, the agreement is evidence rather than an echo.
- **Metrics are tested before they ship.** A research test suite checks every engagement metric against series with known answers, confirms results are identical when viewers are reordered, and expresses every timing rule in seconds so a 30-second ad and a two-hour feature sit on the same scale. Split-half reliability is measured on every metric family.
- **Calibration is published.** FishDog ran its screening system on 12 well-known film scenes with 12 synthetic viewers and published the results: nine matches and three partial matches to the expected interpretation, along with the failures found and fixed on the way (FishDog, n.d.).
- **Response style is measured and corrected.** Synthetic viewers, like people, differ in how high they score everything. Across 13 titles watched by the same 12 viewers, the viewer explained 52 percent of the variation in average engagement, the title 33 percent and their interaction 15 percent; a viewer's average on one film correlated 0.89 with their average on an unrelated film. The shape of a response over time and the ranking of moments within one run are unaffected. Comparisons of absolute levels across titles or audiences run on a fixed reference panel of 100 viewers, balanced to ACS microdata, with each viewer's stored offset removed.

Well-known scenes carry a known risk. Language models memorise books that appear often on the web, and perform much better on those books in downstream tasks (Chang et al., 2023). Famous films carry the same risk. FishDog's system answers from the material supplied: the segment rule bars the film record from describing anything not visible in the footage, and the decisive tests use material the model has never seen.

7. Client history adds a calibration layer

A studio or network holds years of creative material with known outcomes: completion rates, recommendation, sales. Run through FishDog with the right permissions, that archive calibrates readings to the client's own audiences and decisions. A reading that distinguishes the titles that performed from the ones that did not, on held-out titles, earns a level of trust no general benchmark can provide.

Commercial outcomes need care. Marketing spend, release timing, franchise familiarity and competition all move results independently of how an audience felt (Eliashberg et al., 2007). A client calibration therefore compares FishDog against the forecasting the client already uses, and measures what the emotional reading adds.

This layer is offered as a client engagement, not a standing feature, and depends on the client's rights to its material and outcome data.

8. Where to use it today

- **Before a recruited screening:** to find where attention dips and which moments split the audience, so human research time goes to the questions that matter. Marketing researchers place synthetic samples here, in pretesting and pilots, ahead of the main study (Sarstedt et al., 2024).
- **Between cuts:** to compare versions against the same recruited audience and see which moments moved.
- **Across audiences:** to see how a core fan and an occasional viewer read the same scene, without fielding two panels.
- **At acquisition:** to get an audience reading on a finished film within a day, when the decision window is days, not weeks.
- **For recall:** to ask what an audience remembers an hour after watching.

A reading supports these decisions directly. Revenue forecasts and claims about every member of an audience need paired validation against human data first.

9. Conclusion

Audience emotion has been measured in labs for 30 years. It follows from cues on the screen and from who is watching, and a small group's moment-by-moment response predicts how a population responds. FishDog puts those findings into a system that watches the material, recruits a defined audience from a digital twin of the US population, and returns a moment-by-moment reading within a day.

Every input is recorded, every viewer answers independently, and every run can be tested against people. That is the standard a buyer should hold any emotional reading to: say whose response is modeled, show what the system received, keep the disagreement, and prove it on the buyer's own material.

References

Academic studies explain the science and the feasibility of the approach; they are not independent validations of FishDog.

1. Achlioptas, P., Ovsjanikov, M., Haydarov, K., Elhoseiny, M., & Guibas, L. (2021). [ArtEmis: Affective language for visual art](#). CVPR 2021.
2. Argyle, L. P., Busby, E. C., Fulda, N., Gubler, J. R., Rytting, C., & Wingate, D. (2023). [Out of one, many: Using language models to simulate human samples](#). *Political Analysis*, 31(3), 337–351.
3. Asanuma, H., Koide-Majima, N., Nakamura, K., Horii, T., Nishimoto, S., & Oizumi, M. (2025). [Correspondence of high-dimensional emotion structures elicited from video clips between humans and multimodal LLMs](#). *Scientific Reports*, 15, 32175.
4. Barratt, D., Rédei, A. C., Innes-Ker, Å., & van de Weijer, J. (2016). [Does the Kuleshov effect really exist?](#) *Perception*, 45(8), 847–874.
5. Baveye, Y., Dellandréa, E., Chamaret, C., & Chen, L. (2015). [LIRIS-ACCEDE: A video database for affective content analysis](#). *IEEE Transactions on Affective Computing*, 6(1), 43–55.
6. Boksem, M. A. S., & Smidts, A. (2015). [Brain responses to movie trailers predict individual preferences for movies and their population-wide commercial success](#). *Journal of Marketing Research*, 52(4), 482–492.
7. Boksem, M. A. S., van Diepen, R., Eijlers, E., Boekel, W., & Smidts, A. (2025). [Do EEG metrics derived from trailers predict the commercial success of movies? A systematic analysis of five independent datasets](#). *Journal of Marketing Research*, 62(4), 703–720.
8. Boltz, M. G. (2001). [Musical soundtracks as a schematic influence on the cognitive processing of filmed events](#). *Music Perception*, 18(4), 427–454.
9. Chang, K. K., Cramer, M., Soni, S., & Bamman, D. (2023). [Speak, memory: An archaeology of books known to ChatGPT/GPT-4](#). EMNLP 2023.
10. Chapekis, A., et al. (2026). [Can AI stand in for human survey-takers? Not really](#). Pew Research Center, 30 September.
11. Cowen, A. S., & Keltner, D. (2017). [Self-report captures 27 distinct categories of emotion bridged by continuous gradients](#). *PNAS*, 114(38), E7900–E7909.
12. Cutting, J. E., DeLong, J. E., & Nothelfer, C. E. (2010). [Attention and the evolution of Hollywood film](#). *Psychological Science*, 21(4), 432–439.
13. Del Vecchio, M., Kharlamov, A., Parry, G., & Pogrebna, G. (2021). [Improving productivity in Hollywood with data science: Using emotional arcs of movies to drive product and service innovation](#). *Journal of the Operational Research Society*, 72(5), 1110–1137.
14. Dmochowski, J. P., Bezdek, M. A., Abelson, B. P., Johnson, J. S., Schumacher, E. H., & Parra, L. C. (2014). [Audience preferences are predicted by temporal reliability of neural processing](#). *Nature Communications*, 5, 4567.
15. Eliashberg, J., Hui, S. K., & Zhang, Z. J. (2007). [From story line to box office: A new approach for green-lighting movie scripts](#). *Management Science*, 53(6), 881–893.
16. Falk, E. B., Berkman, E. T., & Lieberman, M. D. (2012). [From neural responses to population behavior: Neural focus group predicts population-level media effects](#). *Psychological Science*, 23(5), 439–445.
17. FishDog. (n.d.). [How we calibrated Screening Room with famous film scenes](#).

18. Fu, C., et al. (2025). Video-MME: The first-ever comprehensive evaluation benchmark of multi-modal LLMs in video analysis. CVPR 2025.
19. Gabrielsson, A. (2001). Emotion perceived and emotion felt: Same or different? *Musicae Scientiae*, 5(1 suppl), 123–147.
20. Gales, P., & Duess, A. (2026). Synthetic populations: How FishDog builds, calibrates, and validates them.
21. Gross, J. J., & Levenson, R. W. (1995). Emotion elicitation using films. *Cognition & Emotion*, 9(1), 87–108.
22. Hanjalic, A., & Xu, L.-Q. (2005). Affective video content representation and modeling. *IEEE Transactions on Multimedia*, 7(1), 143–154.
23. Hasson, U., Landesman, O., Knappmeyer, B., Vallines, I., Rubin, N., & Heeger, D. J. (2008). Neurocinematics: The neuroscience of film. *Projections*, 2(1), 1–26.
24. Hewitt, L., Ashokkumar, A., Ghezae, I., & Willer, R. (2024). Predicting results of social science experiments using large language models. Working paper.
25. Hu, T., & Collier, N. (2024). Quantifying the persona effect in LLM simulations. ACL 2024.
26. Johnson, B. K., & Rosenbaum, J. E. (2015). Spoiler alert: Consequences of narrative spoilers for dimensions of enjoyment, appreciation, and transportation. *Communication Research*, 42(8), 1068–1088.
27. Larsen, J. T., McGraw, A. P., & Cacioppo, J. T. (2001). Can people feel happy and sad at the same time? *Journal of Personality and Social Psychology*, 81(4), 684–696.
28. Leavitt, J. D., & Christenfeld, N. J. S. (2011). Story spoilers don't spoil stories. *Psychological Science*, 22(9), 1152–1154.
29. Li, P., Castelo, N., Katona, Z., & Sarvary, M. (2024). Frontiers: Determining the validity of large language models for automated perceptual analysis. *Marketing Science*, 43(2), 254–266.
30. McDuff, D., el Kaliouby, R., Cohn, J. F., & Picard, R. W. (2015). Predicting ad liking and purchase intent: Large-scale analysis of facial responses to ads. *IEEE Transactions on Affective Computing*, 6(3), 223–235.
31. Mobbs, D., Weiskopf, N., Lau, H. C., Featherstone, E., Dolan, R. J., & Frith, C. D. (2006). The Kuleshov effect: The influence of contextual framing on emotional attributions. *Social Cognitive and Affective Neuroscience*, 1(2), 95–106.
32. Moon, S., Abdulhai, M., Kang, M., Suh, J., et al. (2024). Virtual personas for language models via an anthology of backstories. EMNLP 2024.
33. Nowicki, F., et al. (2026). Visual affect analysis: Predicting emotions of image viewers with vision-language models. arXiv:2602.00123. Preprint.
34. Oliver, M. B. (1993). Exploring the paradox of the enjoyment of sad films. *Human Communication Research*, 19(3), 315–342.
35. Park, J. S., et al. (2026). LLM agents grounded in self-reports enable general-purpose simulation of individuals. arXiv:2411.10109v3.
36. Radford, A., et al. (2021). Learning transferable visual models from natural language supervision. ICML 2021.
37. Reagan, A. J., Mitchell, L., Kiley, D., Danforth, C. M., & Dodds, P. S. (2016). The emotional arcs of stories are dominated by six basic shapes. *EPJ Data Science*, 5, 31.
38. Rentfrow, P. J., Goldberg, L. R., & Zilca, R. (2011). Listening, watching, and reading: The structure and correlates of entertainment preferences. *Journal of Personality*, 79(2), 223–258.
39. Sarstedt, M., Adler, S. J., Rau, L., & Schmitt, B. (2024). Using large language models to generate silicon samples in consumer and marketing research. *Psychology & Marketing*, 41(6), 1254–1270.
40. Schaefer, A., Nils, F., Sanchez, X., & Philippot, P. (2010). Assessing the effectiveness of a large database of emotion-eliciting films. *Cognition & Emotion*, 24(7), 1153–1172.
41. Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being. *Journal of Personality and Social Psychology*, 45(3), 513–523.

42. Soto, C. J., & John, O. P. (2017). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology*, 113(1), 117–143.
43. Teixeira, T., Wedel, M., & Pieters, R. (2012). Emotion-induced engagement in internet video advertisements. *Journal of Marketing Research*, 49(2), 144–159.
44. Wang, A., Morgenstern, J., & Dickerson, J. P. (2025). Large language models that replace human participants can harmfully misportray and flatten identity groups. *Nature Machine Intelligence*, 7(3), 400–411.
45. Zillmann, D. (1988). Mood management through communication choices. *American Behavioral Scientist*, 31(3), 327–340.
46. Zlatintsi, A., et al. (2017). COGNIMUSE: A multimodal video database annotated with saliency, events, semantics and emotion. *EURASIP Journal on Image and Video Processing*, 2017, 54.

Disclosures

This paper combines published research, FishDog's published methodology and calibration material, and a description of the production system. Academic studies explain the science and the feasibility of the approach; they are not independent validations of FishDog. No paired human validation results are reported in this paper. FishDog's emotional readings are modeled estimates of a defined audience's response, not testimony from people who watched the material. Synthetic viewer quotations are labeled as modeled.

FishDog · fish.dog · October 2026