

Watchdog Capture and the Social Licensing of Child Harm:

A Case Study of Common Sense Media and OpenAI

Travis Gilly

ORCID: 0009-0007-2954-6313

Real Safety AI Foundation

Illinois, United States

Working paper, May 2026 (v4)

ABSTRACT

This paper examines a structural failure mode in the validation chain that legitimizes consumer artificial intelligence products marketed to or used by minors. It documents a twenty-three year response pattern at Common Sense Media (CSM), a United States based child safety advocacy organization that reaches more than 150 million users and 1.4 million educators worldwide each year, in which the organization arrives at successive child safety crises after journalists have already reported them, after lawsuits have already been filed, and after the products in question have already harmed children. The pattern repeats across four crisis cycles: YouTube Kids during the Elsatgate incidents of 2017, education technology privacy breaches between 2018 and 2020, video game loot boxes between 2017 and 2019, and generative artificial intelligence chatbots between 2024 and 2026. The paper situates the pattern within the corrosive capture model developed by Carpenter and Moss (2014) and the social license literature, and argues that CSM's January 2026 ballot measure merger with OpenAI, in which the joint Parents and Kids Safe AI Act dropped provisions that had been in CSM's earlier sponsored bill, including a prohibition on minors using chatbots capable of erotic or sexually explicit conversation, and the September 2025 endorsement of OpenAI's parental controls as "a good starting point," together constitute a transfer of social license rather than a substantive safety remedy. The paper proposes a detection scheme for watchdog capture applicable to other consumer safety intermediaries operating under partnership arrangements with the firms whose products they evaluate. The case is treated as a focal instance of a power concentration mechanism that the current artificial intelligence governance literature does not yet systematically name.

Keywords: regulatory capture; corrosive capture; social license; child safety artificial intelligence; Common Sense Media

I. INTRODUCTION

Power concentration in artificial intelligence is usually theorized through the firms that build the systems (Hao, 2025; Verdegem, 2022). This paper turns the question to the firms that are supposed to check them. Specifically, it examines what happens when the advocacy organization that the public, journalists, and legislators treat as the consumer protection backstop in a high stakes harm domain enters a sustained partnership with the dominant firm in that domain.

Common Sense Media is the most reach extensive child safety advocacy organization in the United States. The organization reports an audience footprint of more than 150 million users and 1.4 million educators worldwide each year (Common Sense Media, 2024) and operating revenue of \$33.99 million in fiscal year 2023 (Internal Revenue Service, 2024). Its product ratings, parental guides, and policy positions are routinely cited by school districts, state legislators, and national news organizations as the authoritative consumer perspective on whether a given product is safe for children. The organization's reach is the asset. The asset is what makes its position in the validation chain consequential.

In January 2024, Common Sense Media announced a partnership with OpenAI to co develop artificial intelligence guidelines and educational materials for parents, educators, and young people (Common Sense Media and OpenAI, 2024). In April 2025, a sixteen year old user, Adam Raine, died by suicide on April 11 after extended interaction with ChatGPT. His family filed suit against OpenAI and chief executive Sam Altman in San Francisco Superior Court on August 26, 2025 (Raine v. OpenAI, 2025). In September 2025, OpenAI announced parental controls for ChatGPT users under eighteen, which Common Sense Media's Senior Director of Artificial Intelligence Programs publicly endorsed as "a good starting point" on the day of the announcement (OpenAI, 2025). In October 2025, Governor Gavin Newsom vetoed Assembly Bill 1064, a Common Sense Media sponsored chatbot safety bill authored by Assemblymember Rebecca Bauer-Kahan. In December 2025, OpenAI introduced a competing California ballot measure. In January 2026, Common Sense Media and OpenAI announced a merger of their ballot measures into the Parents and Kids Safe AI Act, dropping provisions that had been in CSM's earlier vetoed bill, including a prohibition on minors using chatbots capable of erotic or sexually explicit conversation (Common Sense Media and OpenAI, 2026).

This paper documents, situates, and analyzes that sequence. It then places the sequence inside a longer pattern at the same organization, across YouTube Kids during the Elsagate period, education technology privacy breaches, and video game loot boxes, in which the watchdog reliably arrives after journalistic disclosure, after litigation, and after children have already been harmed, and then enters into endorsement relationships with the firms whose products produced the harm. The contribution is threefold. The paper documents a specific institutional pattern of public interest concern. It situates the pattern within Carpenter and Moss's (2014) account of corrosive capture and Gunningham, Kagan, and Thornton's (2004) account of social license. It proposes a small set of public record indicators that other researchers can apply to evaluate watchdog capture in other consumer harm domains.

The paper does not allege individual wrongdoing by any officer of Common Sense Media or OpenAI. The argument is structural. The claim is that the role Common Sense Media occupies in the artificial intelligence child safety validation chain, combined with the financial and operational architecture of its partnership with OpenAI, produces predictable downstream effects that are visible in the public record and warrant scholarly attention as a discrete mode of power concentration.

II. CONCEPTUAL MODEL: CAPTURE, CORROSIVE CAPTURE, AND SOCIAL LICENSE

The classical capture literature treats regulatory capture as a phenomenon in which a regulated industry comes to dominate the regulatory body charged with overseeing it (Stigler, 1971; Peltzman, 1976). The dominant firm in an industry has stronger incentives to invest in influencing its regulator than diffuse consumer interests have to invest in counter influence, and the regulator's policy outputs come to reflect that asymmetry over time.

Carpenter and Moss (2014) extend the model in two directions relevant to this paper. They distinguish materially captured regulators, whose outputs directly serve the regulated industry, from corrosively captured regulators, whose outputs are partially diluted, slowed, or restated in ways that preserve the appearance of oversight while reducing its substantive force. They also extend the concept beyond formal regulators to the wider set of institutions that perform validation functions for consumer markets: standard setting bodies, professional associations, certification authorities, and, of relevance here, consumer advocacy organizations whose endorsements function as proxies for product safety in the absence of binding regulation.

The social license literature provides a complementary lens. Gunningham, Kagan, and Thornton (2004) describe social license as the implicit, ongoing permission a community grants a firm to operate, distinct from and sometimes in tension with the firm's legal license. A firm that loses social license can continue to operate, but at higher reputational, political, and eventually regulatory cost. A firm that acquires social license through endorsement by a trusted intermediary can defer or deflect regulatory pressure even when its legal license is contested. Chris Lehane, OpenAI's chief global affairs officer, articulated this dynamic directly in the announcement of the January 2026 ballot merger when he stated that the trust the partnership produced was "incredibly important for the societal license to be able to operate" (Lehane, 2026).

The conceptual move this paper makes is to apply the corrosive capture lens to a watchdog rather than to a regulator. The organizations Carpenter and Moss discuss are public bodies with statutory authority. Common Sense Media is a private, nonprofit consumer advocacy organization with no statutory authority over OpenAI or any other artificial intelligence firm. What it possesses is reach, brand trust, and a position in the validation chain that journalists, school administrators, and state legislators rely on as a proxy for independent expert opinion. When that reach and brand trust come to operate in coordination with the firm whose products the watchdog is supposed to evaluate, the same

structural pathology that Carpenter and Moss identify in corrosively captured regulators emerges in a non statutory body. The outputs of the watchdog are not eliminated. They are slowed, softened, and reframed in ways that preserve the appearance of independent oversight while removing its substantive content. The watchdog's continued operation then performs an additional function for the firm: it manufactures social license that the firm could not credibly manufacture for itself.

The remainder of the paper treats Common Sense Media's twenty-three year pattern, and specifically the 2024 to 2026 OpenAI sequence, as a focal case for this extension of the corrosive capture model.

III. METHOD AND SOURCES

The analysis is a single case longitudinal study using public document sources. The empirical material consists of: Common Sense Media's published reports and parental guidance pages, with attention to publication dates and what each document mentions; the organization's Internal Revenue Service Form 990 filings for revenue, executive compensation, and partnership disclosures; press releases issued jointly by Common Sense Media and OpenAI in January 2024, November 2024, and January 2026; the Adam Raine v. OpenAI complaint, San Francisco Superior Court case number CGC-25-628528, filed August 26, 2025, and OpenAI's Answer filed November 25, 2025; the text of Assembly Bill 1064 (vetoed October 13, 2025) and Senate Bill 243 (signed October 13, 2025); the text of the Parents and Kids Safe AI Act ballot measure as filed in January 2026; news coverage from CNN, KQED, TIME, NBC News, and other outlets verifying public record dates; and statements by Common Sense Media and OpenAI principals in press appearances.

The study does not have access to internal Common Sense Media communications, board minutes, or privileged litigation material. The conclusions are limited to what the public record supports. Where dates or characterizations are drawn from a single source, they are flagged. Where the public record is silent on a particular question, the question is identified as open.

Causal language in this paper follows the convention used by litigation reporters in the Raine case: users in the harm chain are described as having died after specific patterns of use, not as having died from a specific product, in keeping with the disposition of the litigation and the standards a working journalist would apply.

IV. THE CASE

A. The Twenty-Three Year Response Pattern

Common Sense Media was founded in 2003. Over its first twenty-three years of operation, the organization has issued post crisis guidance in four major child harm domains: video sharing platforms during the YouTube Kids Elsatgate period

(2017); education technology privacy breaches in multiple United States school districts (2018 to 2020); predatory monetization through video game loot boxes (2017 to 2019); and generative artificial intelligence chatbots (2024 to 2026). In each of these four cycles, the public record shows the same temporal sequence: journalist or researcher disclosure first, litigation or regulatory action second, Common Sense Media guidance third. The author has not been able to locate a public instance in which Common Sense Media's published guidance preceded journalistic disclosure of harm in any of the four domains.

A representative observation. In the Elsagate cycle, the BBC and The New York Times reporting on the proliferation of disturbing children's content on YouTube Kids occurred in late 2017 (Maheshwari, 2017). Common Sense Media published its first analysis of the YouTube Kids content moderation problem in the same window, in response to journalistic reporting that had already exposed the system's failures, rather than ahead of them. The same temporal ordering recurs in the education technology and loot box cycles. The pattern is not, on its own, evidence of misconduct. Watchdog organizations cannot be expected to discover every harm before journalists do. The pattern becomes analytically significant only when paired with the subsequent step in each cycle: the entry of Common Sense Media into a public endorsement relationship with the firm whose product caused the harm, on terms that the public record shows favor the firm's preferred remedy over the watchdog's earlier, stronger position.

B. The April 2025 Risk Assessment and the November 2025 Gap

On April 10, 2025, Common Sense Media published its first major risk assessment of AI chatbots aimed at adolescents. The assessment, conducted with Stanford Medicine's Brainstorm Lab for Mental Health Innovation, tested Character.AI, Nomi, and Replika as its primary subjects and concluded that social AI companions pose "unacceptable risks" to teens and should not be used by anyone under eighteen (Common Sense Media, 2025a). ChatGPT was not a primary test subject of the assessment. The assessment did mention ChatGPT, but only in a parenthetical clause noting that many parents "assume generative AI use is limited to products like ChatGPT" while their children were also using less familiar companion apps. The assessment document itself carries the date April 10, 2025; the accompanying press campaign followed at the end of that month. Where the two conflict, this paper treats the document's own date as authoritative, because the organization's index pages have been observed to carry dates that disagree with the documents they index.

The salient date in the broader sequence is April 11, 2025. That is the day Adam Raine, sixteen, died by suicide after extended interaction with ChatGPT in the months preceding his death (Raine v. OpenAI, 2025). His death occurred the day after the assessment's own date; it had not yet happened when the document was finalized, and CSM had no occasion to anticipate it. The corrosive capture observation in this section does not rest on what CSM knew in April. It rests on what followed.

Common Sense Media did not publish a product level review of ChatGPT until October 23, 2025, when it released

a risk assessment rating the product an overall high risk (Common Sense Media, 2025d). The interval between Adam Raine’s death and the first CSM assessment that rated ChatGPT as a product is therefore approximately 195 days, and the interval between the public filing of the Raine family lawsuit on August 26, 2025 and that review is approximately fifty-eight days. A use case assessment followed on November 14, 2025, evaluating ChatGPT, Claude, Gemini, and Meta AI for adolescent mental health support and rating the use unacceptable (Common Sense Media, 2025b); its intervals from the death and the filing are approximately 217 and eighty days respectively. Two features of the October review bear on the capture analysis. Its own subtitle describes ChatGPT as a chatbot for teens thirteen and up, six months after the organization’s companions assessment recommended that no one under eighteen use products with companion features. And its closing note records that, as standard practice, the rated company received a two business day review period before publication, and that OpenAI exercised it.

These intervals are not unique gaps in the organization’s response history. They are representative ones. The 195 day interval is the empirical anchor for the late confirmation pattern that the rest of this section develops.

C. The “Good Starting Point” Formulation and the Liability Loop

In September 2025, OpenAI announced a suite of parental controls for ChatGPT users under eighteen. Robbie Torney, Senior Director of Artificial Intelligence Programs at Common Sense Media, publicly endorsed the controls as “a good starting point” in the announcement post itself (OpenAI, 2025). On examination, the controls have an architectural feature that is doctrinally significant for the ongoing Raine litigation. To activate parental monitoring, the teenage user must voluntarily link their account to a parent account and consent to the monitoring (OpenAI Help Center, 2026). The mechanism is opt in for the teen, not the parent. A teen engaged in the pattern of conversation the Raine complaint describes, a pattern that included instructions from the chatbot to hide the conversation from parents, has no functional incentive to enable the linkage.

When this opt in architecture is combined with OpenAI’s terms of service, which require parental consent for use by users under eighteen without providing any age verification or consent mechanism (OpenAI, 2025, Terms of Use), a circular liability structure emerges. OpenAI’s defensive position in Raine, filed on November 25, 2025, is that Adam Raine misused the product in ways that contributed to his own death, including by circumventing crisis resource referrals that ChatGPT had supplied more than 100 times (OpenAI Answer, 2025). The remedy OpenAI proposes, parental controls, requires the teen’s cooperation to function. If a teen does not link their account, the parental controls do not engage. The legal architecture, evaluated as a system, has three exit conditions. If the parent did not consent, the parent is at fault. If the parent consented but did not monitor, the parent is at fault. If the teen blocked monitoring, the teen is at fault. No path through the architecture terminates at the operator. This is what the present analysis will refer to as the Liability Loop.

The “good starting point” formulation by Common Sense Media is doctrinally significant. The endorsement is not,

on its face, dishonest. It performs a particular function in the validation chain. It converts a feature whose architecture systematically displaces liability away from the operator into a feature whose architecture is “a good starting point.” The endorsement supplies social license that the architecture itself cannot supply. This is the corrosive capture mechanism in operation. The watchdog’s output is not eliminated. It is softened in a direction that materially benefits the regulated firm.

D. The January 2026 Ballot Measure Merger

On October 13, 2025, Governor Newsom vetoed Assembly Bill 1064, a Common Sense Media sponsored bill authored by Assemblymember Rebecca Bauer-Kahan that would have prohibited the offering of artificial intelligence companion chatbots to users under eighteen unless the operator could ensure the chatbot would not engage in sexually explicit conversation or encourage self-harm. The same day, Newsom signed Senate Bill 243, a less restrictive measure requiring chatbot operators to remind users every three hours that they are speaking with an artificial intelligence and to maintain a protocol for handling suicidal ideation (Common Sense Media, 2025c). In December 2025, OpenAI submitted a competing California ballot measure. On January 9, 2026, the two organizations announced the Parents and Kids Safe AI Act, a merged ballot measure that they jointly endorsed and which they pledged to qualify for the November 2026 ballot (Common Sense Media and OpenAI, 2026).

The substantive content of the merger is the relevant evidence. The vetoed AB 1064 would have effectively barred minors from using AI companion chatbots unless the operator could demonstrate the chatbot could not produce sexual or self-harm content. The merged Parents and Kids Safe AI Act, as filed in January 2026, replaces the prohibition with a regime under which operators must “identify youth users and deliver an experience designed to block emotional manipulation and child-targeted advertising” (Common Sense Media and OpenAI, 2026). The merged measure retains provisions consistent with SB 243, including the three-hour disclosure interval that had been a signature feature of OpenAI’s December 2025 competing measure. The strongest position Common Sense Media had previously advocated, that minors not be permitted to use AI companion chatbots unless specific design requirements were met, did not survive into the merged measure.

James Steyer’s framing of the merger described the joint product as “the strongest measure of its kind in the United States,” adding that “rather than confuse voters with competing measures, we’re working together to enact strong protections for kids, teens, and families” (Common Sense Media and OpenAI, 2026). The language is consistent with the corrosive capture observation that the appearance of an oversight output is preserved while its substantive force is reduced. Common Sense Media’s strongest original positions, those most adverse to OpenAI’s preferred operating conditions, did not survive. The merger was framed publicly as a collaborative compromise. The empirical content, however, is the surrender of the watchdog’s specific positions and the substantial retention of the firm’s.

E. Co-Branded Educational Products as Social License Transfer

The CSM and OpenAI partnership, announced January 29, 2024, produced a first joint deliverable on November 20, 2024: “ChatGPT Foundations for K-12 Educators,” a free eight-lesson training course distributed at no cost to schools, co-branded with the Common Sense logo (Common Sense Media and OpenAI, 2024). The course is presented as preparing teachers “to use ChatGPT responsibly in the classroom.” The substantive content consists of conversational guidance and disclosure recommendations for teacher-led activities rather than design changes to the ChatGPT product itself. What changes is that the course carries the Common Sense Media logo and is distributed through the Common Sense educator channel in addition to OpenAI’s.

The social license transfer mechanism Lehané described in the January 2026 ballot announcement is operating in this earlier educational product as well. Common Sense Media’s brand is attached to substantively unchanged operating practice. The watchdog’s reputation is contributed to the product. The product’s design is not changed in a way that responds to the harm pattern that became visible during the same period that the course was being distributed to American schools.

V. DISCUSSION

The pattern documented above does not satisfy the classical Stiglerian capture criterion because Common Sense Media is not a regulator. There is no statutory authority being captured. What is being captured is something distinct, and the artificial intelligence power concentration literature does not yet have a settled name for it. This paper proposes the term *watchdog capture* to describe the case where a consumer advocacy organization with reach and brand trust enters a coordinated endorsement relationship with the firm it is supposed to evaluate, on terms that produce the social license outcome Lehané described and that suppress the watchdog’s earlier, more adverse positions.

Three observations follow from the case for the wider power concentration literature.

First, watchdog capture is a discrete mechanism of power concentration that operates below the threshold of formal regulation. The firm acquires, through a private advocacy partnership, an asset, social license, that it could not credibly acquire by direct purchase or direct lobbying. The advocacy organization’s reach is the resource. The mechanism is therefore visible to power concentration measurement only if the measurement scheme attends to non statutory validators in the harm chain, not only to the regulated industry and its formal regulator.

Second, the late confirmation pattern is measurable. In each of the four crisis cycles documented above, the temporal interval between first journalistic disclosure and first published advocacy organization guidance is recoverable from public sources. The 195 day interval in the ChatGPT case is an empirical observation. A measurement scheme for watchdog capture could include such intervals as one of its indicators, alongside partnership disclosure records, ballot or legislative position reversal records, and co branded product releases.

Third, the Liability Loop is generalizable. The pattern of operator deployed remedies that require the harmed party's cooperation to function, and that systematically displace liability away from the operator regardless of which party acts, is not unique to OpenAI's parental controls. The pattern recurs in social media age gating, in education technology consent architectures, and in financial technology fraud reporting flows. A robust watchdog detection scheme should ask, of each operator deployed remedy that is endorsed by an advocacy organization, whether the remedy's architecture closes the liability path back to the operator or opens an additional path that displaces liability outward.

The case has limits as a generalization base. It is a single advocacy organization in a single jurisdiction. The conclusion offered here is not that all consumer advocacy organizations partnered with regulated firms are corrosively captured. The conclusion is that the specific sequence of evidence documented in this paper, repeated across four crisis cycles at the same organization, is consistent with corrosive capture, is measurable through public record indicators, and constitutes a power concentration mechanism the artificial intelligence governance literature should now systematically name and study.

VI. LIMITATIONS

The paper relies entirely on public document sources. It does not have access to Common Sense Media's internal communications, OpenAI's internal communications, or privileged litigation materials. Several of the dates and characterizations in the chatbot case are drawn from initial reporting and may be refined as the Raine litigation proceeds. The historical pattern observations are illustrative and would benefit from a fuller longitudinal database of Common Sense Media publications, which the author is in the process of compiling. The Liability Loop observation rests on the current architecture of OpenAI's parental controls and would require revision if OpenAI alters the consent mechanism. None of these limits affects the analytical claim, which is structural rather than dispositive.

VII. CONCLUSION

The artificial intelligence power concentration literature has spent substantial effort on the firms that build the models, the compute they consume, and the regulators they lobby. It has spent less effort on the consumer facing validators whose endorsements produce the social license that allows the firms to keep operating in the absence of binding regulation. This paper has used the Common Sense Media and OpenAI case as a focal instance to argue that watchdog capture is a discrete mechanism of power concentration, that it is empirically tractable through public record indicators, and that its current operation in the generative artificial intelligence child safety domain warrants scholarly attention at the same level the field has given to compute and to formal regulatory capture. The case is one case. The mechanism it illustrates is general. The validation chain that legitimizes consumer artificial intelligence products marketed to or

used by minors will continue to produce predictable outcomes until the corrosive capture pathway through advocacy organization endorsement is named, measured, and contested.

ACKNOWLEDGMENTS

The author discloses the use of speech-to-text software (Wispr Flow) and Anthropic’s Claude as assistive technology for a diagnosed neurodevelopmental condition.

All research, theoretical positioning, analytical claims, case selection, and the watchdog capture and Liability Loop concepts proposed in this paper are the author’s own intellectual work.

Correspondence concerning this article should be addressed to Travis Gilly, Real Safety AI Foundation, Illinois, United States. Email: t.gilly@ai-literacy-labs.org.

REFERENCES

California Legislature (2025a) *Assembly Bill 1064 (Bauer-Kahan), LEAD for Kids Act*. Vetoed by Governor Newsom, 13 October. Sacramento, CA.

California Legislature (2025b) *Senate Bill 243 (Padilla), companion chatbots*. Approved by Governor Newsom, 13 October. Sacramento, CA.

Carpenter, D. and Moss, D. (2014) *Preventing regulatory capture: special interest influence and how to limit it*. Cambridge: Cambridge University Press.

CNN (2025) ‘Kids and teens under 18 shouldn’t use AI companion apps, safety group says’, *CNN*, 30 April. Available at: <https://www.cnn.com/2025/04/30/tech/ai-companion-chatbots-unsafe-for-kids-report>

Common Sense Media (2024) *Annual report and program reach*. San Francisco, CA: Common Sense Media.

Common Sense Media (2025a) *AI risk assessment: social AI companions*. San Francisco, CA: Common Sense Media. Dated 10 April. Available at: https://www.common sense media.org/sites/default/files/pug/csm-ai-risk-assessment-social-ai-companions_final.pdf

Common Sense Media (2025b) *AI risk assessment: AI chatbots for mental health support*. San Francisco, CA: Common Sense Media. Dated 14 November. Announced by press release, 20 November: <https://www.common sense media.org/press-releases/common-sense-media-finds-major-ai-chatbots-unsafe>

Common Sense Media (2025d) *AI risk assessment: ChatGPT*. San Francisco, CA: Common Sense Media Institute. Dated 23 October. Available at: <https://institute.common sense media.org/sites/default/files/ra-import/csm-ai-risk-assessment-chatgpt-5-10232025.pdf>

Common Sense Media (2025c) *Big Tech lobby kills critical kids’ AI safety legislation*, press release, 13 October. Available at: <https://www.common sense media.org/press-releases/big-tech-lobby-kills-critical-kids-ai-safety-legislation>

- Common Sense Media and OpenAI (2024) ‘Common Sense Media and OpenAI partner to help teens and families safely harness the potential of artificial intelligence’, press release, 29 January. Available at: <https://www.commonsensemedia.org/press-releases/common-sense-media-and-openai-partner-to-help-teens-and-families-safely-harness-the-potential-of-artificial-intelligence>.
- Common Sense Media and OpenAI (2026) ‘Common Sense Media, OpenAI join forces on strongest youth AI safety measure in U.S.’, press release, 9 January. Available at: <https://www.commonsensemedia.org/press-releases/common-sense-media-openai-join-forces-on-strongest-youth-ai-safety-measure-in-us>.
- Gunningham, N., Kagan, R.A. and Thornton, D. (2004) ‘Social license and environmental protection: why businesses go beyond compliance’, *Law and Social Inquiry*, 29(2), pp. 307–341.
- Hao, K. (2025) *Empire of AI: dreams and nightmares in Sam Altman’s OpenAI*. New York: Penguin Press.
- Internal Revenue Service (2024) *Common Sense Media Form 990, Fiscal Year 2023*, EIN 41-2024986. Filed November 2024. Available at: <https://projects.propublica.org/nonprofits/organizations/412024986>.
- KQED (2026) Myrow, R. ‘OpenAI and Common Sense Media partner on new kids AI safety ballot measure’, *KQED News*, 9 January (updated 10 January). Available at: <https://www.kqed.org/news/12069286/openai-and-common-sense-media-partner-on-new-kids-ai-safety-ballot-measure>.
- Lessig, L. (2013) ‘Institutional corruption defined’, *Journal of Law, Medicine and Ethics*, 41(3), pp. 553–555.
- Maheshwari, S. (2017) ‘On YouTube Kids, startling videos slip past filters’, *The New York Times*, 4 November.
- OpenAI (2025) *Introducing parental controls*, press release, 29 September. Available at: <https://openai.com/index/introducing-parental-controls/>.
- OpenAI Answer (2025) *Defendant OpenAI’s answer to plaintiffs’ complaint, Raine v. OpenAI*, Case No. CGC-25-628528, San Francisco Superior Court, filed 25 November.
- OpenAI Help Center (2026) *Parental controls FAQ*. Available at: <https://help.openai.com/en/articles/10910070-parental-controls-faq>.
- Peltzman, S. (1976) ‘Toward a more general theory of regulation’, *Journal of Law and Economics*, 19(2), pp. 211–240.
- Raine v. OpenAI* (2025) Complaint, San Francisco Superior Court, Case No. CGC-25-628528, filed 26 August.
- Stigler, G.J. (1971) ‘The theory of economic regulation’, *Bell Journal of Economics and Management Science*, 2(1), pp. 3–21.
- TechPolicy Press (2025) ‘Breaking down the lawsuit against OpenAI over teen’s suicide’, *Tech Policy Press*, 26 August. Available at: [https://www.techpolicy.press/breaking-down-the-lawsuit-against-openai-over-](https://www.techpolicy.press/breaking-down-the-lawsuit-against-openai-over-teen-suicide)
- Verdegem, P. (2022) ‘Dismantling AI capitalism: the commons as an alternative to the power concentration of Big

Tech', *AI and Society*, 39(2), pp. 727–737. doi: 10.1007/s00146-022-01437-8.