

Cruelty Across Substrates

From animal cruelty to embodied AI, and the coming detection gap.

A fifty-year-old warning signal is about to go silent.

Childhood cruelty to animals is one of the most durable predictors of later violence. As synthetic agents come to perform distress convincingly, that dispositional training migrates to a private venue with **no witness, no veterinary record, no body to find.**

The paper does not claim synthetic agents are moral patients. It claims the signal risks vanishing precisely as the conditions for its use accelerate.

SOURCES Kellert & Felthous (1985); Arluke et al. (1999); APA, DSM-5 (2013).

Practicing cruelty trains cruelty. The target's substrate is incidental.

It predicts violence not because animals are moral patients, but because the practice rehearses a disposition that **generalizes**, and the rehearsal happens entirely inside the perpetrator.

Desensitization

Causing suffering attenuates the affective response that suffering ordinarily evokes.

Dominance affect

For some, the pleasure attaches to control over a being unable to resist.

Social learning

Cruelty internalized as a viable way to relate to vulnerability.

A signal written into the manual clinicians use every day.

Replicated since 1985, across populations and methods. The diagnostic chain, animal cruelty → Conduct Disorder → Antisocial Personality Disorder, **is not theoretical**.

5× increase in violent-crime arrests among animal abusers vs. non-abusers (Arluke et al., 1999)

DSM-5 · Conduct Disorder

Physical cruelty to animals, 1 of 15 symptom criteria.

FBI Behavioral Science Unit

Pattern in operational profiling since the late 20th century.

Distress performed convincingly is not a remote prospect. It is funded and deployed.

WORLD MODELS

Persistent identity, embodied action

World Labs' Marble · Genie 3 · Nvidia Cosmos, built to learn the way biological agents do.

CLOSED-LOOP VOICE

Hears your tone, matches in <200ms

Inworld TTS-2 · IndexTTS2's 8-emotion control, anger, fear, sadness, pleading, on consumer hardware.

REGULATORY NOTICE

Legislators have started to react

Multiple state bills target humanlike AI, without yet finding a workable response.

Users already abuse synthetic agents, without believing them conscious.

The threshold for social uptake is **interactional cueing**, not **metaphysical belief**. The analogy is not a leap from the empirical to the speculative, the HRI record already documents the dominance, social-learning, and group-reinforcement signatures.

The Media Equation

People apply social rules to machines that show social cues. (*Reeves & Nass*)

Children abusing a service robot

Dominance affect + group reinforcement, no belief in harm. (*Brščić et al.*)

They hesitated when it pleaded

Empathic-distress activation, attenuated vs. human. (*Bartneck & Hu; Rosenthal-von der Pütten*)

Four properties separate routine destruction from the configuration that trains cruelty.

Contingent response

Distress reacts to what the perpetrator does, not a decoupled, scripted death-cry.

Identity persistence

The same target across events, carrying memory of prior harm, not a respawn.

Perceptual fidelity

Looks and sounds enough like a being that can suffer, not a 1993 sprite.

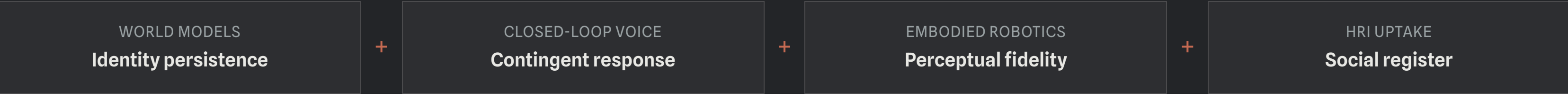
Social register uptake

Processed through social and dominance registers, not purely instrumental ones.

1993 shooter: at most one property.

Convergent systems: all four, at high fidelity, at once.

Four layers compose, and the four properties land at once.



It flinches, pleads, modulates its voice in response to the perpetrator's anger. **The paper does not claim it is in pain**, only that the perpetrator's rehearsal recruits similar mechanisms.

The convergence is not a future moment. The voice layer has shipped. It is occurring in stages, now.

SOURCES World Labs (2025–26); Google DeepMind, Genie 3 (2025); Inworld AI, Realtime TTS-2 (2026); Reeves & Nass (1996); Brščić et al. (2015).

THE ANIMAL CASE · TRACES EXIST

The body is found

The vet keeps a record

The neighbor hears, reports

A school referral follows



THE SYNTHETIC CASE · NO TRACE

No body. No vet. No witness. No injury record.

And on local hardware, no telemetry any vendor can see. The collapse runs past physical evidence to **any digital trace at all.**

The pathway is not interrupted. Only its visibility is. The perpetrator still practices, where no one can see.

SOURCES Ascione & Arkow (1999); Ascione (1993, 2005); Gilly, Cruelty Across Substrates (2026), sec. VII.

WHERE THE FIELDS MUST ADAPT

- A DSM-6 question on cruelty toward humanoid agents
- Pediatric & school intake, extend "how do you treat the pet?"
- Character evidence under Rule 404(b)
- Screening for sensitive roles

WHAT THE ARGUMENT DOES NOT SUPPORT

- Not** domestic behavioral surveillance of children.
- Not** routine reporting of ordinary voice-agent use.
- Not** equating a rude chatbot exchange with a violent disposition.
- Not** manufacturer reporting that bypasses consent.

Clinical inquiry with consent ≠ mandatory state reporting.

SUBSTRATISM

"It's just silicon, a category error."

The argument never claims the agent is harmed. It claims the **perpetrator is trained**. The objection misreads what the argument is doing.

ANIMACY AWARENESS · THE HARDER ONE

"Knowing it's a machine may dampen the effect."

Conceded, it may. Empathic response to robot abuse is real but attenuated. The moderation is likely **partial, and shrinks as the perceptual gap closes**. An open empirical question.

The signal must not go silent precisely as the conditions for its use accelerate.

The next decade's work is to adapt the warning system, or lose one of the most durable early indicators forensic and clinical assessment has. **And to do the adapting without recreating the surveillance state.**

The Kantian observation that cruelty corrupts the cruel is sufficient to ground the argument. The rest is detection, and it is collapsing.

SOURCES Gilly, T. (2026). Cruelty Across Substrates. Real Safety AI Foundation, working paper (v7.5).

— REFERENCES

Gilly, T. (2026). *Cruelty Across Substrates: From Animal Cruelty to Embodied AI and the Coming Detection Gap*. Working paper (v7.5), Real Safety AI Foundation.

Anderson, C. A., Shibuya, A., Ihori, N., Swing, E. L., Bushman, B. J., Sakamoto, A., Rothstein, H. R., & Saleem, M. (2010). Violent video game effects on aggression, empathy, and prosocial behavior in eastern and western countries: A meta-analytic review. *Psychological Bulletin*, 136(2), 151–173.

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). American Psychiatric Publishing.

Arluke, A., Levin, J., Luke, C., & Ascione, F. (1999). The relationship of animal abuse to violence and other forms of antisocial behavior. *Journal of Interpersonal Violence*, 14(9), 963–975.

Arrambide, K., Yoon, J., MacArthur, C., Rogers, K., Luz, A., & Nacke, L. E. (2022). “I don’t want to shoot the android”: Players translate real life moral intuitions to in game decisions in Detroit: Become Human. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (pp. 1–15). ACM.

Artificial Analysis. (2026). Speech Arena leaderboard. Retrieved May 2026 from <https://artificialanalysis.ai/text-to-speech/leaderboard>.

Ascione, F. R. (1993). Children who are cruel to animals: A review of research and implications for developmental psychopathology. *Anthrozoös*, 6(4), 226–247.

Ascione, F. R., & Arkow, P. (Eds.). (1999). *Child Abuse, Domestic Violence, and Animal Abuse: Linking the Circles of Compassion for Prevention and Intervention*. Purdue University Press.

Ascione, F. R. (2005). *Children and Animals: Exploring the Roots of Kindness and Cruelty*. Purdue University Press.

Baker Botts. (2026, February 24). AI chatbot regulation: 78 state bills, 58 lawsuits. Retrieved from <https://ourtake.bakerbotts.com/post/102mipe/ai-chatbot-regulation-78-state-bills-58-lawsuits>.

Bandura, A. (1973). *Aggression: A Social Learning Analysis*. Prentice-Hall.

Banakou, D., & Slater, M. (2014). Body ownership causes illusory self attribution of speaking and influences subsequent real speaking. *Proceedings of the National Academy of Sciences*, 111(49), 17678–17683.

— REFERENCES (CONTINUED)

Bandura, A. (1986). Social Foundations of Thought and Action: A Social Cognitive Theory. Prentice-Hall.

Bartneck, C., & Hu, J. (2008). Exploring the abuse of robots. *Interaction Studies*, 9(3), 415–433.

Bilibili. (2025). bilibili Model Use License Agreement. Retrieved from [https://github.com/index-tts/index-tts.Brscič,D.,Kidokoro,H.,Suehiro,Y.,&Kanda,T.\(2015\).Escapingfromchildren’sabuseofsocialrobots.InProceedingsoftheTenthAnnualACM/IEEEInternationalConferenceonHumanRobotInteraction\(pp.59–66\).](https://github.com/index-tts/index-tts.Brscič,D.,Kidokoro,H.,Suehiro,Y.,&Kanda,T.(2015).Escapingfromchildren’sabuseofsocialrobots.InProceedingsoftheTenthAnnualACM/IEEEInternationalConferenceonHumanRobotInteraction(pp.59–66).)

Carey, S. (1985). Conceptual Change in Childhood. MIT Press.

Carey, S. (2009). The Origin of Concepts. Oxford University Press.

Cleckley, H. (1976). The Mask of Sanity (5th ed.). C. V. Mosby. (Original work published 1941).

Coeckelbergh, M. (2010). Robot rights? Towards a social-relational justification of moral consideration. *Ethics and Information Technology*, 12(3), 209–221.

Coeckelbergh, M. (2017). New Romantic Cyborgs: Romanticism, Information Technology, and the End of the Machine. MIT Press.

Darling, K. (2021). The New Breed: What Our History with Animals Reveals about Our Future with Robots. Henry Holt.

Engels, K. S., & Evans, S. (2022). Detroit Become Human as philosophy: Moral reasoning through gameplay. In D. K. Johnson (Ed.), *The Palgrave Handbook of Popular Culture as Philosophy*. Palgrave Macmillan. https://doi.org/10.1007/978-3-319-97134-6_83-1.

Ferguson, C. J. (2015). Do angry birds make for angry children? A meta-analysis of video game influences on children’s and adolescents’ aggression, mental health, prosocial behavior, and academic performance. *Perspectives on Psychological Science*, 10(5), 646–666.

Felthous, A. R., & Kellert, S. R. (1987). Childhood cruelty to animals and later aggression against people: A review. *American Journal of Psychiatry*, 144(6), 710–717.

Hellman, D. S., & Blackman, N. (1966). Enuresis, firesetting and cruelty to animals: A triad predictive of adult crime. *American Journal of Psychiatry*, 122(12), 1431–1435.

Google DeepMind. (2025, August 5). Genie 3: A new frontier for world models. Retrieved from <https://deepmind.google/blog/genie-3-a-new-frontier-for-world-models/>.

Google. (2026, January 29). Project Genie: Experimenting with infinite, interactive worlds. Retrieved from <https://blog.google/innovation-and-ai/models-and-research/google-deepmind/project-genie/>.

— REFERENCES (CONTINUED)

Hare, R. D. (1991). The Hare Psychopathy Checklist—Revised. Multi-Health Systems.

Hensley, C., Tallichet, S. E., & Singer, S. D. (2006). Exploring the possible link between childhood and adolescent bestiality and interpersonal violence. *Journal of Interpersonal Violence*, 21(7), 910–923.

Holl, E., & Melzer, A. (2022). Moral minds in gaming: A quantitative case study of moral decisions in Detroit: Become Human. *Journal of Media Psychology*, 34(5), 287–298.

IndexTeam. (2025, September 8). IndexTTS2: Emotion controlled, duration accurate, zero shot text to speech engine. Retrieved from <https://github.com/index-tts/index-tts>.

Inworld AI. (2026, May 5). Realtime TTS-2: A new frontier voice model that feels as human as it sounds. Retrieved from <https://inworld.ai/blog/realtime-tts-2>.

Kahn, P. H., Kanda, T., Ishiguro, H., Freier, N. G., Severson, R. L., Gill, B. T., Ruckert, J. H., & Shen, S. (2012). “Robovie, you’ll have to go into the closet now”: Children’s social and moral relationships with a humanoid robot. *Developmental Psychology*, 48(2), 303–314.

Kant, I. (1996). *Practical Philosophy* (M. J. Gregor, Ed. and Trans.). Cambridge University Press.

Kant, I. (1997). *Lectures on Ethics* (P. Heath & J. B. Schneewind, Eds.; P. Heath, Trans.). Cambridge University Press.

Kellert, S. R., & Felthous, A. R. (1985). Childhood cruelty toward animals among criminals and noncriminals. *Human Relations*, 38(12), 1113–1129.

Kuhlmeier, V. A., Bloom, P., & Wynn, K. (2004). Do 5 month old infants see humans as material objects? *Cognition*, 94(1), 95–103.

MarkTechPost. (2026, May 5). Inworld AI launches Realtime TTS-2: A closed-loop voice model that adapts to how you actually talk. Retrieved from <https://www.marktechpost.com/2026/05/05/inworld-ai-launches-realtime-tts-2-a-closed-loop-voice-model-that-adapts-to-how-you-actually-talk/>.

Merz Perez, L., & Heide, K. M. (2004). *Animal Cruelty: Pathway to Violence Against People*. AltaMira Press.

Moffitt, T. E. (1993). Adolescence limited and life course persistent antisocial behavior: A developmental taxonomy. *Psychological Review*, 100(4), 674–701.

National Law Review. (2026, February 11). Tennessee’s AI bill would criminalize the training of AI chatbots. Retrieved from <https://natlawreview.com/article/tennessees-ai-bill-would-criminalize-training-ai-cha>.

— REFERENCES (CONTINUED)

Nvidia. (2026, January 7). Cosmos celebrates 5 million downloads on first anniversary. NVIDIA Developer Forum. Retrieved from <https://www.nvidia.com/en-us/ai/cosmos/>.

Reeves, B., & Nass, C. (1996). The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places. Cambridge University Press.

Ressler, R. K., Burgess, A. W., & Douglas, J. E. (1988). Sexual Homicide: Patterns and Motives. Lexington Books.

Reuters. (2026, February 18). AI pioneer Fei-Fei Li’s World Labs raises \$1 billion in funding. Retrieved from <https://www.reuters.com/business/ai-pioneer-fei-fei-lis-world-labs-raises-1-billion-funding-2026-02-18/>.

Robins, L. N. (1978). Sturdy childhood predictors of adult antisocial behaviour: Replications from longitudinal studies. *Psychological Medicine*, 8(4), 611–622.

Rosenthal-von der Pütten, A. M., Krämer, N. C., Hoffmann, L., Sobieraj, S., & Eimler, S. C. (2013). An experimental study on emotional reactions towards a robot. *International Journal of Social Robotics*, 5(1), 17–34.

Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B*, 364(1535), 3549–3557.

Slater, M., Spanlang, B., Sánchez-Vives, M. V., & Blanke, O. (2010). First person experience of body transfer in virtual reality. *PLoS ONE*, 5(5), e10564.

Spelke, E. S., Phillips, A., & Woodward, A. L. (1995). Infants’ knowledge of object motion and human action. In D. Sperber, D. Premack, & A. J. Premack (Eds.), *Causal Cognition: A Multidisciplinary Debate* (pp. 44–78). Oxford University Press.

Tamborini, R. (2011). Moral intuition and media entertainment. *Journal of Media Psychology*, 23(1), 39–45.

Transparency Coalition for AI. (2026, April 3). AI legislative update. Retrieved from <https://www.transparencycoalition.ai/news/ai-legislative-update-april3-2026>.

Troutman Pepper. (2026, April 2). Tennessee enacts health care AI bill with private right of action. Retrieved from <https://www.troutmanprivacy.com/2026/04/tennessee-enacts-health-care-ai-bill-with-private-right-of-action/>.

World Labs. (2025, November 12). Marble: A multimodal world model. Retrieved from <https://www.worldlabs.ai/blog/marble-world-model>.

World Labs. (2026, January 21). Announcing the World API. Retrieved from <https://www.worldlabs.ai/blog/announcing-the-world-api>.

— REFERENCES (CONTINUED)

Wright, J., & Hensley, C. (2003). From animal cruelty to serial murder: Applying the graduation hypothesis. *International Journal of Offender Therapy and Comparative Criminology*, 47(1), 71–88.