Affordances for Harm

How Offenders Misuse Platform Capabilities to Exploit Children, and Where to Intervene

Mrinaal Ramachandran

Graduate Student, Department of Computer Science
University of Massachusetts Amherst

mramachandra@umass.edu · *Independent Research*

Corpus	7,426 cases 80,000+ features 56 sources
Coverage	61 task forces 2,864 law enforcement agencies
Timespan	2002–2026
Platforms	30+ analyzed

Abstract

For years, the history of the internet has been championed through the lens of platforms, growth, and innovation. Far less often do we systematically examine the harms that have scaled alongside that same infrastructure.

From large-scale content distribution and global connectivity to encrypted messaging and AI-generated imagery, the capabilities that make technology worth using have also made it exploitable. The same infrastructure that connected billions of people quietly rewired the economics of child exploitation—not because these technologies were designed to cause harm, but because offenders are adaptive: some pursue children deliberately, others are capitalizing on access and opportunity the internet has made possible for the first time.

This paper draws on 7,426 curated Internet Crimes Against Children case records spanning 2002 to 2026 and asks: which platform capabilities offenders exploit most consistently, how technology is weaponized within specific offense subtypes—grooming, sextortion, production of CSAM, coordinated criminal networks—and where disruption is most realistic. Across 30+ platforms and 61 task forces, the same capability types surface again and again—anonymity, disappearing content, file distribution, contact discovery, trust-building at speed—regardless of which platform is in the headlines. The technology changes. The exploitation mechanics do not.

Contents

1. Exploitation Is a Human Choice

The capabilities that make the internet valuable to hundreds of millions of legitimate users are the same capabilities offenders have learned to turn against children. Every case in this corpus begins with the same decision: a person chose to exploit a child.

Across the twenty-four years and 7,426 enforcement records this corpus spans, what technology changed was not that decision but its operational scope.

Child sexual exploitation is not a problem the internet invented. It is among the oldest documented harms to children, and the children it targets have always been vulnerable for the same reasons: age, developmental stage, and dependency on adults. What the internet changed is the capability available to the person who decides to act on that vulnerability. The offender in 1990 operated inside constraints that no longer exist in the same form. Physical proximity was required to make contact. Geography bounded victim access. Material was difficult to produce, store, and distribute. Detection risk was comparatively high at every step. In 2026, none of those constraints hold. A device in a pocket provides global reach. Anonymity is default on the most-used platforms in the world. Content can be produced, encrypted, distributed, and erased in seconds. The barriers did not fall because technology made people worse. They fell because technology raised the capability ceiling for everyone — including those who use it to harm.

This is the amplification problem, and it is distinct from causation. No platform built a feature to enable child exploitation. The features exist for hundreds of millions of legitimate users, and bad actors have learned to use the same capabilities that make the technology valuable. That is not a design failure unique to any company; it is the reality of the modern digital society. It is also why the question *which platform did this?* is less useful than the question this paper asks: *which capabilities are consistently reached for, and why do they appear across every platform in the enforcement record?*

Offenders in this corpus range from individuals who systematically identified and groomed multiple victims across years to those who acted on an available opportunity technology placed in front of them. Both types are in the enforcement record. Both made a choice. Understanding how malicious users transform platform affordances into misuse surfaces is the prerequisite for designing against them.

The landscape of exploitation continues to evolve alongside technology. The dominant image—a predator posing as a peer, deceiving a child into a meeting

that ends in assault—is one offense type in the enforcement record. It is not the only one. Sextortion, in which offenders coerce victims through threats to expose images, emerged as a distinct and rapidly growing typology only in the last decade, made possible by the same digital infrastructure that connects everyone online. Coordinated networks of offenders operating across gaming platforms represent another. Each capability introduced has, in turn, produced offense patterns that did not previously exist at scale. Offending profiles and capabilities are not static. They expand with technology.

The harm in these cases is not abstract. Livingstone and Smith [12] document the inseparability of digital and physical risk for children: what happens on a platform shapes what happens offline, and the boundary between them is not meaningful to the child living inside it. Contact made in a chat room migrates. Material produced in one context circulates in others long after the offense ends. The platform did not manufacture predatory intent. It provided the infrastructure through which that intent found a child.

This paper asks three sequential questions about that infrastructure. Which capabilities do offenders consistently reach for, across offense types, platforms, and years? How do those affordances map to harm vectors across the technological eras this corpus spans? Where in that infrastructure does intervention become realistic? The affordance framework developed in §2 provides the analytical language: a way to ask not which platform was involved but which capability was used, and what that capability made possible. The answer turns out to be stable across three decades of platform evolution. Anonymity. Ephemerality. Distribution infrastructure. Contact discovery. Trust architecture at speed. These capabilities appear in the enforcement record because they are structurally useful: they translate platform properties into offense mechanics.

Exploitation is a human choice. The technology does not make that choice. But it shapes, in ways this paper measures at scale for the first time, the harm signatures of modern offenders.

2. The Infrastructure of Harm

2.1 Prior Work: CSEA, Platforms, and the Enforcement Ecosystem

Online child exploitation is one of the most extensively documented crime categories in the federal enforcement ecosystem. It is also one of the least studied at the case level. Understanding why requires understanding what the existing

infrastructure does, and challenges the system faces.

How ICAC Enforcement Works. The primary law enforcement response in the United States is organized through the Internet Crimes Against Children Task Force Program: 61 federally coordinated task forces representing over 5,400 federal, state, and local agencies, established in 1998 alongside the NCMEC CyberTipline and expanded continuously since. In fiscal year 2024, ICAC task forces conducted approximately 203,467 investigations and arrested more than 12,600 suspected offenders. Investigations are classified as proactive—investigators initiating contact in undercover capacity—or reactive, triggered by a tip or complaint. The FBI's Violent Crimes Against Children program and Homeland Security Investigations operate in parallel for multi-jurisdictional cases.

The statutory framework has three main layers. The PROTECT Our Children Act of 2008 codified the ICAC program and established mandatory reporting requirements. 18 U.S.C. § 2258A requires electronic service providers to report apparent CSAM to the CyberTipline. The REPORT Act of 2024 extended those obligations to online enticement and child sex trafficking for the first time. In 2024, NCMEC received approximately 20.5 million CyberTipline reports — equivalent to 29.2 million separate incidents — from platforms operating globally. Against that volume, 12,600 arrests represents successful case resolution for a fraction of what is reported. This is not a failure of law enforcement. It is a scale problem: the volume of suspected exploitation arriving through the CyberTipline vastly exceeds the number of investigators available to act on it, and triage quality determines which offenders are caught and which victims get reached.

The Response Infrastructure. The technology ecosystem for child protection operates on three planes.

The first is content detection. PhotoDNA, developed by Microsoft in 2009, generates perceptual hashes of images and compares them against a database of known CSAM, detecting matches even after modification. It is deployed by Google, Meta, Twitter, Reddit, Discord, and dozens of other platforms. Google's CSAI Match performs the equivalent function for video. Meta's open-sourced PDQ and TMK+PDQF algorithms extend perceptual hashing to smaller platforms that cannot build detection infrastructure independently. Thorn's Safer platform takes a complementary approach: perceptual hashing against a database of over 82 million verified CSAM hashes combined with an AI-based classifier designed to identify novel material that has never been hashed, with cross-platform hash-sharing that propagates newly identified CSAM across the detection ecosystem without delay. According to the Tech Coalition's 2023 annual survey, 89% of member companies

deploy at least one image hash-matcher; 59% use video hash-matching.

This infrastructure is the backbone of CSAM detection at scale, and its effectiveness for known material is well-established [19]. It also has four structural limits that become more significant as the threat landscape evolves. Hash-matching requires a prior identification — new material, including AI-augmented content, has no hash to match. Detection only occurs on cooperating platforms — encrypted messaging services, smaller platforms, and dark web networks are outside the detection envelope. Hash-matching is also increasingly vulnerable to adversarial attack: recent research has demonstrated that gradient-based image modifications can defeat PhotoDNA detection while leaving content visually intact. And hash-matching answers exactly one question: is this image known illegal material? It cannot answer any of the investigatively important questions — which platform capability made this offense possible, who created this material, how does this case relate to the hundreds investigated last year across the same platform.

*The Behavioral Research Base.*The second major body of work is empirical. The Crimes Against Children Research Center at the University of New Hampshire produced the most systematic research on online CSEA offending. Wolak et al. [23] documented the dynamics of internet-initiated sex crimes through a national law enforcement survey: who offends, how online relationships develop, how offenses progress from digital contact to physical harm. Wolak et al. [24] established that most internet-initiated sex offenses do not fit the predator-deception model that dominates prevention messaging — most offenders did not disguise their age or intentions but cultivated relationships over time, exploiting adolescent developmental vulnerability rather than deceiving victims about who they were. Wolak and Finkelhor [22] and Wolak et al. [25] characterized sextortion as an emerging and distinct offense typology with its own dynamics and victim population. Kloess et al. [10] provided a comprehensive framework for the process of online sexual grooming across internet communication platforms: prevalence, offense stages, offender characteristics, and legal responses across jurisdictions. Kloess et al. [11] mapped offense processes specifically through internet communication platforms, tracing how offenders move victims from initial contact through escalation. Livingstone and Smith [12] documented the intertwining of digital and physical risk for children at national scale. Mitchell et al. [13] characterized the law enforcement response to internet-facilitated commercial sexual exploitation using a nationally representative sample of agencies.

This body of work established the behavioral landscape of online CSEA. It did not — because the data and tools did not yet exist — analyze the relationship between specific platform capabilities and the offense patterns those capabilities make

possible across thousands of cases. The affordance lens has itself been applied to CSEA conceptually — by psychologists modeling technology-mediated abuse [17], and by criminologists theorizing the mechanics through which technological design "invites" technology-facilitated violence [26]. What no prior work has done is connect behavioral patterns to platform capabilities via an affordance-level abstraction, operationalized against enforcement case records at scale.

*Operational Forensics.*The third body of work is forensics. After a device is seized, investigators use enterprise forensics platforms to examine it: Nuix for large-scale e-discovery and timeline reconstruction, Magnet Axiom for mobile and cloud forensics, Cellebrite UFED for physical and logical extraction. The workflow is precise — physical images of storage media, file carving to recover deleted content and fragments, parsing of application databases and encrypted chat logs, artifact recovery from unallocated space, timeline reconstruction from access metadata and system logs. These platforms are powerful, expensive, and infrastructure-heavy. They ask what is on this phone. Not what this case shares with five hundred cases investigated last year across the same platform.

All three layers of this infrastructure carry an acute human cost. Investigators in child exploitation units routinely encounter material that has no equivalent in other areas of law enforcement. Digital forensics analysts do not merely view this material — they analyze it, build timelines, and construct the evidentiary record that prosecution requires, with consequences including emotional numbing, hypervigilance, intrusive ideation, and accelerated burnout. The volume of evidence has grown as the threat landscape has expanded — a single device seizure now routinely contains hundreds of thousands of images and terabytes of cloud-synced material. The cognitive and psychological load scales accordingly. Triage tools that reduce manual exposure are not only an efficiency argument. They are a wellbeing argument.

The harm for victims in these cases does not end with the offense. Material produced circulates long after investigations close. Survivors navigate criminal proceedings, civil systems, and the ongoing reality that digital content does not disappear. Investigators carry the cases they have worked. Families carry what those cases revealed about someone they trusted. These are not incidental facts about a difficult field. They are part of why the harms this paper addresses matter. Every unworked CyberTip is a case that may go uninvestigated. Every missed cross-case pattern and unreported platform capability is an offense trajectory that could have been identified earlier. The infrastructure described here — detection, enforcement, behavioral research, operational forensics — is staffed by capable, motivated professionals working under immense constraints of scale, an evolving

threat landscape, and the sustained weight of radioactive material. The following section introduces the ontological and affordance-based framework that makes cross-case, platform-level analysis possible at corpus scale — without the direct exposure that makes engagement in this field so difficult to sustain.

2.2 What Technology Makes Possible

"Platform" is the wrong unit of analysis for this problem.

Platforms emerge, scale, and disappear. Myspace is gone. Omegle was shut down. Discord did not exist fifteen years ago. If the analytical unit is the platform, every new platform generation requires research from scratch. The offense mechanics do not restart with each platform generation. The capabilities that appear consistently in the enforcement record — anonymity, ephemerality, unmonitored communication, contact discovery, distribution infrastructure — appear on every platform, implemented differently but functioning in similar manners. The unit that holds across platform generations is affordance.

The concept of affordance originates in ecological psychology. Gibson [5] proposed that organisms do not perceive their environment as a collection of objects with fixed properties. They perceive action-possibilities: what the environment affords the organism. A door handle affords gripping. A cliff edge affords a scenic viewpoint for a hiker, a jumping point for a daredevil, and a fatal fall for the unwary. The affordance is relational: it exists between what the environment makes available and what the organism can do with it. Norman [16] translated this to design: artifacts communicate possibilities for action independent of the designer's intent. Children can press a button before anyone explains what it does — the design affords the action. Hutchby [9] extended this to technology: artifacts have material properties that shape what users can do with them. However, affordances emerge from the interaction between what the artifact makes available and what the actor brings to it, not from the designer's intent alone. Bucher and Helmond [2] applied the framework to social media platforms specifically, distinguishing high-level affordances — what the platform as a whole makes possible — from low-level affordances embedded in specific interface features.

The translation to this domain is direct. End-to-end encryption, designed to protect private communication for billions of legitimate users, affords unmonitored contact: communication that no platform can scan and no investigator can access without a court order. That affordance exists on WhatsApp, Signal, Telegram, and in the private message systems of every major platform. Timer-based content deletion, designed to make casual sharing feel lower-stakes, affords ephemerality: communication without a persistent record. On Snapchat, on Instagram Stories,

on WhatsApp — material sent disappears. In the enforcement record, this is evidence destruction by design. Account creation flows that require no verified identity, built to reduce onboarding friction, afford anonymity: action without accountability. File hosting and distribution infrastructure, built for sharing photos and documents, affords mass distribution of any content to anyone with a link. Location-based discovery features, built to surface nearby users for social or dating purposes, afford proximity contact with strangers. Generative AI tools, built to expand computing possibilities and revolutionize content creation, afford the production of synthetic exploitative material, the fabrication of convincing personas at scale, and the automation of contact and grooming that previously required sustained human attention — capabilities that did not exist at this level of accessibility five years ago.

Different platforms. Expanding capabilities. Same function in the enforcement record.

Platforms are not designed to be exploitation-resistant. They are designed for engagement, retention, and scale [7]. The product decisions that produce each of these affordances are not made especially for offenders, or for any specific subset of the hundreds of millions of users. The capability exists for everyone. Misuse is not a design failure in the conventional sense; rather, it is an emergent consequence of building for everyone when the population includes people who have already decided to harm.

The affordance names what the platform offers and the enforcement record provides evidence for how offenders misappropriated them. Section 3 presents both at scale.

3. Evidence at Scale: Analysis of 7,426 Exploitation Cases

3.1 Collection, Processing, and Validation

The corpus underlying this analysis consists of 7,426 publicly available Internet Crimes Against Children case records collected from 56 law enforcement sources spanning 2002 to 2026. Sources include task force press releases, agency annual reports, DOJ CEOS announcements, NCMEC public outputs, and federal court filings. All records are publicly available. No private case data, investigative files, or victim-identifying information is included in any form. The collection and analysis protocol operates in accordance with HRPO NHSR Determination #7668; this research does not contain private or personally identifiable information under

federal regulations [45 CFR 46.102(f)(1),(2)].

Source selection targeted the full range of the U.S. ICAC enforcement ecosystem: 61 task forces, federal pipelines including Army CID, Customs and Border Protection, ICE/HSI, the U.S. Secret Service, and the U.S. Marshals Service, and 2,864 unique law enforcement agencies. Collection expanded from ICAC-specific queries to broader child exploitation search terms, yielding 56 sources and a corpus representing 24 years of enforcement activity.

Processing operates through a hybrid deterministic and interpretive machine-learning pipeline across three stages.

The first stage is regex-based extraction. Deterministic rules extract structured fields from case narratives across seven categories:

- **Perpetrator signals:** age, registered sex offender status, gender, and multi-defendant flags
- **Victim signals:** age, age ranges, stated victim count, and gender after validation
- **Relationships:** kin, role, or stranger when not stated
- **Prosecution:** charge phrases with counts and booking stage from arrest through sentencing; sentence durations
- **Evidence volume:** image, video, storage, and message counts when quantified in text
- **Platform identification:** named applications and surfaces — social media, messaging, gaming, file hosting, livestreaming, early-era chat clients, and generative AI tools when cited — alongside generic labels (online, chat, social media) when no product match is found
- **Technology signals** (stored separately from the platform list): investigation tooling (PhotoDNA, hash-matching, CyberTipline language), anonymization infrastructure (Tor, dark web, cryptocurrency), and P2P clients

The second stage is pattern-based classification. Each case receives topic labels (production, possession, distribution, trafficking, CSAM, AI-generated CSAM, sextortion, hands-on versus online-only, family versus stranger, international, multi-state) and severity indicators (infant, very young, under 12, sexual abuse, multiple perpetrators). Severity phrases used in triage priority scoring are extracted separately: language indicating active, ongoing, or escalating abuse is weighted in the priority model.

The third stage is ML enrichment. When the ML stack is enabled, NER via Stan-

ford NLP (Stanza) adds organizations, locations, dates, and ages, merged with regex output under a precedence model that preserves deterministic extractions when conflicts arise. A victim-age gate drops decoy and headline ages introduced by undercover investigation language. Semantic sentence scoring attaches concept weights to each case; grooming severity tags and strong possession or AI-generation language reinforce corresponding case topics. Concept scores are retained for downstream analysis; a curated subset are merged into main fields.

Extraction is context-aware throughout, distinguishing investigation types, victim from perpetrator gender, and perpetrator admissions where explicitly present. Agency normalization resolves over 2,800 unique agency string variants to canonical identifiers using a combination of pattern matching and NER, enabling cross-source aggregation across jurisdictions that name the same agency differently across releases.

The primary evidence base for platform harm analysis (Q1) is a candidate pool selected by automated filter, not manual case selection. A case enters the pool if its structured metadata contains at least one non-generic platform label — excluding generic placeholders ("online," "internet," or generic tags). This yields 1,875 candidate cases (25.1% of the corpus) spanning 54 named platforms; the remaining 74.9% lack a named platform and fall outside affordance-level analysis due to insufficient signals for supporting claims at the level of a specific platform capability. This is partly a property of the source material itself: public case documents systematically under-report platform and offense detail to avoid influencing offending behavior, compromising ongoing investigations, or revealing prosecutorial strategy (discussed in Section 7.4). Restricting Q1 to cases with a named platform trades corpus breadth for evidentiary precision as affordance-level analysis requires knowing which platform, and which of its features, was actually present.

Each candidate is processed by an evidence extractor that strips PSA and awareness-campaign boilerplate, extracts a supporting quote, assigns an offense role, and classifies evidence strength at the platform–case level. The candidate-to-evidence pipeline yields a final Q1 evidence base of 1,875 cases across 3,128 platform–case records.

Evidence strength is tiered. *Stated* records link a platform instrumentally to offense conduct in the same sentence — through an explicit construction (*through*, *via*, *using* a platform) or an offense keyword in close proximity, excluding awareness framing and enumerated app lists. *Inferred* records show platform and offense language co-occurring without instrumental linkage. *Named-only* records carry a metadata platform tag without qualifying textual support. At the case level,

856 cases (45.8%) contain at least one stated record, 134 (7.2%) carry inferred-only evidence, and 881 (47.1%) are named-only. The affordance labels applied in Q1 are a manual analytical layer over this automated extraction: the pipeline supplies quotes, roles, and tiers; affordance semantics are assigned by hand.

The full pipeline — extraction rules, classifier weights, normalization dictionaries, and reproducibility documentation — is released as open-source software under the MIT License at github.com/mrinaalr/CaseLinker.

3.2 Ontologies for Modeling Radioactive Data

CaseLinker extracts over 80,000 structured features across 10 dimensions and stores thousands of cases in a PostgreSQL database for analysis. However, tabular representations are insufficient for the questions this paper asks. A case record is not a row. It is a set of relationships: an offender reached a victim through a platform, used a capability to produce material, and distributed it through infrastructure. A relational table can store the tags, actions, and data. It cannot natively represent the structure and relationships between fields. The questions Q1's affordance-based platform harm analysis asks — which capabilities cluster with which offense types, which platform features co-occur across cases, how offense mechanics distribute across a 24-year enforcement record — are relational questions. They require a data model that treats relationships as first-class objects.

Knowledge graphs provide that model. Each case in this corpus is represented as a directed, typed graph: nodes for entities (offenders, victims, platforms, charges, agencies), edges for relationships (`contacted_via`, `charged_under`, `investigated_by`), and typed properties for attributes. This representation enables the pattern queries Q1 requires: offense events, platforms, agencies, and actors as pieces of an interconnected system, not isolated records.

The vocabulary that models these cases is the Crimes Against Children (CAC) Ontology, shepherded by Project VIC International as an interoperable standard for representing crimes-against-children investigations. The stack is layered: gUFO provides foundational ontology primitives. UCO (Unified Cyber Ontology) models cyber investigation objects and relationships. CASE (Cyber-investigation Analysis Standard Expression) defines how investigation narratives are expressed as knowledge graphs. CAC extends that stack for crimes against children specifically — platforms, victims, offenders, investigations, and outcomes as typed entities rather than free text.

CaseLinker maps each extracted case feature to a CAC class. Platform mentions become typed `cac:Platform` nodes with semantic relationships to offense events.

Statutory charges map to identifiers linked to offense type taxonomies. Perpetrator behaviors map to grooming stage classes. The mapping is explicit and auditable: every ontological assertion traces to a source extraction and a source text span. There are no black-box inferences. Structural validation is performed using SHACL (Shapes Constraint Language) against a curated subset of the CAC module stack; each case graph is validated prior to inclusion in the Q1 evidence base, and non-conforming graphs are flagged and excluded rather than silently retained.

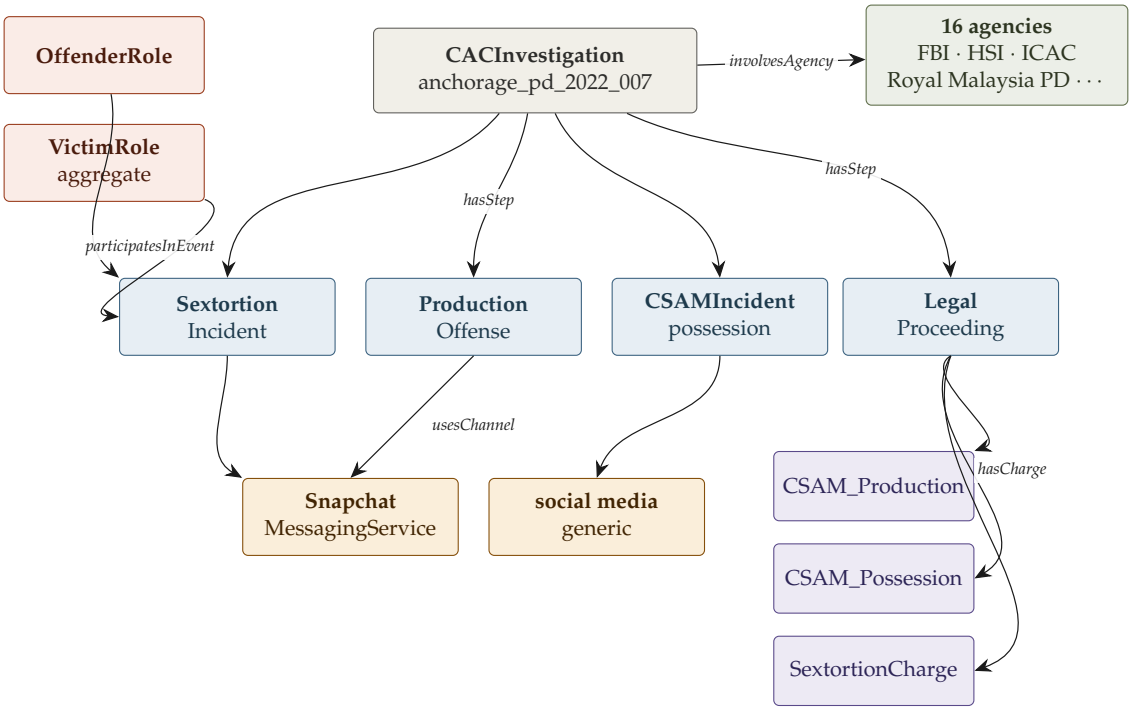

Figure 1: CAC ontology representation of a single case as emitted by the `case2cac` pipeline, prior to court filling enrichment (case `anchorage_pd_2022_007`).

The result is a corpus of 1,500+ SHACL-validated case graphs forming the evidence base for Q1 analysis. Graph construction at corpus scale is handled by the CaseLinker MCP (Model Context Protocol) server: 34 case2cac tools for automated graph construction, validation, traversal, and export from press-release and case-record sources.

Primary-source court record representations — dockets, statements of offense, and federal indictments retrieved from primary court filings — are handled by the CASE-UCO SDK directly. The CASE-UCO SDK is a typed, multi-language library (Python, C#, Java, Rust) implementing 428+ ontology classes across the CASE/UCO standard, designed for constructing and validating JSON-LD knowledge graphs for digital forensics and cyber-investigation workflows. Where the

case2cac tools operate over structured fields at scale, the CASE-UCO SDK is used for precision representation: extracted facts from primary federal court documents are mapped to typed ontology classes, producing ground-truth graphs that anchor corpus-level pattern analysis. The specific facts-to-graph process for each PACER case is detailed in Section 5.

3.3 Research Questions and Analytical Approach

Three empirical and one theoretical question organize the analysis. Each draws on a distinct evidence layer: the 7,426-case corpus for breadth, the 1,500+ validated knowledge graphs for structured pattern analysis, and four primary-source federal court records for ground-truth harm-signature anchoring.

Q1 — Platform Affordances. Which platform capabilities do offenders exploit most consistently, across case records, offense types, and the full period this corpus covers? Q1 operates at the platform level, drawing on every case with at least one identified platform in the enforcement record. The 30+ platform affordance table maps each platform to its documented misuse surfaces and harm vectors across stated, inferred, and named-only evidence tiers. The knowledge graph representation enables this analysis without reducing to frequency counts: a query across the graph can ask which capability types co-occur with which offense types, which platforms share affordance profiles, and how those profiles shift across offender typologies and agencies the corpus represents.

Q2 — Exploitation Life Cycle. How is technology implicated across the four offense subtypes — grooming and enticement, sextortion, CSAM production, and coordinated criminal networks — from first contact through prosecution? Q2 operates at the harm-signature level. The primary evidence base is four federal district court cases, retrieved from Public Access to Court Electronic Records (PACER) and modeled via the CAC Ontologies:

- *United States v. Rehman* (D.D.C. 1:23-cr-00064) — §2422(b) enticement; platform-mediated grooming and relationship formation
- *United States v. Amin* (D. Alaska 3:22-cr-00055) — 13-count §2252A(g) enterprise; sextortion at scale across more than 80 Snapchat and 40 Instagram accounts
- *United States v. Pathmanathan* (D.D.C. 1:22-cr-00150-JEB) — §2251(a),(e) production; Instagram-to-Messenger platform migration and live stream directed self-production
- *United States v. Bermudez et al.* (E.D.N.Y. 1:25-cr-00361) — six-defendant §2252A(g) coordinated network; Discord and gaming platform infrastructure

Each case was mapped to the CAC ontology independently of the automated pipeline, anchoring corpus-level patterns against primary federal court records.

Q3 — Disruption. Where in the platform stack and enforcement pipeline do realistic opportunities to disrupt exploitation exist? Q3 synthesizes Q1 and Q2: given which capabilities offenders consistently reach for, and how those capabilities combine into recognizable offense mechanics, which affordances are most amenable to design-level, policy-level, or enforcement-level intervention? Q3 is the operational output of the analysis — the translation from documented pattern to actionable target.

Q4 — Formal Framework. Can the relationship between platform affordances, offender goals, and victim-facing harm be expressed in a generalized mathematical model for exploitation? Q4 operates at the level of abstraction, formalizing the patterns the corpus documents. The mathematical framework in Section 7 maps offender goals to affordance trajectories (φ), exploitation types to harm sets (η), and affordance classes to co-occurring harms (ψ) — producing a marginal exploitation utility measure $u(a_{\text{new}}, g)$ that is estimable before a feature enters the enforcement record.

Q1 characterizes the affordance landscape. Q2 anchors it in the court-verified mechanics of four distinct offense types. Q3 asks what can be done. Q4 builds the vocabulary to reason about exploitation before it occurs.

Practitioners and policy readers may proceed directly to Section 4 for the platform analysis and Section 6 for intervention recommendations; the formal framework in Section 7 is self-contained and can be read independently.

4. Platform Affordances and Misuse Surfaces

The enforcement record names 54 distinct platform labels across the Q1 evidence base. Analyzed individually, they appear chaotic: products from different eras, built by different companies, serving different purposes. Analyzed by affordance, they resolve into a small number of recurring signatures. A platform's affordance profile predicts its harm profile. The same chat affordance, implemented on a platform from 2008 or one from 2025, produces the same offense pattern in the record.

Table 1 presents the platform manifest: every analyzed platform with sufficient case representation, its affordance signature, the misuse surface that offenders utilize, and the victim-facing harm vectors documented in the corpus. Three

definitions discipline the table. An **affordance** is a neutral capability the platform's design makes possible. A **misuse surface** is how an offender turns that affordance toward an offense; it is offender-facing. A **harm vector** describes what happens to a child; it is victim-facing, drawn from a fixed set established across the corpus, and ranked by documented prevalence. Every platform with documented harm to more than two children receives analysis: each represents real victims and a real affordance pattern.

Platform affordance manifest. Each platform's defining affordance signature, the misuse surface it creates, and the victim-facing harm vectors documented in the corpus, ranked by prevalence.[1]

Platform	Affordance Signature / Misuse Surface	Documented Harm Vector
Messaging Services		
Snapchat 126 cases	*Ephemerality, camera-first capture.* Evidence destruction by design; real-time coercion of imagery; ephemeral solicitation with reduced forensic trail.	Child contacted and exploited through ephemeral messaging Child's abuse imagery distributed/re-shared Child groomed toward contact
Kik 124 cases	*Anonymity, group discovery.* Pseudonymous contact without traceable identity; closed-group trading; account churn to evade detection.	Child's abuse imagery distributed/traded (re-victimization) Child contacted through anonymous messaging Child groomed toward contact
Discord 29 cases	*Server/community structure, persistent contact.* Closed servers as low-visibility spaces; coordination among offenders; sustained contact enabling coercion.	Child's abuse imagery distributed/traded (re-victimization) Child contacted through servers and DMs Child sextorted under threat Child groomed toward contact
Telegram 8 cases	*[affordances pending]*	*[pending analysis]*
WhatsApp 6 cases	*[affordances pending]*	*[pending analysis]*

continued on next page

[1]The manifest measures *documented offender use of platform capabilities*, not ground-truth harm frequency. Possession and distribution of CSAM dominate the record partly because they are the most detectable evidence of harm and play a significant charging role after being surfaced through the CyberTip pipeline; contact and grooming are systematically under-detected and under-documented in public records. Affordances and misuse surfaces are derived from analysis of each platform's capabilities; harm vectors are drawn only from offenses the corpus actually documents. Platform counts are a lower bound on involvement, not a complete picture.

Table 1 (continued)

Platform	Affordance Signature / Misuse Surface	Documented Harm Vector
Facebook Messenger 6 cases	*[affordances pending]*	*[pending analysis]*
IRC 6 cases	*[affordances pending]*	*[pending analysis]*
Skype 4 cases	*[affordances pending]*	*[pending analysis]*
Social Media Platforms		
Facebook 45 cases	*Real-name profiles, friend graph, minimal verification.* Fake/impersonation accounts; profile-mining to target; Messenger as private channel.	Child's abuse imagery distributed/re-shared Child contacted through profiles and Messenger Child groomed toward contact
Instagram 21 cases	*Visual-first profiles, DM, discoverability.* Visual profiles to select targets; DM coercion; disappearing messages reduce forensic trail.	Child contacted through DMs and visual profiles Child's abuse imagery distributed/re-shared Child sextorted under threat
Twitter / X 15 cases	*[affordances pending]*	*[pending analysis]*
Reddit 7 cases	*[affordances pending]*	*[pending analysis]*
TikTok 4 cases	*Algorithmic discovery, DM, public-by-default.* For You feed surfaces minors without active search; public videos to select targets; DM to move contact off the public feed.	Child contacted and exploited through messaging Child groomed toward contact or meeting Child solicited for explicit imagery
File Hosting Services		
Dropbox 20 cases	*Cloud storage, public link sharing, desktop sync.* Shared links distribute CSAM collections without direct transfer; desktop sync client mirrors local library to cloud automatically; handoff layer — link shared via separate contact channel.	Child's abuse imagery possessed or stored in offender's cloud account Child's abuse imagery uploaded for distribution or collection Child's abuse imagery distributed or re-shared via shared links (re-victimization)
Google Drive 12 cases	*Cloud storage, shareable links, persistent hosting.* Shared links distribute CSAM without direct file transfer; folder-level bulk sharing; handoff layer — material hosted on Drive, link shared via separate contact channel.	Child's abuse imagery possessed or stored in offender's cloud account Child's abuse imagery uploaded for distribution or collection Child's abuse imagery distributed or re-shared (re-victimization)

continued on next page

Table 1 (continued)

Platform	Affordance Signature / Misuse Surface	Documented Harm Vector
Mega.nz 4 cases	*[affordances pending]*	*[pending analysis]*
Anonymous Chat Platforms		
Whisper 3 cases	*Anonymity, location-based discovery.* Anonymous contact with minors without traceable identity; location proximity used to surface nearby targets; DM moves contact off public feed into private channel.	Child contacted and exploited through anonymous messaging Child groomed or enticed toward sexual contact or meeting Child solicited for explicit imagery through anonymous posts or direct messages
Omegle 2 cases	*Anonymous stranger pairing, unmonitored video.* Random pairing reaches minors without targeting effort; zero-identity architecture leaves no forensic trail; unmonitored video channel used for real-time solicitation and exhibition; migration target to persistent platforms.	Child contacted and exploited through anonymous random video or text chat Child groomed or enticed toward sexual contact or meeting Child solicited for explicit imagery or subjected to real-time sexual exhibition in video chat
AI Services		
Gen AI 14 cases	*Generative synthesis, likeness manipulation.* Synthetic CSAM from prompts; nudification of real photos; model fine-tuning on a target's images.	Real child's likeness manipulated/synthesized into abuse imagery without contact Child's abuse imagery (incl. AI-generated) distributed/possessed
Additional Surfaces		
P2P (BitTorrent) 3 cases	*[affordances pending]*	*[pending analysis]*
Video Streaming 9 cases	*[affordances pending]*	*[pending analysis]*
Gaming Platforms 7 cases	*Interactive user community, built-in chat, low-friction contact.* In-game chat used to initiate contact with minors under cover of shared gameplay; pseudonymous accounts with no identity verification; migration target — contact initiated in-game, exploitation moves to Kik, Whisper, or anonymous messaging platforms.	Child contacted and exploited through in-game chat or messaging Child groomed or enticed toward sexual contact or meeting Child's exploitation migrates off-platform to anonymous messaging after initial in-game contact

continued on next page

Table 1 (continued)

Platform	Affordance Signature / Misuse Surface	Documented Harm Vector
Online Dating 2 cases	*Contact discovery, identity misrepresentation.* Dating platforms used to initiate contact with minors through fake or age-misrepresenting profiles; age-gate bypass on platforms where minors are present; migration target — contact initiated on dating surface, exploitation moves to messaging or storage platforms.	Child contacted and exploited through dating platform messaging Child groomed or enticed toward in-person meeting Child's abuse imagery solicited or exchanged following platform-initiated contact

4.1 Contact and Approach Affordances

4.2 Possession and Trade Affordances

4.3 Coordination Affordances

4.4 Production Affordances: The Generative Evolution

5. Harm Signatures: How Technology Shapes Exploitation

The platform manifest in Section 4 maps capabilities to misuse surfaces. It answers a static question: what does each platform make possible? This section asks the harder one. Across four offense subtypes — grooming and enticement, sextortion, the production of abuse material, and coordinated criminal enterprises — how do those affordances actually combine during the mechanics of an offense, from first contact through prosecution? Each subsection develops a **harm signature** for its offense type: a generalized exploitation lifecycle mapping the steps an offender takes at each stage to the victim-facing harms those actions leave behind. The evidence base is four federal district court cases retrieved from PACER and read in full, each mapped to the CAC ontology independently of the automated pipeline. They are ground-truth anchors within the same analytical framework: primary-source court records that document, in precise legal language, exactly which capabilities were used, in which sequence, to reach which victims. The corpus supplies breadth. These four cases supply the mechanism.

5.1 Grooming and Enticement

Grooming is not one offense type among four. It is the substrate on which every offense type in this section is built. The sustained contact that makes production possible is grooming. The deception and contact that make sextortion's first image possible draw on the same mechanics as grooming. The recruitment mechanics that allow a coordinated enterprise to scale across victims are grooming. What changes across offense types is what the offender does after the relationship is established. The relationship itself is always the same structure: a child brought, through sustained deception or threats, to a state of compliance she would not have reached otherwise.

The grooming trajectory is also the most probabilistic of the four. A sextortion offender who holds retained imagery has a structural advantage that does not depend on the victim's continued willingness. A producer who has already directed explicit conduct has created evidence that binds the victim through fear. A grooming offender has none of that leverage in the early stages. Every request is a gate. Every gate is a compliance decision the victim can refuse. What the offender is managing, across every contact and every escalation, is a probability distribution over victim responses — and the grooming mechanics exist precisely to shift that distribution. Flattery raises the probability of the next compliance. Reciprocal disclosure lowers the victim's sense of asymmetry. Persistence reduces the exit window. Exploitation of disclosed vulnerability — depression, isolation, prior trauma — narrows the space between compliance and refusal.[2]

The chain documented in this case runs from first contact through flattery and low-stakes requests, through escalating solicitation and directed production, to meeting arrangement and hands-on contact — all within approximately five weeks on a single platform. At no point in this chain did the offender use threats, quotas, or distribution leverage. Compliance was obtained entirely through relationship investment: sustained presence, flattery calibrated to disclosed vulnerability, and the normalization of escalating requests through reciprocal disclosure. Each compliance gate passed lowered the probability of exit at the next one. The trajectory did not require coercion in the technical sense because the relationship cultivation had already done the coercive work before any explicit demand was made.

The terminal harm state here is physical contact, not perpetual online coercion. This distinguishes the enticement-to-contact trajectory from sextortion, where the

[2] *United States v. Rehman*, D.D.C. 1:23-cr-00064-CJN, Statement of Offense (filed Nov. 21, 2024). Structural observations in this subsection derive from the Statement of Offense read in full. No victim-identifying information appears here.

terminal state is retained material held under constant threat of release. In this case, the CSAM produced during the online phase is a precursor to the contact goal — a stage in escalation, not the instrument of continued exploitation. The offense does not cycle. It progresses to a terminus, and that terminus is physical. Four documented exit points existed in the chain. The offender navigated all four through the same mechanism: sustained relationship investment targeted at a child whose vulnerability had been disclosed and whose trust had been deliberately cultivated.

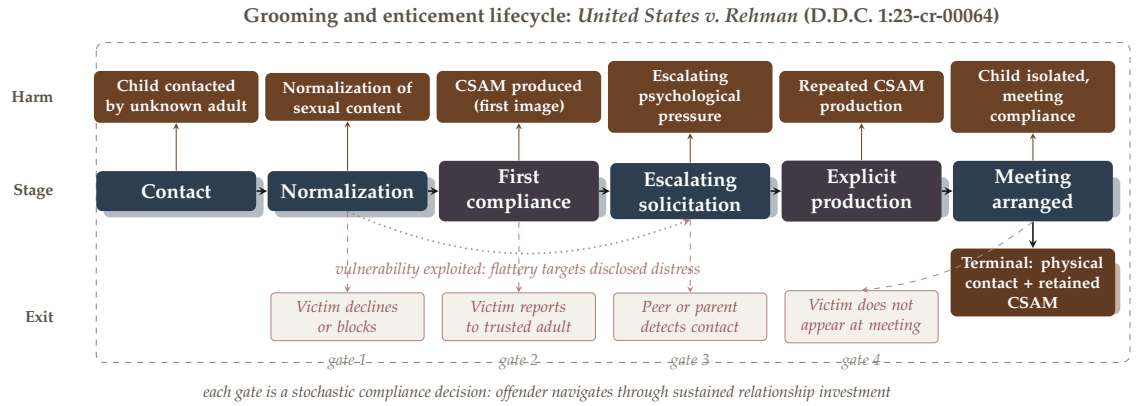

Figure 2: Grooming and enticement lifecycle: *United States v. Rehman* (D.D.C. 1:23-cr-00064) mapped onto the generalized exploitation lifecycle. Gate nodes (purple) mark stochastic compliance thresholds where the trajectory could have collapsed. Exit branches below the spine show documented disruption opportunities at each gate. The vulnerability arc denotes offender adaptation to disclosed victim distress across the escalation phase.

5.2 The Production of Abuse Material

Production is the offense of directed creation of CSAM. The material does not preexist it. The offender engineers its existence by coercing a victim to perform explicit conduct on camera, or by filming in-person abuse directly. The act of exploitation becomes a permanent artifact the moment capture occurs. The production event is not the contact, the escalation, or the coercion preceding it — it is the conversion of an abuse event into persistent material. The lifecycle proceeds in two phases.

The remote, online-directed capture is the dominant production modality in this corpus. The first phase is access. The offender establishes contact on a public surface and escalates through the grooming and social engineering pathways documented in Section 5.1. Communcation then shifts to a channel capable of

image or video production. The contact surface is for access. The production channel is for capture.

The second phase is directed production. The offender issues commands during the live session. The victim performs them on camera. The offender records the session.[3] When the victim resists, the offender threatens to distribute previously captured material or escalates threats and compliance resumes. When a victim blocks contact entirely, the offender creates a new account and returns to the contact stage. The lifecycle does not break. It resets.

The production lifecycle terminates in an archive. Material captured across sessions is organized and held outside any channel the victim can access. Each session adds to it. The harm does not close when the final session ends. It persists as leverage for every contact that follows.

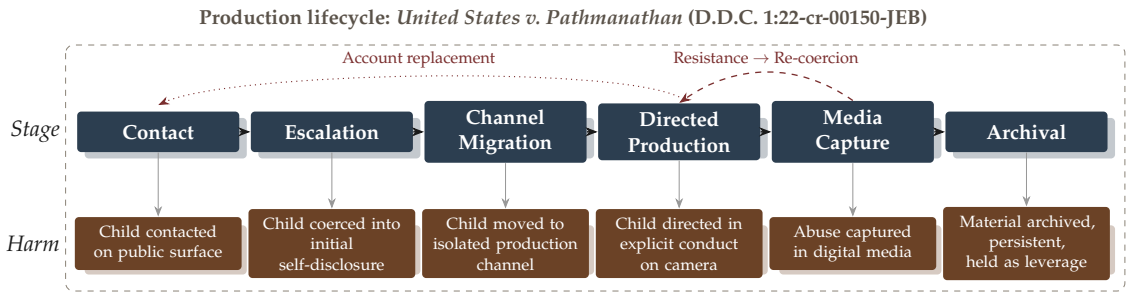

Figure 3: Production lifecycle: *United States v. Pathmanathan* (D.D.C. 1:22-cr-00150-JEB) mapped onto the generalized exploitation lifecycle. Dashed arc denotes the resistance and re-coercion loop within an active session; dotted arc denotes account replacement when a victim blocks contact entirely.

5.3 Sextortion: Coercion After Contact

Sextortion is structurally distinct from every other offense type in this corpus. The harm does not begin with the production of material — it begins before, in the act of deception and coercion that makes production possible. And unlike grooming, which moves toward a physical meeting, or production, which terminates when material exists, sextortion is self-sustaining: the material produced in the first exchange becomes the instrument of the next. The chain feeds itself. Every image a victim sends under threat becomes leverage for demanding another.

The mechanism is coercion under retained threat. A first image is obtained —

[3] *United States v. Pathmanathan*, D.D.C. 1:22-cr-00150-JEB, Statement of Offense (filed Jan. 30, 2026). Structural observations in this subsection derive from the Statement of Offense read in full. No victim-identifying information appears here.

through impersonation, false promises, or social engineering. Once that image exists, the offender's position is asymmetric: the victim cannot undo the disclosure, and the offender controls whether it spreads. That asymmetry is the engine of the offense. Threats to expose the image to the victim's social network — friends, family, schoolmates — convert a single act of disclosure into a sustained compliance mechanism. The victim is not coerced once. They are coerced continuously, each demand backed by the same retained material and the constant threat of its release.[4]

What sustains the loop is not new leverage — it is the same leverage, redeployed. Daily quotas of images and videos, directed conduct on live calls, demands for increasingly explicit material: each round of compliance produces new material, which extends the leverage, which enables the next round. The offense does not escalate because the offender acquires more power. It escalates because the victim has less exit. Platform bans do not break the chain. When accounts are disabled, new ones are created and victims are told to reconnect. The coercion loop persists across platform interventions because the leverage — the retained imagery — exists outside any single account.

The lifecycle documented in this case extends beyond the coercion loop into collection, distribution, and exposure. Material is organized and stored; links are shared with co-conspirators who must harass victims before receiving access. Victims' images are sent to their friends and schoolmates when they attempt to stop. And the chain propagates: victims under active coercion are directed to recruit other minors, supply follower lists, and warn peers that they too will be exposed. One victim helped identify roughly fifteen others. The sextortion lifecycle does not terminate when the offender stops. It terminates, if it terminates at all, when the retained material is no longer reachable or the offender is apprehended.

[4] *United States v. Amin*, D. Alaska 3:22-cr-00055-SLG-KFR (indictment filed Oct. 5, 2022). Structural observations in this subsection derive from the federal indictment read in full. No victim-identifying information appears here.

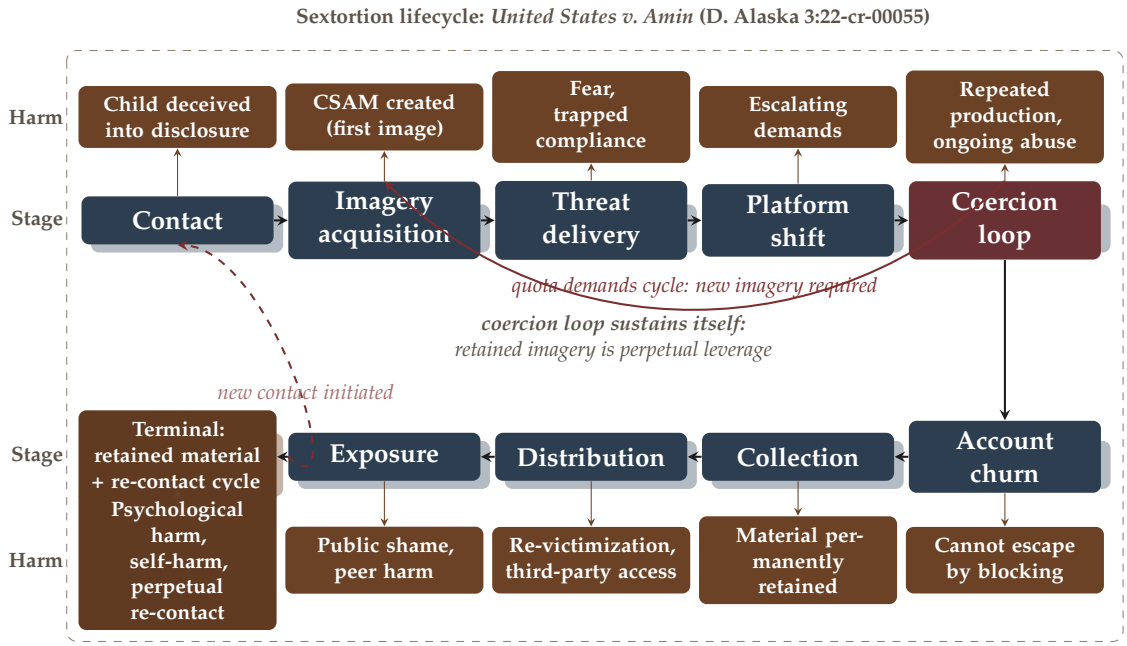

Figure 4: Sextortion lifecycle: *United States v. Amin* (D. Alaska 3:22-cr-00055) mapped onto the generalized exploitation lifecycle. The coercion loop — from threat delivery back to imagery acquisition — is the defining mechanic: retained material is perpetual leverage. Accent arcs denote self-sustaining cycles; the recruitment arc denotes propagation back to the contact stage.

5.4 Coordinated Criminal Enterprises

The preceding three subsections each describe a single offender pursuing a goal through a sequence of actions. Grooming and enticement require sustained contact and trust architecture. Sextortion requires imagery acquisition and coercion leverage. Production requires a victim, a camera, and a channel. In each case, one person traces a trajectory through A toward a goal in $\mathcal{G}$. Coordinated criminal enterprises are structurally different in one respect: the trajectory is parallelized across actors. The affordance set is the same. What changes is that no single offender executes the full sequence. Recruitment, production, distribution, and coercion become roles, each filled by a different member of the enterprise, each operating on a different phase of the same trajectory simultaneously. The platform does not change what exploitation requires. It changes how many people can share the work of doing it.

The coordination affordance makes this possible. Discord's server and channel architecture — designed to let communities organize around shared interests, with granular permission controls over who can access which spaces — functions

in the enterprise context as an organizational infrastructure.[5] Locked channels become internal operational security: the same permission system that lets a gaming community separate beginner and advanced discussion lets an enterprise separate its membership from its operations. Multiple-platform architecture — gaming platforms as recruitment funnel, communication channels as operational headquarters — is a key signature of exploitation that has scaled past what any single offender can execute.

What the enterprise structure enables that solo offending cannot is scale and role specialization. A solo offender who produces, distributes, and coerces is executing each phase sequentially, bounded by individual bandwidth. An enterprise distributes those phases across members, executing them in parallel across multiple paths simultaneously.

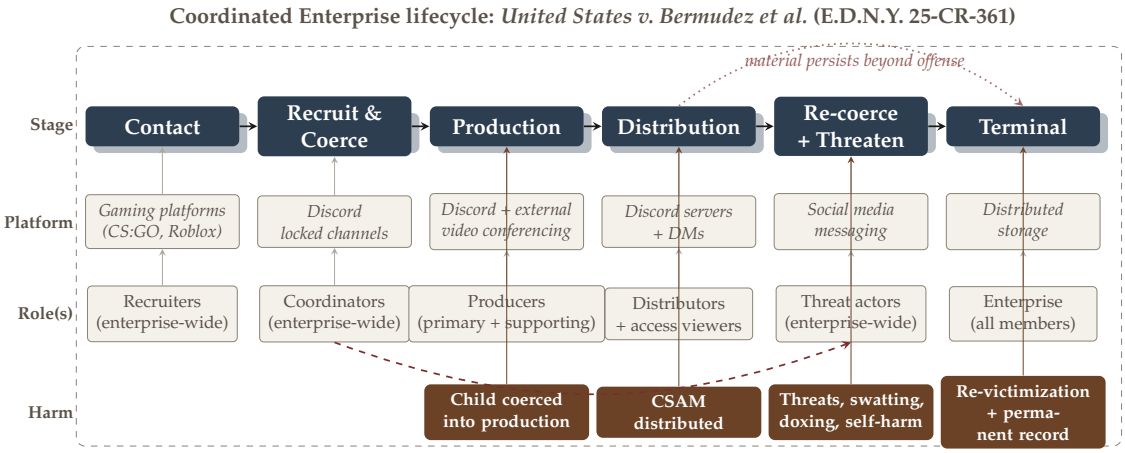

Figure 5: Coordinated enterprise lifecycle: *United States v. Bermudez et al.* mapped onto the generalized exploitation lifecycle. The enterprise distributes a single offense trajectory across role-specialized actors operating in parallel. Dashed arc denotes parallel execution across role-specialized members; dotted arc denotes material persistence beyond the production event.

Across the four harm signatures documented in this section, a pattern holds. The offense mechanics applied differ — grooming operates through sustained contact, sextortion through coercion leverage, production through directed imagery, coordinated enterprises through parallelized roles — but the affordances that enable them are drawn from the same finite set. Anonymity, ephemerality, unmonitored communication, distribution infrastructure, coordination architecture. The same

[5] *United States v. Bermudez et al.*, E.D.N.Y. 1:25-cr-00361 (indictment filed Nov. 18, 2025). Structural observations in this subsection derive from the federal indictment read in full. No victim-identifying information appears here.

capabilities appear at different stages of different trajectories, repurposed by different offender types toward exploitative goals. Section 6 explores where in those sequences disruption becomes possible.

6. Disrupting Offense Mechanisms

6.1 Where Intervention Has Traction

The enforcement infrastructure described in Section 2.1 was built for a specific function: detect known illegal material, report it, and prosecute the offender in possession of it. It performs that function at scale. 35.9 million incidents of suspected CSAM were reported to NCMEC in 2023, alongside 104 million+ related files reported by registered Electronic Service Providers [20]. PhotoDNA, deployed across dozens of major platforms, has been described as highly effective at detecting known CSAM content, with low false positive and false negative rates. The detection layer works.

The investigation layer strains under what detection produces.

The limitation is structural, not operational. Hash-matching answers exactly one question: is this image known illegal material? It cannot answer the questions that determine whether a case is workable — which platform capability made this offense possible, whether the offender is in contact with an active victim, how this case connects to others investigated last year across the same platform. The core challenge is not volume or capacity alone: documented evidence shows that law enforcement officers struggle to accurately triage and prioritize CyberTipline reports, as low report quality makes it difficult to distinguish cases that will lead nowhere from those that could uncover ongoing abuse [8, 15]. That is not a failure of the agencies receiving them. It is a consequence of building detection infrastructure faster than investigation infrastructure, in a threat landscape that has expanded faster than either.

Systematic analysis of the corpus and its affordance patterns reframes where intervention pressure is most productively applied. Three intervention points emerge from the record, each operating under distinct legal and operational constraints.

The first is the detection gap. Intervention at this point is legally well-defined: content that matches a known-illegal hash triggers mandatory reporting under 18 U.S.C. § 2258A and has led to the identification of hands-on abusers, trafficking networks, and ongoing abuse that would otherwise have gone undetected. The

infrastructure is mature. Detection-threshold intervention is necessary but insufficient. It catches the distribution of known material. It cannot reach the contact, grooming, or production phases that precede it.

The second is the policy and investigative layer. Proactive investigation, multi-agency sting operations targeting offenders who travel to meet minors, and peer-to-peer network monitoring has historically produced some of the most significant prosecutions in the enforcement record. The statutory framework has expanded to match the threat landscape: the REPORT Act of 2024 extended mandatory reporting obligations to online enticement and child sex trafficking for the first time, closing a gap that the earlier framework did not address. Policy intervention at this layer is constrained by the same architecture that protects legitimate users — law enforcement cannot monitor communications without legal process, platforms cannot act on suspicion without evidence of an act. Those constraints exist for reasons that extend well beyond this domain. Working within them requires intelligence about modern offense patterns, so that enforcement policy and investigative resources concentrate where the record shows the greatest potential to reach active harm.

The third intervention point is pre-deployment design review. This is the point at which no legal constraint prevents action, no investigative threshold must be crossed, and no harm has yet occurred. Before a feature ships, a platform can ask: does this capability appear in the enforcement record? What exploitation pattern does it produce? What is the marginal exploitation utility of this capability for the most plausible offense goals? The mathematical framework developed in Section 7.2 provides the formal structure for that question. The corpus provides the evidence base. The answers are not speculative — thirty platforms and twenty-four years of enforcement records document what those answers look like across every affordance class in the record.

Pre-deployment intervention is not surveillance. It does not profile users, flag accounts, or produce investigative referrals. It is an engineering question asked before a capability enters the environment, when the cost of mitigation is lowest and the range of options is widest. Age verification at account creation costs less to implement before a platform scales than after. Ephemerality with preserved access for legal process is a design choice that becomes increasingly difficult to reverse once a user base has formed around the privacy expectation. The window for design-level intervention is pre-deployment. After that window closes, the remaining intervention points are reactive by necessity.

The enforcement record does not show that any single intervention point is sufficient. Detection without investigation produces unworkable volume. Inves-

tigation without detection misses the scale of the distribution problem. Policy without design-level mitigation addresses harm after the platform has been in the environment long enough to generate cases. What the record shows is a sequencing problem: intervention has historically arrived after the affordance, after the platform, after the harm. The framework developed in this paper provides the vocabulary to reverse that sequence — to ask the exploitation question before deployment rather than after prosecution. That reversal does not require new law. It does not require new investigative authority. It requires that the question become standard practice in the engineering of new capabilities.

The following subsections address each intervention point in depth.

6.2 The Pre-Deployment Window: Building Against Exploitation

6.3 Policy and Investigative Levers

6.4 Detection, Triage, and Cross-Case Intelligence

7. Designing Against Predation

7.1 The Adaptive Adversary

Most offenders in this corpus are rational adversaries. Not in the moral sense, nothing about this conduct is reasonable, but in the operational one. A rational adversary has goals, holds a working model of the environment those goals operate in, and chooses actions that advance the goals given that model [18]. The offender wants to reach a child, obtain or distribute material, and avoid being caught. He knows platforms detect known material. He knows investigators are searching. He acts accordingly.

This is visible throughout the enforcement record, and it is the part designing against predation has to deliberate. Offenders do not use platforms naively. They use them the way one who knows he is being hunted and is actively hunting using them. On Instagram and Snapchat, accounts are discarded and recreated to evade blocks and outrun identification. Contact initiated on a public, monitored surface migrates to an encrypted or ephemeral one before the offense escalates, reducing the evidence trail and detection surface for law enforcement. Exploitative communities share tradecraft: which platforms are monitored, which detection methods are in use, grooming scripts that work, and how to avoid being caught. None of these are incidental choices. Offenders model detection and consistently aim to move around defenses and towards their goal.

That is the structure the affordance framework exposes. An offender failing to reach their goals on one platform does not stop. He relocates to a platform whose affordances better serve evasion, trading reach for anonymity, persistence for ephemerality, an open channel for an encrypted one.

This elevates what it means to design against predation, and to ship new technology with novel capabilities.

A defense built for a static offender — one who picks a platform and stays there, who does not anticipate detection, who does not adapt — fails, because the record does not consistently model that offender. It contains one who migrates channels, embraces new technologies, and reaches for the tool that defeats the detection method in use. Designing against predation means designing against adaptation.

7.2 Rational Exploitation and Affordance Trajectories

Section 7.1 established that offenders are rational adversaries who model their environment and optimize against it. The platform affordance is the instrument of offenders. What follows is a mathematical framework mapping offender goals, affordances, and harm vectors.

Let $\mathcal{A} = \{a_1, a_2, \ldots, a_n\}$ denote the set of capabilities a platform makes available: anonymity, ephemerality, contact discovery, unmonitored communication, distribution infrastructure, generative synthesis. An offender with exploitation goal g selects a trajectory through $\mathcal{A}$:

$$L^*_{g,\mathcal{A}} \;=\; \arg \max_{L \in \mathrm{Seq}(\mathcal{A})} \; \mathbb{E}\big[U_g(L)\big] \tag{1}$$

where $L^*_{g,\mathcal{A}} \in \mathrm{Seq}(\mathcal{A})$ is the implied optimal path an offender pursuing goal g traces through the affordance environment, U_g is the utility function for goal g, $\mathrm{Seq}(\mathcal{A})$ denotes the set of all finite ordered sequences over $\mathcal{A}$, and $\mathbb{E}[\cdot]$ is taken over stochastic outcomes (victim compliance, detection risk, platform intervention) that render $U_g(L)$ a random variable [18]. Different goals produce different optimal trajectories. An enticement offender maximizes over sustained-contact affordances — anonymity, trust architecture, unmonitored channels. A sextortion offender maximizes over affordances that make coercion possible — imagery acquisition, distribution infrastructure, encrypted payment channels. A trafficking offender explores a chain of sequential events, from contact to exploitation. $L^*_{g,\mathcal{A}}$ is goal-dependent. $\mathcal{A}$ is shared.

Define $\mathcal{H} = \{h_1, h_2, \ldots, h_k\}$ as the finite set of victim-facing harm types docu-

mented in this corpus: a child groomed toward in-person abuse; a child coerced into producing self-generated imagery; a child's material distributed and recirculated across networks; a child sextorted under threat of exposure; a child trafficked for commercial sexual exploitation; a child's likeness synthesized into abuse imagery without contact. This paper does not measure harm severity or rank outcomes within $\mathcal{H}$. The analytical objective is narrower: map which trajectories through $\mathcal{A}$ produce which elements of $\mathcal{H}$.

Define $\varphi\colon \mathcal{L} \to \mathcal{E}$ as the mapping from offense lifecycle to exploitation type, where $\mathcal{L}$ is the space of possible trajectories through $\mathcal{A}$ and $\mathcal{E} = \{\text{enticement, sextortion, trafficking, CSAM production, } \dots\}$ is the set of exploitation types. A complete trajectory terminates to a single type. Enticement and sextortion are distinct elements of $\mathcal{E}$, reached along distinct trajectories:

$$\varphi\left(L^*_{\text{enticement},\mathcal{A}}\right) \neq \varphi\left(L^*_{\text{sextortion},\mathcal{A}}\right) \quad \text{in general} \tag{2}$$

The trajectories diverge because the goals diverge, and therefore the affordances selected diverge.

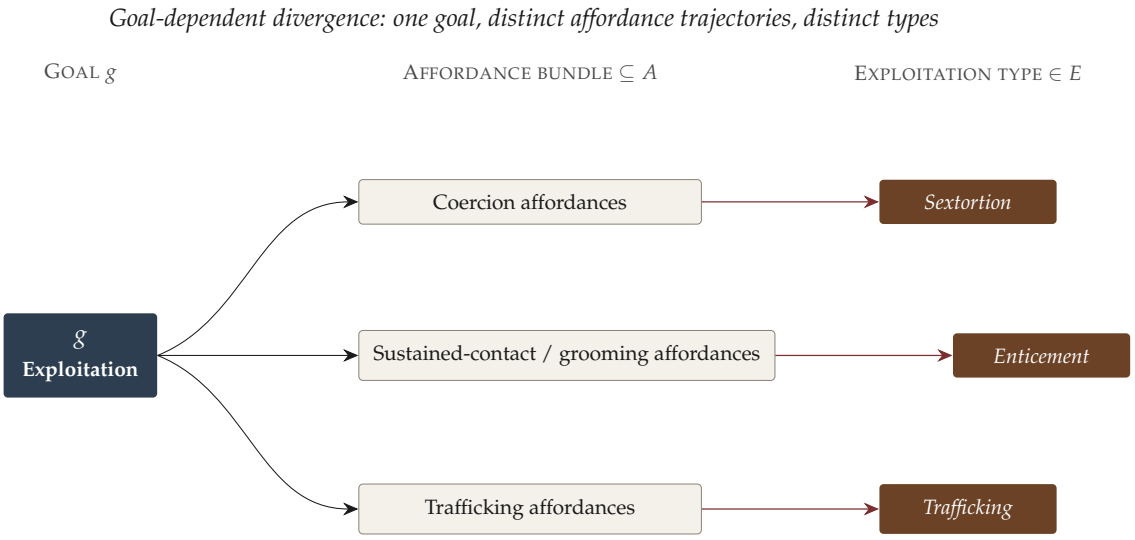

Figure 6: Goal-dependent divergence. A single exploitation goal g resolves through distinct affordance trajectories to distinct exploitation types in $\mathcal{E}$

Each exploitation type is constituted by a set of victim-facing harms. Define $\eta\colon \mathcal{E} \to 2^{\mathcal{H}}$ mapping each exploitation type to the harms it comprises, where $\mathcal{H}$ is the harm set defined above. Sextortion comprises coerced imagery production and exploitation under threat of exposure; enticement comprises grooming toward

in-person abuse. Trafficking comprises a sequence of stages spanning contact, targeting, grooming, material distribution, and commercial sexual exploitation, each with its set of victim-facing harms.

The harm content of a trajectory is the composition $\eta(\varphi(L)) \subseteq \mathcal{H}$: a path resolves to an exploitation type, and an exploitation type resolves to the harms that constitute it.

Beneath the trajectory, each affordance carries its own harm associations. Define $\psi \colon \mathcal{A} \to 2^{\mathcal{H}}$ mapping each affordance class $a_n \in \mathcal{A}$ to the set of harms across all trajectories in which a_n appears, recoverable from the enforcement record by observing which harm types co-occur with a_i across the corpus. Where φ resolves a trajectory to its exploitation type and η unpacks that type into harms, ψ records the harms each individual affordance participates in. Table 2 reports ψ.

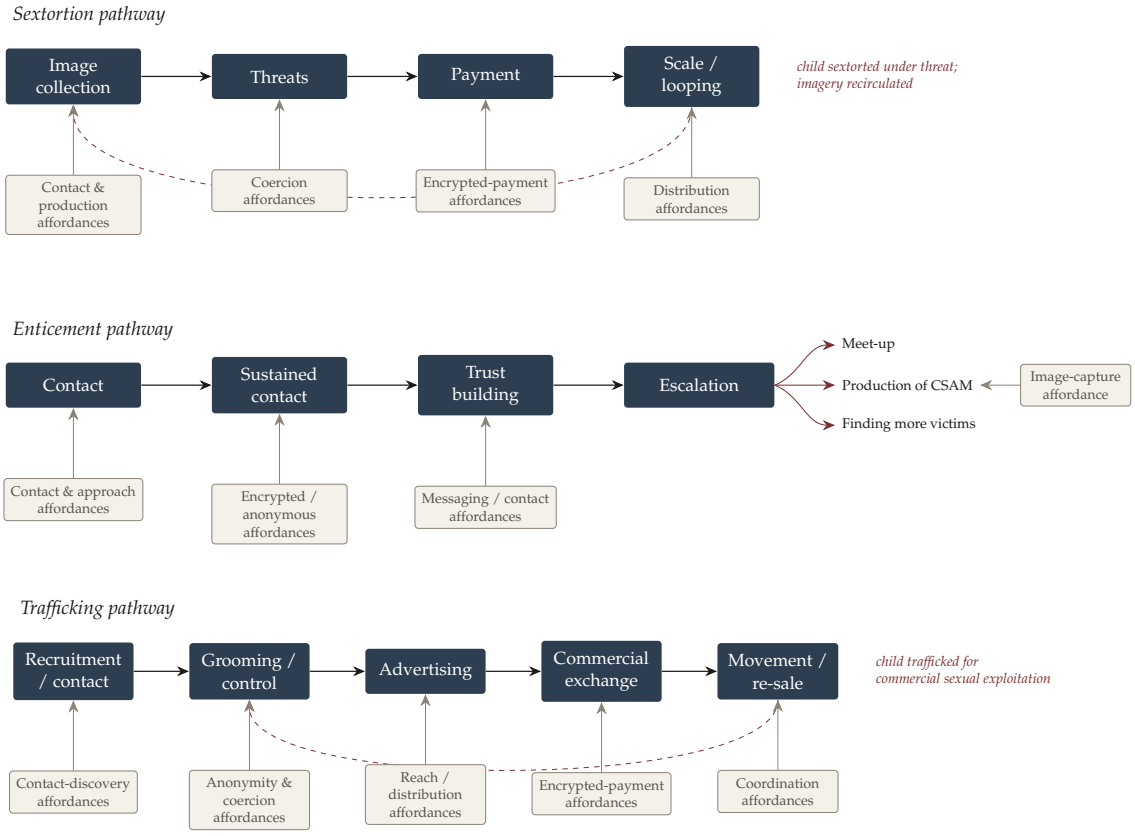

Figure 7: Affordance trajectories for three exploitation types. Each pathway traces a sequence of offense stages with the affordance class enabling each stage feeding upward.

New platforms expand $\mathcal{A}$. For a fixed goal g, the optimal trajectory shifts and the

set of affordances available to optimize over expands. The efficiency of exploitation changes. The exploitation type does not. Every path still resolves to an exploitation type in $\mathcal{E}$, and every type to a subset of $\mathcal{H}$. $\mathcal{H}$ is closed.[6]

The design implication is equally direct. For any proposed affordance $a_{new} \notin \mathcal{A}$, define its exploitation utility as the marginal change in an offender's optimal expected payoff under goal g when a_{new} enters the affordance environment:

$$u(a_{new}, g) \;=\; \mathbb{E}\Big[U_g\Big(L^*_{g,\,\mathcal{A}\cup\{a_{new}\}}\Big)\Big] \;-\; \mathbb{E}\Big[U_g\Big(L^*_{g,\,\mathcal{A}}\Big)\Big] \tag{3}$$

where $L^*_{g,\mathcal{A}}$ is the implied optimal path for goal g over the current affordance set and $L^*_{g,\mathcal{A}\cup\{a_{new}\}}$ the implied optimal path when a_{new} enters the affordance environment.

Table 2: Affordance-to-harm mapping. For each affordance class $a_i \in \mathcal{A}$, $\psi(a_i) \subseteq \mathcal{H}$ is the set of victim-facing harms across all trajectories in which a_i appears in the CaseLinker corpus.

Affordance class a_i	**Harms** $\psi(a_i) \subseteq \mathcal{H}$
Anonymity	h_1 groomed toward in-person abuse; h_2 coerced into self-generated imagery; h_4 sextorted under threat of exposure
Ephemerality	h_2 coerced into self-generated imagery; h_4 sextorted under threat of exposure
Contact discovery	h_1 groomed toward in-person abuse
Unmonitored communication	h_1 groomed toward in-person abuse; h_2 coerced into self-generated imagery; h_4 sextorted under threat of exposure
Distribution infrastructure	h_3 material distributed and recirculated across networks
Generative synthesis	h_6 likeness synthesized into abuse imagery without contact

An affordance with high $u(a_{new}, g)$ for any exploitation goal g is a structural

[6] See Appendix X for the theoretical basis of this claim.

risk before it is a documented harm. Ephemerality carries high u for sextortion. Unverified account creation carries high u across nearly all g. Generative synthesis carries high u for CSAM production and persona fabrication at scale. These values are estimable before a feature ships. The enforcement record provides a systemic record to model harm.

Platforms that do not model $u(a_{new}, g)$ prior to deployment leave that quantity unmeasured and therefore unmitigated. The underlying corpus assembled in this paper provides an instrument for that measurement. It documents which affordances rational agents with exploitation goals consistently reach for, and what happens to children when they do.

The next platform will introduce affordances that do not yet exist. Some will carry high u. Twenty-four years of enforcement records document what $\varphi(L^*_{g,A})$ looks like when $u(a_{new}, g)$ goes systemically unmeasured before deployment.

7.3 The Foreseeable Harm Standard

The enforcement record is not a surprise. It is a retroactive document comprised of foreseeable harm.

Every affordance class that appears consistently across this corpus — anonymity, ephemerality, unmonitored communication, contact discovery, distribution infrastructure — had a documented exploitation pattern before most of the platforms currently using it existed. Anonymous contact without verified identity on Kik and Telegram in the current enforcement record were documented misuse surfaces on IRC before either platform existed. The distribution infrastructure that defines Dropbox and Google Drive in this corpus appeared first in Tor and early peer-to-peer networks. The capability classes are not new. The exploitation mechanics they produce are not new. The cases are new. The children in them are new. The platforms are new. The harm signature is not.

This matters because of what foreseeability means. In law, a harm is foreseeable if a reasonable person could have anticipated it in advance. The standard does not require certainty. It does not require a specific victim, a specific offender, or a specific platform. It requires that the risk was knowable — that someone paying attention could have seen it coming. Thirty platforms and twenty-four years of enforcement records establish, at scale, that the risk was always knowable. The affordance classes in this corpus carried documented exploitation histories before the product decisions that introduced them to new platforms were made. The harm was foreseeable because the pattern already existed.

The industry has operated on a different assumption. The implicit posture — expressed in policy statements, legal filings, and public communications — has been that misuse is unpredictable: that no reasonable platform team could have anticipated that their features would be turned against children. This corpus does not support that assumption. It documents the same five capability types appearing in the enforcement record across every platform generation, implemented differently each time and producing the same offense mechanics each time. That is not an unpredictable pattern. It is a stable one. Stable patterns are foreseeable. Foreseeable harms that go unmitigated are choices, not accidents.

This paper does not argue platform liability. That is a legal question with its own framework, its own evidentiary standards, and its own venue. What this paper argues is narrower and fundamental: if harm is estimable before a feature ships — and Section 7.2 establishes that it is, using twenty-four years of enforcement records as the evidence base — then a platform that ships a high affordance-misuse-harm pathway without mitigation has made a measurable choice. The framework makes that choice visible. It does not assign guilt. It removes the defense of ignorance.

What foreseeable harm review looks like in practice is not complicated. Before a feature ships, ask: does this affordance class appear in the enforcement record? What exploitation pattern does it produce? What is misuse risk for the most plausible exploitation goals? If risk is high, what mitigation exists at the design level — not in the terms of service, not in the content moderation queue, but in the feature itself? These are engineering questions. They are not currently standardized or required. They could be.[7]

Every case in this corpus was foreseeable. Not the specific child or offender. The harm type, the affordance class, the exploitation mechanic — those were foreseeable, because the pattern was in the record before the platform was in the market. The next feature will introduce capabilities that do not yet exist in the enforcement record. Some will carry high risk. The corpus provides the instrument to find them and design against exploitation before an offender does.

7.4 Limitations and Scope

8. Exploitation Is Not Inevitable

[7] See Appendix Z for a walkthrough example on a novel technology

Victimization begins with a decision to exploit a child. That decision is not a variable platform design controls. It has existed across history, on every platform, with every affordance configuration. The offender's goal is stable.

What changes with each platform generation is the affordances and capabilities available to each person who has already made that decision. Platforms choose which affordances to ship. Anonymity is a design choice. Ephemerality is a design choice. Distribution infrastructure is a design choice. Each carries exploitation utility — a measurable contribution to an offender's ability to reach their goal — whether or not it is measured.

For twenty-four years, the enforcement model has been retrospective measurement and reactive attribution. A feature ships. Offenders reach for it. Cases accumulate. The pattern becomes visible in enforcement records reviewed years later by researchers, not by the teams or platforms who built the feature.

This does not have to be the sequence.

The framework developed in this paper provides the vocabulary to run that measurement before a feature ships rather than after a child is harmed. For each proposed affordance, ask what a rational agent with an exploitation goal does with it. The answer is not speculative. Thirty platforms, twenty-four years, and 7,426 cases document what that answer looks like across every affordance class.

Platforms can be innovative. They can scale. They can ship features that hundreds of millions of legitimate users depend on without shipping features whose exploitation utility is unknown. The mechanism is not esoteric: model the adversarial agent before deployment. Treat high-risk affordances as structural risks requiring mitigation, not edge cases requiring legal review after the fact.

Child exploitation is a problem neither produced nor intended by the platforms through which it spreads. It is a problem that has followed every technological generation because the question of what offenders do with a new feature has not yet become a standard part of engineering.

The next platform can ship its most ambitious feature without generating another case in this corpus.

9. Data, Code, and Public Research Artifacts

All data, code, and research artifacts underlying this analysis are publicly available. The corpus, extraction pipeline, and knowledge graph infrastructure are released

under the MIT License.

Code and Pipeline

The full CaseLinker pipeline — extraction rules, analysis tools, and reproducibility documentation — is available at https://github.com/mrinaalr/CaseLinker.

Live Platform

The CaseLinker interface, including the platform harm dashboard and case search tools, is accessible at https://caselinker.up.railway.app.

Related Work

This paper is part of the CaseLinker Technical Report Series. Prior reports (#1–5) are available on GitHub and the initial preprint is available at https://arxiv.org/abs/2603.18020. Additional research and project documentation is maintained at https://end-child-exploitation.com.

API and MCP Access

REST API and MCP access is available for bulk data export and programmatic corpus analysis. Access is rate-limited by default for financial and data stewardship reasons. Researchers, practitioners, and institutional collaborators seeking full access may contact the author directly — a free key will be provided upon request.

Human Research Determination

This research operates under HRPO NHSR Determination #7668. The corpus contains no private or identifiable information, investigative files, or victim-identifying data under federal regulations [45 CFR 46.102(f)(1), (2)]. All case records are publicly available enforcement documents.

Contact

Mrinaal Ramachandran
mramachandra@umass.edu
University of Massachusetts Amherst

Appendix X: On the Closure of $\mathcal{H}$

This paper treats $\mathcal{H}$ as a closed finite set on theoretical, not legal, grounds. The closure of $\mathcal{H}$ is not asserted on its own. It follows from the closure of the goal set beneath it.

An offender acts on a goal: reach a child, obtain or produce material, coerce, distribute, sell. Call the set of these goals $\mathcal{G}$. The claim is that $\mathcal{G}$ is finite and historically stable. The goals in the enforcement record are the same goals that predate the internet, because they are defined by what one person can want to do to a child, not by the technology used to do it. A new affordance does not add a goal to $\mathcal{G}$. It changes which trajectory realizes a goal, and how cheaply.

The rest follows by composition. The exploitation type $\mathcal{E}$ is the image of $\mathcal{G}$ under the offender's choice of trajectory: every type in $\mathcal{E}$ is the realization of some goal $g \in \mathcal{G}$. The harm set $\mathcal{H}$ is the image of $\mathcal{E}$ under η: every harm in $\mathcal{H}$ is constitutive of some type in $\mathcal{E}$. If $\mathcal{G}$ is closed, $\mathcal{E}$ inherits its closure, and $\mathcal{H}$ inherits it in turn. Technology acts on the trajectories, never on the sets. It changes how often a goal is reached and how cheaply, never which goals exist.

Two cases make this concrete, and both are the kind a reader might first read as counterexamples.

Sextortion emerged as a named typology only in the last decade, which can look like a new element of $\mathcal{H}$. It is not. The goal, coerce a child through the threat of exposure, is old. What changed is cost. A camera in every pocket, disappearing messages, and pseudonymous contact collapsed the price of acquiring the leverage and delivering the threat. The affordances made the goal cheap and therefore frequent. They did not invent it. Sextortion became more prevalent. It did not become new.

Generative synthesis is the harder case, because it produces abuse imagery without physical contact, which has no prior analogue in mechanism. But the harm it activates, a child's likeness used in abuse imagery, predates the mechanism by decades. The affordance is new. The element of $\mathcal{H}$ it activates is not. Synthesis is a cheaper, contactless path to a harm the record already contained.

In both cases the pattern is the same. The affordance is new. The trajectory is new. The efficiency is new. The goal, the type, and the harm are not. This is the empirical content of the closure claim, and the corpus is its basis: twenty-four years of enforcement records across thirty platforms, and no case producing a harm not representable as an element of $\mathcal{H}$ as defined in Section 7.2.

A note on the status of the claim. Closure of $\mathcal{G}$ is an empirical and historical observation, not a theorem. No proof rules out a future technology that creates a goal irreducible to the existing set. What the record establishes is that across twenty-four years and every platform generation in the corpus, none has. The categories of harm are bounded by the finite ways an offense can be committed against a child's person, and the corpus has not yet produced an exception. The claim is falsifiable by a single case that does. None appears.

Appendix Y: Establishing Theorems, Abstracting Exploitation, and Disproving Claims

The framework in Section 7.2 is offered as an instrument, not a universal law. Its claims are formal and therefore falsifiable, and the author explicitly invites researchers, particularly behavioral criminologists and offender-behavior specialists, to use the apparatus to test, extend, or break the claims it makes.

The author would also like to note that the modeling is limited by two bounds:

These are theorems over the observable universe, not laws over the full one. The corpus is drawn from *successful* enforcement: exploitation that was detected, charged, and made public. Trajectories that evade detection entirely are absent by construction, and some affordances may be valuable to offenders precisely because they leave no enforcement trace. The closure claims therefore hold over the space of *detected* exploitation. Whether they hold over the full space is an open empirical question the public record cannot settle, because the unobserved portion is, by definition, unobserved. A harm type that exists today only in undetected offenses would not appear in $\mathcal{H}$, and the framework would not show it. This is the sharpest limitation of the work and the most direct route to extending it: a method that surfaces undetected trajectories would test closure where this corpus cannot.

Abstraction discards detail that matters. Reducing an offense to a trajectory through an affordance set abstracts essential details that matter. It does not model the child's experience, the specific dynamics of a given case, the offender's psychology beyond the thin rationality of goal-directed action, or the social and developmental context in which an offense occurs. The framework is deliberately narrow: it maps capability to misuse to harm, and nothing else. It is an engineering instrument for pre-deployment reasoning, not a theory of why offenders offend or

of what victims endure. Used as the latter, it overreaches. The author notes this boundary so the abstraction is not mistaken for the territory.

How to disprove this work. The claims are stated formally in Section 7.2, and each is structured to be broken by evidence rather than by argument. The author encourages their direct test. A single documented case is sufficient to falsify any of the closure claims: an offense whose goal does not reduce to the goal set, a harm not representable in $\mathcal{H}$, or an affordance whose introduction produces a genuinely new exploitation type rather than a cheaper path to an existing one. The type-invariance and utility claims are testable against the corpus itself, which is public (Section 9): a trajectory that resolves to a type outside $\mathcal{E}$, or a demonstration that observed misuse is uncorrelated with the affordance properties the framework scores, breaks them. None of this requires agreement with the author's framing. It requires a case, or a correlation, that the model cannot hold. A criminologist who produces one using this machinery advances the field as much as a confirmation would, and the author regards that outcome as the framework working as intended.

Appendix Z: A Worked Pre-Deployment Review

The framework in Section 7.2 is meant to be run before a feature ships, not after it generates cases. This appendix demonstrates that on a hypothetical but realistic product. Nothing here is a real platform or feature. The point is to show what foreseeable-harm review looks like when the math is applied to a capability that does not yet exist in the enforcement record.

Z.1 The Product (as the team would write it)

Team Banana is building a consumer VR capture tool, internally **Project Memory**. The pitch, as it would appear in a product brief:

> Memory lets users capture moments as volumetric, navigable scenes rather than flat photos. Point the device at a birthday, a first step, a reunion, and Memory reconstructs the moment in three dimensions. Users can step back into the captured scene in VR, and share it with friends and family who can experience the moment as if they were there. Memories persist in the user's library and can be revisited indefinitely.

The engineering decomposition the team would draw up:

- **Volumetric capture.** Multi-sensor reconstruction of a physical scene, including any people in it, into a navigable 3D asset.
- **Real-likeness representation.** The captured asset is a photoreal reconstruction of real people, viewable from angles the original capture did not directly frame.
- **Spatial / experiential sharing.** Scenes are shareable as immersive experiences another user enters, not as files they receive.
- **Artifact persistence.** Captured scenes are stored indefinitely in the user library and re-enterable at any time.

Read as a product, this is benign and obviously valuable. That is precisely the condition under which the framework earns its place: the capability does not look like a weapon. The review below asks what a rational agent with an exploitation goal does with it anyway.

Z.2 Mapping the affordances to the existing sets

Each capability above is an affordance $a \in \mathcal{A}$. The first question is whether any is genuinely new or whether it reinforces an affordance already in the record.

- Volumetric capture is a new *form* of an existing affordance, image-capture / production, extended from 2D to 3D. It does not create a new harm category; it deepens an existing one.
- Real-likeness representation overlaps the generative-synthesis affordance already in $\mathcal{A}$. A photoreal 3D reconstruction viewable from un-captured angles is, functionally, likeness synthesis: the asset depicts the subject in configurations the camera did not record.
- Spatial sharing is a new form of distribution infrastructure: it moves an experiential asset rather than a file, but its function in the record, sharing of captured material to other parties, is the existing distribution affordance.
- Artifact persistence maps to durable storage already in the record: indefinite, re-enterable retention of captured material is functionally the cloud-storage and possession affordance the corpus documents on file-hosting platforms; the asset is held, distributable, and re-accessible rather than transient.

This is the closure claim doing work in advance. A genuinely novel technology decomposes into affordances that resolve to the existing $\mathcal{A}$, and therefore to the existing harm set $\mathcal{H}$. Project Memory introduces no harm type not already in $\mathcal{H}$. It

introduces new *trajectories* to harms the record already contains.

Z.3 Comparative exploitation utility

For each affordance and each plausible goal $g \in \{\text{enticement, sextortion, production}\}$, the review estimates $u(a, g)$, the marginal contribution of the affordance to an offender's optimal trajectory under that goal (Eq. 3). The values are comparative, not fitted: the framework ranks affordances by the leverage they add, it does not assign a cardinal payoff. What follows is the structure u recovers.

Affordance	Enticement	Sextortion	Production
Volumetric capture	low	moderate	**high**
Real-likeness representation	low	**high**	**high**
Spatial / experiential sharing	low	moderate	**high**
Artifact persistence	low	**high**	moderate

The structure, read off the table:

Production carries high u across three of four affordances. Volumetric capture, likeness representation, and spatial sharing each materially extend an offender's ability to produce and circulate abuse material. The combination is the concern: a tool that captures a real child volumetrically, renders that likeness from un-captured angles, and distributes the result as an immersive asset is, in the production trajectory, a more efficient path to harms h_3 (distribution) and h_6 (likeness synthesized into abuse imagery without contact) than any affordance currently in the record. u is high here because the affordances compose: each is moderate alone and high together.

Enticement is uniformly low. None of these affordances materially improves contact, trust-building, or sustained-channel access, which is what the enticement trajectory optimizes over. The tool is not a contact surface. This is as important a finding as the high values: it tells the team where *not* to spend mitigation effort.

Real-likeness representation carries high u for sextortion. A detailed, manipulable likeness is more coercive material than a flat image, it raises the credibility and threat-value of what the offender can hold over the victim.

Artifact persistence carries *high* u for sextortion. Sextortion is sustained coercion: the threat is only as durable as the material behind it. A persistent, re-enterable asset keeps the leverage live indefinitely, the offender can re-threaten, escalate, and re-extort against the same captured scene over time, rather than holding a

one-time threat that degrades. Persistence is what converts a single capture into renewable coercive leverage.

Z.4 Mitigations during review

The mitigations are not brainstormed against the product in general. Each addresses a specific high-u affordance, at the design level, in the feature itself, before code.

- **Against likeness representation (high u, sextortion and production).** Constrain reconstruction to captured viewpoints. If the asset cannot be rotated to angles the camera did not frame, the affordance stops functioning as likeness synthesis and reverts to ordinary capture. This is a capture-geometry constraint, decided at the reconstruction-engine level, and it removes the single highest-u affordance from the synthesis trajectory.
- **Against the capture-render-share composition (high u, production).** Because u is high only when the three affordances compose, break the composition. Gate volumetric capture of identifiable minors, or require on-device-only processing for scenes containing them, so the captured asset never becomes a shareable server-side object. The mitigation targets the link between capture and distribution, which is where the composed utility lives.
- **Do not over-invest against enticement.** u is uniformly low across the contact trajectory. The tool does not need a contact-layer mitigation it was never going to require. Mitigation effort concentrates on production and synthesis, where the record says the leverage is.

Z.5 What the review produced

Before a line of capture code was written, the review identified that Project Memory's exploitation risk concentrates in the production/synthesis trajectory, not the contact one; that the risk is compositional, living in the capture-render-share chain rather than any single affordance; that the highest-leverage mitigation is a capture-geometry constraint at the reconstruction engine; and that the mitigation considerations can be documented and implemented. None of these conclusions required a shipped product, a reported case, or a legal threshold. They required the affordance decomposition, the existing harm set, and a comparative estimate of u. That is the sequence the paper argues for, run once, on one feature, before deployment.

References

[1] Danah Boyd. *It's Complicated: The Social Lives of Networked Teens*. Yale University Press, New Haven, CT, 2014.

[2] Taina Bucher and Anne Helmond. The affordances of social media platforms. In Jean Burgess, Alice Marwick, and Thomas Poell, editors, *The SAGE Handbook of Social Media*, pages 233–253. SAGE Publications, London, 2018. doi:10.4135/9781473984066.n14.

[3] Elie Bursztein, Einat Clarke, Michelle DeLaune, David M. Elifff, Nick Hsu, Lindsey Olson, John Shehan, Marina Thiry, Kurt Thomas, and Travis Bright. Rethinking the detection of child sexual abuse imagery on the Internet. In *Proceedings of the World Wide Web Conference (WWW 2019)*, pages 2601–2607, New York, NY, 2019. ACM. doi:10.1145/3308558.3313482.

[4] Danielle Keats Citron and Mary Anne Franks. Criminalizing revenge porn. *Wake Forest Law Review*, 49:345–391, 2014.

[5] James J. Gibson. *The Ecological Approach to Visual Perception*. Houghton Mifflin, Boston, MA, 1979.

[6] Tarleton Gillespie. The politics of "platforms". *New Media & Society*, 12(3): 347–364, 2010. doi:10.1177/1461444809342738.

[7] Tarleton Gillespie. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. Yale University Press, New Haven, CT, 2018. doi:10.12987/9780300235029.

[8] Shelby Grossman, Riana Pfefferkorn, David Thiel, Sara Shah, Renée DiResta, John Perrino, Elena Cryst, Alex Stamos, and Jeff Hancock. The strengths and weaknesses of the online child safety ecosystem. Technical report, Stanford Internet Observatory, Stanford Digital Repository, 2024. URL `https://purl.stanford.edu/pr592kc5483`.

[9] Ian Hutchby. Technologies, texts and affordances. *Sociology*, 35(2):441–456, 2001. doi:10.1177/S0038038501000219.

[10] Juliane A. Kloess, Anthony R. Beech, and Leigh Harkins. Online child sexual exploitation: Prevalence, process, and offender characteristics. *Trauma, Violence, & Abuse*, 15(2):126–139, 2014. doi:10.1177/1524838013511543.

[11] Juliane A. Kloess, Catherine E. Hamilton-Giachritsis, and Anthony R. Beech. Offense processes of online sexual grooming and abuse of children via Internet communication platforms. *Sexual Abuse: A Journal of Research and Treatment*, 31(1):73–96, 2019. doi:10.1177/1079063217720927.

[12] Sonia Livingstone and Peter K. Smith. Annual research review: Harms experienced by child users of online and mobile technologies—the nature, prevalence and management of sexual and aggressive risks in the digital age. *Journal of Child Psychology and Psychiatry*, 55(6):635–654, 2014. doi:10.1111/jcpp.12197.

[13] Kimberly J. Mitchell, Lisa M. Jones, David Finkelhor, and Janis Wolak. Internet-facilitated commercial sexual exploitation of children: Findings from a nationally representative sample of law enforcement agencies in the United States. *Sexual Abuse: A Journal of Research and Treatment*, 23(1):43–71, 2011. doi:10.1177/1079063210374347.

[14] National Center for Missing & Exploited Children. CyberTipline 2023 report. Technical report, National Center for Missing & Exploited Children, Alexandria, VA, 2023. URL https://www.missingkids.org/gethelpnow/cybertipline.

[15] National Center for Missing and Exploited Children. Testimony before the house energy and commerce committee, subcommittee on innovation, data, and commerce, March 2025. URL https://www.congress.gov/119/meeting/house/118066/witnesses/HHRG-119-IF17-Wstate-SourasY-20250326.pdf. Hearing on child online safety legislation.

[16] Donald A. Norman. *The Design of Everyday Things*. Basic Books, New York, NY, revised and expanded edition, 2013.

[17] Ethel Quayle. Affordances, social media and the criminogenic nature of the internet: Technology-mediated child sexual abuse. In Ernesto Caffo, editor, *Online Child Sexual Exploitation*, pages 33–48. Springer, Cham, 2021. ISBN 978-3-030-66654-5. doi:10.1007/978-3-030-66654-5_4.

[18] Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Pearson, Hoboken, NJ, 4th edition, 2021.

[19] Martin Steinebach. An analysis of PhotoDNA. In *Proceedings of the 18th International Conference on Availability, Reliability and Security (ARES 2023)*, page 8 pages, Benevento, Italy, August 2023. ACM. doi:10.1145/3600160.3605048. URL https://dl.acm.org/doi/fullHtml/10.1145/3600160.3605048.

[20] Thorn. Child sexual abuse material (CSAM), 2024. URL https://www.thorn.org/research/child-sexual-abuse-material-csam/. Statistics sourced from NCMEC 2023 CyberTipline data.

[21] Bryce G. Westlake, Martin Bouchard, and Richard Frank. Finding the key players in online child exploitation networks. *Policy & Internet*, 9(1):104–125, 2012. doi:10.2202/1944-2866.1126.

[22] Janis Wolak and David Finkelhor. Sextortion: Findings from a survey of 1,631 victims. Technical report, Crimes Against Children Research Center, University of New Hampshire, Durham, NH, 2016. URL https://www.unh.edu/ccrc/sites/default/files/media/2022-02/key-findings-from-a-survey-of-sextortion-victims-revised-8-9-2016.pdf.

[23] Janis Wolak, David Finkelhor, and Kimberly J. Mitchell. Internet-initiated sex crimes against minors: Implications for prevention based on findings from a national study. *Journal of Adolescent Health*, 35(5):424.e11–424.e20, 2004. doi:10.1016/j.jadohealth.2004.05.006.

[24] Janis Wolak, David Finkelhor, and Kimberly J. Mitchell. Online "predators" and their victims: Myths, realities, and implications for prevention and treatment. *American Psychologist*, 63(2):111–128, 2008. doi:10.1037/0003-066X.63.2.111.

[25] Janis Wolak, David Finkelhor, Wendy Walsh, and Leah Treitman. Sextortion of minors: Characteristics and dynamics. *Journal of Adolescent Health*, 62(1):72–79, 2018. doi:10.1016/j.jadohealth.2017.08.014.

[26] Mark A. Wood, Matthew Mitchell, Flynn Pervan, Briony Anderson, Tully O'Neill, Jackson Wood, and Will Arpke-Wales. Inviting, affording and translating harm: Understanding the role of technological mediation in technology-facilitated violence. *The British Journal of Criminology*, 63(6):1384–1404, 2023. doi:10.1093/bjc/azac095.