

**PARENTAL RESPONSIBILITY FOR A MINOR'S AI-GENERATED
DEFAMATORY CONTENT: TESTING THE "KNEW OR OUGHT TO HAVE
KNOWN" STANDARD UNDER INDIAN LAW**

by

Yashash KS

ABSTRACT

This paper examines whether existing Indian tort and defamation principles can adequately determine parental responsibility when a minor uses generative artificial intelligence to create or disseminate defamatory content, particularly where the parent had no realistic opportunity to detect the conduct. Indian law does not make parents automatically liable for the torts of their children; responsibility depends on an independent showing of parental fault, ordinarily negligent supervision, judged by whether the parent knew or ought reasonably to have known of the risk. The paper tests this standard against six constructed factual scenarios and argues that generative artificial intelligence does not defeat the standard as such, but sharply increases the incidence of genuinely undetectable harm, a condition the paper terms digital invisibility. Where such invisibility is real, the paper argues that responsibility is more defensibly located in the due diligence obligations recently imposed on intermediaries handling synthetically generated information, and in an as yet undeveloped Indian conception of generative AI developer responsibility, than in any expansion of parental fault. Drawing on a comparative examination of France, Mauritius, and the United Kingdom, the paper proposes a judicially administrable "reasonable digital supervision" standard for Indian courts, evaluated against four alternative reform options.

I. INTRODUCTION

A fourteen-year-old opens a generative artificial intelligence application on the family phone, types a short prompt describing a classmate's teacher, and within seconds produces a written passage, or an image, that falsely attributes dishonest or immoral conduct to that teacher. The content is posted to a closed group chat before the parent, asleep in the next room, has any occasion to notice. By morning the material has been forwarded beyond the group, the teacher's reputation has suffered concrete harm, and the parent is asked, in effect, to answer for a wrong committed with a tool the parent did not choose, using a device the parent may have supplied for homework, in a manner the parent had no realistic opportunity to observe.

This scenario is not exceptional. It is an ordinary operating condition of generative artificial intelligence on consumer devices, and it exposes a structural problem in Indian civil liability law that predates the technology but is sharpened by it. Indian tort doctrine, inherited substantially from English common law, does not impose liability on a parent merely because a child has committed a tort.² Responsibility attaches to the parent only where an independent basis of fault, ordinarily negligent supervision, can be shown: the parent knew or ought reasonably to have known of a risk and failed to act on that knowledge. This standard was built for a world in which wrongdoing by children left some trace, a phone call, a bruised knuckle, a broken window, that a reasonably attentive parent could be expected to notice or anticipate. Generative artificial intelligence removes the trace. A defamatory statement or a synthetic image can be conceived, generated, and published within a single unremarkable-looking session on a shared device, without any of the behavioural or physical warning signs on which the "knew or ought to have known" standard has always silently relied.

This paper asks a narrow but under-examined question: when, if at all, should a parent bear civil responsibility for defamatory content that a minor creates or disseminates using generative artificial intelligence, and how should Indian law determine parental responsibility where the parent had limited or no reasonable opportunity to detect, prevent, or control the conduct? The paper takes no starting position that parents should or should not be liable. It instead tests the existing doctrinal apparatus, personal negligence, negligent supervision, the liability of the minor

²*Moon v. Towers* (1860) 8 C.B. (N.S.) 611, holding that "a father is not, as such, liable for the torts of his infant child," a rule received into Indian jurisprudence through the general reception of English common law and still cited in Commonwealth jurisdictions.

personally, and the residual responsibility of platforms and of the developers of the underlying tools, against a set of constructed factual scenarios, and asks where the apparatus succeeds, where it strains, and where a different allocation of responsibility might be more defensible.

The central claim developed here is a qualified one. Indian negligence principles are not, in their abstract formulation, incapable of addressing AI-generated defamatory content produced by minors: the "knew or ought to have known" inquiry is flexible enough in theory to accommodate new fact patterns. What is genuinely strained is the evidentiary and normative content that Indian courts and commentators have, until now, implicitly given to that standard, content that assumes observability. The paper argues that Indian law needs not a new cause of action but a more precise, technologically literate account of what a reasonable parent can be expected to know and do in a household where a smartphone is simultaneously a homework tool and an unsupervised content-generation studio. It further argues that, in a meaningful subset of cases, the most defensible locus of responsibility is not the parent at all but the intermediary that hosted the publication or the developer of the generative tool, and that Indian regulatory attention to synthetically generated information under the Information Technology Rules is more directly responsive to this problem than any plausible reform of parental tort liability.

The paper proceeds as follows. Part II defines key terms. Part III examines the liability of minors and the structure of Indian defamation law. Part IV subjects the negligent-supervision standard to six constructed scenarios. Part V examines why generative artificial intelligence specifically destabilises the visibility premise underlying that standard. Part VI maps the four-way allocation of responsibility among minor, parent, platform, and AI developer. Part VII undertakes a comparative analysis of France, Mauritius, and the United Kingdom. Part VIII considers the constitutional and child-rights dimensions of any reform. Part IX proposes a framework and evaluates it against four alternatives. Part X concludes.

The paper's method is doctrinal, comparative, and scenario-based. It does not undertake empirical research, and it does not claim that any hypothetical fact pattern used to test the law describes an actual reported incident; each is expressly labelled as a hypothetical. Every legal proposition relied upon below has been checked against a primary source, a statute, rule, or reported judgment, or against an independently verified secondary source; propositions that could not be verified this way have been left out rather than asserted on the strength of an unverified source.

II. CONCEPTUAL AND LEGAL FOUNDATIONS

Precision in terminology matters because the central argument of this paper is that a single phrase, "AI-generated defamatory content," conceals several analytically distinct situations. A *minor* is, for the purposes of Indian civil law, a person who has not completed the age of eighteen years.³ A *parent* is used here to mean the person or persons who exercise custodial and supervisory responsibility over the minor, which in most households is a biological or adoptive parent but may, on the same reasoning, extend to a guardian in fact.

Parental responsibility is used in this paper as the broad, non-technical idea that a parent owes some duty in connection with the conduct of a child in the parent's custody. *Parental liability*, by contrast, denotes the narrower legal conclusion that a parent is answerable in damages for harm the child has caused. The paper is careful throughout not to treat the two as synonymous. *Negligent supervision* refers to the specific tort theory under which a parent is held liable not for the child's wrong as such, but for the parent's own failure to exercise reasonable care in supervising or controlling the child, where that failure is a proximate cause of the harm.⁴

Defamation is the publication, to at least one person other than the claimant, of a statement that lowers the claimant's reputation in the estimation of right-thinking members of society, or that exposes the claimant to hatred, contempt, or ridicule.⁵ *Defamatory content* is used here as a general label for any statement, image, audio, or video that would satisfy this test if published. *AI-generated content* means content produced, in whole or substantial part, through the operation of a generative artificial intelligence system in response to a human prompt. The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026, now supply a statutory anchor for a closely related term, "synthetically generated information," defined as information that is artificially or algorithmically created, generated, modified, or altered

³Indian Majority Act, 1875, s. 3.

⁴Restatement (Second) of Torts § 316 (Am. Law Inst. 1965) (a parent is under a duty to exercise reasonable care to control a minor child so as to prevent harm to others where the parent knows or has reason to know of the ability and the necessity or opportunity to exercise such control), a formulation consistent with the negligent-supervision principle applied in Commonwealth jurisdictions, including India.

⁵This is the general common law formulation of the tort, received into Indian civil jurisprudence; the criminal offence is separately and more narrowly defined in the Bharatiya Nyaya Sanhita, 2023, s. 356 (which replaced the Indian Penal Code, 1860, s. 499).

using a computer resource in a manner that appears reasonably authentic or true.⁶ *Synthetic media* and *deepfake* are narrower still: synthetic media denotes any media artefact wholly or partly generated by algorithmic means, while a deepfake, strictly speaking, denotes synthetic audio-visual content that impersonates a specific real person with a high degree of visual or auditory realism.

This paper deliberately resists the assumption that "AI-generated defamatory content" is simply another name for a deepfake. At least four analytically distinct categories deserve separate treatment. First, AI-generated defamatory *text*, such as a chatbot-produced paragraph falsely alleging criminal conduct, raises no issue of visual impersonation and is, in substance, no different in kind from a handwritten defamatory letter, only faster and more fluent. Second, AI-generated defamatory *images* that do not depict a real identifiable person raise different foreseeability and identification questions than genuine deepfakes. Third, true audio-visual *deepfakes* impersonating an identifiable real person raise the sharpest publication and attribution problems, since the realism of the medium amplifies the reputational harm. Fourth, *AI-assisted* content, where the minor supplies substantial original defamatory material and uses the tool only to refine or illustrate it, raises different causation questions than content substantially machine-generated from a bare prompt. These distinctions matter because foreseeability, ease of production, and traceability of the minor's contribution vary sharply across categories, and a framework treating them identically will fit some poorly.

An *intermediary*, finally, is defined under Indian law as any person who, on behalf of another, receives, stores, transmits, or provides a service with respect to an electronic record, including network service providers, search engines, and social media platforms.⁷ This definition, and its conditional safe harbour, is examined in Part VI.

⁶Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026,

G.S.R. 120(E) (10 February 2026, in force 20 February 2026), inserting r. 2(1)(wa) into the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021. This is described as the first legislative definition of its kind in Indian law and is comparable in structure to the definition of AI-generated content in the European Union's AI Act.

⁷Information Technology Act, 2000, s. 2(1)(w).

III. LIABILITY OF MINORS AND THE LAW OF DEFAMATION IN INDIA

Two propositions frame this Part, and both require more care than the literature ordinarily gives them. The first is that a minor is not, by virtue of minority alone, immune from tortious liability under Indian law. The second is that Indian defamation law is bifurcated between a civil tort, largely developed through received common law principles and applied by civil courts awarding damages, and a criminal offence, now codified in the Bharatiya Nyaya Sanhita, 2023.

On the first proposition, Indian courts have long distinguished a minor's incapacity to contract, treated strictly and rendering a minor's agreement void, from a minor's capacity to be sued in tort, treated far more permissively. A minor can, in principle, be held liable for a tort, including defamation, provided the tort does not in substance amount to enforcing a contract the law regards the minor as incapable of making. This doctrine's practical significance is nonetheless limited here for two reasons. First, a minor defendant, even if technically liable, is ordinarily a poor source of recovery, since minors do not typically hold assets or income against which a judgment can be executed. Second, litigation against a minor requires appointment of a guardian ad litem and raises procedural complications that make a direct action unattractive, which is why claimants and scholars gravitate toward the parent as the more practically meaningful target. This is a practical, not a doctrinal, explanation: nothing in Indian tort law converts the minor's own liability into the parent's liability merely because the minor is judgment-proof.

This leads directly to the second and more consequential proposition. Indian civil jurisprudence does not recognise a general doctrine under which a parent is automatically, or vicariously, liable for the torts of a child merely by reason of the parent-child relationship. The foundational common law rule, received into Indian jurisprudence through the general reception of English tort principles, is that a parent is not, as such, liable for the torts of an infant child.⁸ Liability, where it exists, must instead be traced to an independent basis of fault on the parent's own part, most commonly negligent supervision, and occasionally negligent entrustment, where a parent hands a minor an instrument that the parent knows or ought to know is likely to be misused in a harmful way. This paper's own review of reported Indian case law did not locate a Supreme Court or High

⁸*Moon v. Towers* (1860) 8 C.B. (N.S.) 611. The rule has been consistently restated in Commonwealth common law jurisdictions.

Court decision squarely applying a negligent-supervision standard to a parent's liability for a minor's online or digitally generated conduct, and it would overstate the position to describe any part of this area as judicially settled in India. The honest characterisation is that Indian negligent-supervision doctrine exists as an imported common law principle, applied occasionally and without extensive elaboration by Indian courts, rather than as a developed body of domestic jurisprudence, and this absence of directly applicable Indian authority is itself part of the research gap this paper addresses; see Part IX (Limitations).

Turning to defamation, Indian law's civil and criminal tracks must be kept separate, since this paper centres civil parental responsibility, and criminal law questions, while relevant to a complete picture, are not its focus. The criminal offence, previously found in section 499 of the Indian Penal Code, 1860, is now codified in section 356 of the Bharatiya Nyaya Sanhita, 2023, and its constitutional validity, including the compatibility of criminal defamation with the freedom of speech guaranteed by Article 19(1)(a), was upheld by the Supreme Court in *Subramanian Swamy v. Union of India*.⁹ That decision is significant for this paper chiefly for its confirmation that the right to reputation is itself an aspect of the right to life and personal liberty protected by Article 21 of the Constitution, a point that becomes relevant in Part VIII, where the interest of the defamed victim is weighed against the child's and the parent's competing interests.

The civil tort of defamation, by contrast, proceeds independently of the criminal offence, requires no proof of intention to cause harm, and is available to a claimant seeking damages, an injunction, or both. Its elements, defamatory imputation, publication to a third party, and identification of the claimant, are drawn from the general common law and apply without modification to a defamatory statement, image, or video, whatever its technological means of production. Generative artificial intelligence does not, in other words, require a new definition of defamation. It complicates only the questions that sit upstream of that definition: who authored the imputation, who published it, and who, if anyone besides the immediate author, bears responsibility for the fact that it was published. Truth, fair comment, and privilege remain available as defences in the ordinary way, though a defence of fair comment is unlikely to assist a minor who has fabricated, rather than commented upon, a state of affairs, and a synthetic image or video, precisely because it is presented

⁹*Subramanian Swamy v. Union of India*, (2016) 7 S.C.C. 221. The Court held that the offence of criminal defamation under ss. 499 and 500 of the Indian Penal Code, 1860 (now s. 356, Bharatiya Nyaya Sanhita, 2023) is a reasonable restriction on free speech under Article 19(2) of the Constitution, and that the right to reputation forms part of the right to life and personal liberty under Article 21.

as authentic, forecloses reliance on any defence premised on the content being understood as opinion rather than fact.

A further complication specific to synthetic media, developed more fully in Part V, is identification. Civil defamation requires that the claimant be identifiable, by name or description, from the publication. A realistic AI-generated image or video depicting a real, named individual satisfies this requirement straightforwardly. A defamatory text passage that describes its subject only by role, "a teacher at a named school who," may or may not satisfy identification depending on how readily third parties could identify the person, a question that becomes harder to resolve when the statement was AI-generated and its precise wording was not consciously chosen by the minor at all.

IV. NEGLIGENT SUPERVISION AND THE "KNEW OR OUGHT TO HAVE KNOWN" STANDARD

The negligent-supervision standard asks whether a reasonable parent, in the same circumstances, would have known of the risk that the child would engage in the harmful conduct, and would have taken some step, whether supervision, restriction, or intervention, that would probably have prevented the harm. The standard therefore has two limbs: a knowledge or foreseeability limb, and an opportunity-to-intervene limb. Both limbs assume, as their unstated premise, that harmful conduct by a child leaves some observable trace, physical, behavioural, or circumstantial, that a reasonably attentive parent could pick up on before or shortly after the harm occurs. This Part tests that assumption against six constructed hypothetical scenarios, each of which involves a minor using generative artificial intelligence to create defamatory content. None of the scenarios describes an actual reported incident; each is a hypothetical designed to isolate a particular doctrinal variable.

Scenario A. A minor, alone in a bedroom, opens a generative AI application on the family's shared tablet and, within a few minutes, produces and shares a defamatory image of a classmate. The parent was in the next room, uninvolved, and had no prior indication that the child had any interest in, or grievance against, the classmate. Applying the standard, foreseeability is nearly absent: nothing in the parent's prior knowledge would have flagged this specific risk, and the opportunity to intervene, measured in minutes, is negligible. Under conventional negligent-supervision doctrine, liability should not attach. The scenario illustrates the baseline case in which the "knew or ought to have known" standard functions as designed, by screening out cases where the parent genuinely could not have anticipated or interrupted the harm.

Scenario B. The same minor has, on two prior occasions within the preceding month, been confronted by the school for creating unkind or mocking content about other students using an AI tool, and the parent was informed on both occasions. A third incident then occurs. Here, foreseeability is squarely present: the parent had specific, recent, and directly relevant knowledge of a propensity to misuse the same category of tool for a similar kind of harm. The opportunity-to-intervene limb is also more favourable to a claimant, because a reasonable parent with this knowledge could have imposed supervision, removed access, or installed restrictions between the

second and third incidents. This scenario is the strongest candidate for parental liability on orthodox negligent-supervision principles, and it demonstrates that the standard is not incapable of reaching AI-mediated harm; it simply requires the kind of repeated, reported prior conduct that gives the parent actual or constructive notice.

Scenario C. The parent is fully aware that the child uses generative AI applications daily for schoolwork and has, in fact, encouraged this use, but has no reason to suspect that the child has ever used or would use such a tool to produce defamatory content. General awareness of technology use, without more, does not equate to foreseeability of the specific risk of defamatory misuse. This scenario tests, and rejects, an overbroad reading of negligent supervision under which mere parental awareness that a child uses AI tools at all would be enough to found liability; such a reading would functionally convert ordinary, encouraged, educational technology use into a strict-liability trigger, a result inconsistent with the fault-based structure of negligent supervision as a doctrine.

Scenario D. The parent has, on a prior occasion, discovered that the child used an AI tool to produce inappropriate content directed at a sibling, took no corrective action, and continued to provide the child with unrestricted, unsupervised access to internet-connected devices. A second, more serious incident then occurs, this time directed at a person outside the family. This scenario differs from Scenario B in an important respect: the prior notice here did not come from a third party or institution but from the parent's own direct discovery, and the parent's subsequent inaction, continuing unrestricted access despite known risk, is itself an independent act of continuing negligence, not merely a failure to acquire knowledge. Scenario D therefore presents the clearest case for liability of all six scenarios, because it combines actual knowledge, a clear opportunity to intervene, and a deliberate decision not to do so.

Scenario E. The minor creates and publishes defamatory content using a private account, on a device the parent does not access, in a manner that leaves no trace visible to the parent even in principle, for instance because the account uses end-to-end encrypted messaging and the content is deleted from the device after sending. Here the visibility premise collapses entirely. Even an attentive, engaged, non-negligent parent, exercising what would ordinarily be regarded as reasonable supervision, would have no realistic means of detecting or anticipating the conduct. Scenario E is the sharpest illustration of the digital-invisibility problem developed in Part V, and

it is the scenario in which insisting on parental liability would most clearly convert a fault-based standard into something closer to strict liability in substance, because "reasonable" supervision, however diligent, could not have prevented the harm.

Scenario F. The parent discovers the defamatory content only after it has already been published and has begun to circulate, but then fails to take any reasonable step, requesting deletion, contacting the platform, warning the recipient, apologising, or restricting the child's continued access, to limit further dissemination or repetition. This scenario shifts the analytical focus from the creation of the content to its aftermath. Even where a parent could not reasonably have foreseen or prevented the initial publication, as in Scenario A or Scenario E, a distinct and later duty may arise once the parent has actual knowledge of the ongoing harm: a duty to mitigate continuing publication, which is analytically closer to a duty to act once a risk is known than to a duty of anticipatory supervision. Scenario F suggests that Indian negligent-supervision doctrine, if it is to respond coherently to AI-generated harms, may need to distinguish more clearly than it currently does between fault in failing to prevent the initial act and fault in failing to respond once the parent has been made aware of it, since these engage different factual inquiries and arguably different standards of reasonableness.

Read together, the six scenarios do not support the proposition that AI-generated defamatory content by minors defeats the negligent-supervision standard as such. Scenarios B, D, and F show that the standard, applied with ordinary rigour, remains capable of producing a defensible liability outcome wherever the parent had some form of actual or constructive prior notice, or acquired notice after the fact and failed to act on it. What the scenarios do show, most starkly in Scenario E and, to a lesser extent, Scenario A and Scenario C, is that a meaningful category of cases exists in which even a genuinely careful and engaged parent would have no realistic opportunity to detect or prevent the harm, because the technology permits conduct that is instantaneous, privately executed, and leaves no trace of the kind the doctrine has always implicitly required. It is this category, not the doctrine as a whole, that generative artificial intelligence renders genuinely novel, and it is this category that Part IX's proposed framework is chiefly designed to address.

V. AI-GENERATED DEFAMATORY CONTENT AND THE PROBLEM OF DIGITAL INVISIBILITY

The preceding Part isolated the doctrinal variable; this Part explains, in more general terms, why generative artificial intelligence specifically, as distinct from earlier internet technologies, poses this problem with unusual force. Three features of generative AI, taken together, are responsible: speed, the elimination of specialised skill as a precondition for realistic fabrication, and the private, self-contained nature of the generation process.

On speed, earlier forms of online harm by minors, cyberbullying through ordinary text messages, the circulation of an embarrassing but genuine photograph, typically required sustained effort over time, which increased the chance of behavioural warning signs accumulating, or a pre-existing piece of authentic material whose creation or acquisition itself often left some trace. Generative AI collapses the interval between intention and realistic, publishable output to a matter of seconds, and requires no antecedent material beyond a short text prompt. On the elimination of specialised skill, the production of a convincing forged photograph or synthetic audio previously required technical training or specialised software, both of which functioned, imperfectly but genuinely, as informal barriers limiting the population of minors capable of causing this kind of harm. Generative AI applications, many built into everyday consumer software and requiring no more skill than typing a sentence, remove this barrier entirely, a point emphasised by recent comparative scholarship arguing that responsibility should shift toward developers of tools accessible to and usable by children without any meaningful safeguard.¹⁰ On privacy of the generation process, the entire act, prompting, generation, and often immediate publication, occurs within a single application on a personal device, frequently a device the minor also uses for legitimate homework, and frequently at a time of day, or in a room of the house, in which parental presence would not itself be unusual or suspicious. The harmful act is, in other words, camouflaged by the surrounding pattern of ordinary, encouraged technology use, in a way that a late-night departure from the house or a suspicious phone call is not.

¹⁰Ayelet Gordon-Tapiero, Yotam Kaplan & Gideon Parchomovsky, *Deepfake Liability*, 104 N.C. L. Rev. 377 (2026) (arguing that generative AI developers should face products-liability-style accountability because their tools are widely accessible, including to minors, and lack effective safeguards against misuse, and that user- and platform-liability theories alone are an inadequate response to the deepfake problem).

The combined effect of these three features is what this paper calls digital invisibility: a condition in which conduct that negligent-supervision doctrine has always assumed will, at some point, become observable to a reasonably attentive parent simply does not become observable, not because the parent has failed to look, but because there is structurally nothing to see until after the harm has been done. It would be a mistake, however, to conclude that digital invisibility makes the "knew or ought to have known" standard unworkable in the abstract. The standard remains coherent as a legal test; what changes is the factual density of situations to which it will, on a fair application, yield a negative answer. Generative AI does not break the doctrine; it shifts a larger share of real-world fact patterns into the category the standard has always been designed to exclude from liability, harm a reasonable parent could not have anticipated or prevented, the outcome illustrated by Scenario A, Scenario C, and especially Scenario E in Part IV.

This has an important, and somewhat counter-intuitive, implication for law reform. A policy response that focuses solely on tightening the parental negligence standard, for instance by presuming knowledge from the mere fact that a minor had access to a generative AI tool, would not solve the underlying problem; it would instead convert negligence into something close to strict liability for households that did nothing wrong, since, as Scenario E shows, no degree of ordinary diligence would have made the conduct visible. A more coherent response recognises that where digital invisibility is genuine, the doctrinally honest conclusion is that the parent is not the best-situated party to prevent the harm, and that responsibility should look instead to actors who are structurally better placed to prevent or limit it: the platform that hosts publication, which is examined in Part VI, and, per the comparative scholarship noted above, potentially the developer of the generative tool itself, whose design choices, the presence or absence of content filters, age verification, or watermarking, do bear a genuine causal relationship to the ease with which the harm occurred, in a way that the parent's supervision, however diligent, could not.

One further complication, distinct from digital invisibility, deserves a flag rather than resolution here: attribution. Establishing that a defamatory output was substantially the product of the minor's own creative choice, rather than a largely unpredictable output of the model responding to an innocuous prompt, raises evidentiary difficulties that scarcely existed for earlier forms of minor-authored defamatory content. This problem compounds, without duplicating, the visibility problem, and a reform proposal that solves one without addressing the other will remain incomplete.

VI. PARENT, MINOR, PLATFORM AND AI DEVELOPER: ALLOCATION OF RESPONSIBILITY

Having established that a meaningful category of AI-generated defamatory harm by minors falls outside any fair application of negligent-supervision doctrine, this Part maps the four actors among whom responsibility might in principle be distributed, and asks, for each, whether Indian law currently provides, or could principled provide, a basis for responsibility.

The *minor's* own personal liability was addressed in Part III and need only be recalled here: it exists in principle but is of limited practical value to a claimant because of the minor's ordinary lack of assets and the procedural complexity of suing a minor directly. It is nonetheless doctrinally important to keep the minor's liability visible, rather than collapsing it into the parent's, because doing so preserves the important distinction, developed further in Part VIII, between holding the child accountable, which has an educative and dignity-respecting function, and holding the parent liable, which does not serve the same function and carries different costs.

The *parent's* responsibility, as Parts III through V have shown, is properly confined to situations of demonstrated actual or constructive notice, whether arising from a pattern of prior conduct, as in Scenario B, from a parent's own prior discovery followed by continued inaction, as in Scenario D, or from a parent's failure to mitigate harm after acquiring knowledge of it, as in Scenario F. Outside these categories, imposing liability on the parent is not a principled allocation of responsibility; it is a convenient one, chosen because the parent is easier to locate and more likely to have resources than the minor, and Indian law should resist converting a doctrine of fault into a doctrine of convenience.

The *platform*, or intermediary, occupies an intermediate position that has recently been substantially strengthened by regulation. Under section 79 of the Information Technology Act, 2000, an intermediary is conditionally exempt from liability for third-party content it merely hosts or transmits, provided it observes prescribed due diligence.¹¹ The content of that due diligence obligation, originally set out in the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, was substantially amended with effect from 20 February 2026

¹¹Information Technology Act, 2000, s. 79.

specifically to address synthetically generated information.¹² The 2026 amendment inserts a statutory definition of synthetically generated information, requires that every reference to "information" elsewhere in the Rules be read to include synthetic content, and imposes on intermediaries, particularly significant social media intermediaries, obligations to deploy reasonable technical measures to detect and label such content, to preserve labels and embedded metadata against removal, and to comply with substantially shortened takedown timelines, including a compressed window for content involving non-consensual intimate imagery or deepfakes.¹³ This is a materially more direct regulatory response to the problem examined in this paper than anything available under general tort principles, because it targets the moment at which digitally invisible, privately created content becomes publicly visible, which is also the moment at which a platform, unlike a parent, is structurally capable of detecting and acting on it. Where a platform fails to comply with these obligations after content is reported or otherwise comes to its notice, it risks losing the safe harbour under section 79 and becoming exposed to liability in the ordinary way, a mechanism this paper regards as considerably better matched to the digital-invisibility problem than any plausible expansion of parental fault.

The *developer* of the generative AI tool is the most doctrinally undeveloped of the four actors under Indian law, which currently has no dedicated products-liability regime addressing software, let alone generative software, of this kind. The comparative scholarship reviewed for this paper argues, in the American products-liability context, that a generative AI developer's design choices, whether the tool includes content filters, watermarking, or age-appropriate restrictions, bear a genuine causal relationship to the ease with which a minor can produce realistic defamatory content, and that this causal relationship justifies treating the developer as a proper object of liability in appropriate cases, independent of any fault on the part of the immediate user.¹⁴ Indian law has no equivalent doctrine at present, and this paper does not claim that one currently exists; it observes only that the underlying causal logic of the argument, that the party best able to design out a foreseeable, product-inherent risk should bear some responsibility for failing to do so, is not

¹²Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026, G.S.R. 120(E) (10 February 2026).

¹³*Ibid.*, amending r. 3 of the 2021 Rules generally and inserting r. 3(3), which requires prominent disclosure of synthetically generated information at the point of upload and prohibits users or intermediaries from removing or suppressing that disclosure or associated metadata; the same amendment reduces the takedown timeline applicable to certain categories of unlawful synthetic content, including deepfakes and non-consensual intimate imagery.

¹⁴Gordon-Tapiero, Kaplan & Parchomovsky, *supra* note 9.

foreign to Indian law generally, which recognises analogous principles in other contexts, and that this is an area meriting separate, focused research beyond the scope of the present paper.

No single one of these four actors can fairly absorb the entire weight of responsibility for AI-generated defamatory content by minors. The minor's liability is real but practically hollow. The parent's liability is principled only within the bounded category of demonstrated notice. The platform's responsibility, newly strengthened by the 2026 Amendment Rules, is the most operationally effective lever currently available, precisely because it operates at the point of publication rather than private creation. The developer's responsibility is the least developed under Indian law but rests on the soundest causal logic for the digitally invisible category of harm identified in Part V. This four-way, context-dependent allocation, rather than a search for a single liable party, is the analytical foundation for the framework proposed in Part IX.

VII. COMPARATIVE APPROACHES

This Part examines France, Mauritius, and the United Kingdom, selected because each represents a materially different structural answer to the question this paper addresses, rather than because any of the three has yet enacted AI-specific parental liability rules; none has.

France operates under a civil-law regime of presumed parental fault. Article 1242, paragraph 4, of the Civil Code renders parents who jointly exercise parental authority liable, in solidum, for damage caused by a minor child living with them.¹⁵ French jurisprudence has hardened this presumption into liability that arises automatically once the child's act is shown to be the direct cause of the harm, irrespective of any fault by the child, and that can be displaced only by force majeure or the victim's own fault, not by proof that the parents exercised reasonable supervision. The Cour de cassation confirmed this in *Bertrand v. Domingues*, holding that only an external cause or the victim's fault, not evidence of careful supervision, could exonerate a parent, and reaffirmed it in *Levert*.¹⁶ This regime prioritises victim compensation and administrability over any inquiry into individual fault, and it would, applied to Scenario E in Part IV, produce parental liability even though the conduct was genuinely undetectable, exposing the central weakness of a presumed-fault model for the problem this paper examines: it compensates the victim by holding liable a parent who could not, on any realistic view, have prevented the harm.

Mauritius, whose civil code shares a common French-derived heritage, imposes an analogous regime under Articles 1382 to 1384 of its own Civil Code, under which a person is answerable not only for damage caused by their own act but also for damage caused by persons for whom they are responsible, a provision applied by Mauritian courts, and on appeal by the Judicial Committee of the Privy Council, on a presumed basis once the relevant relationship of responsibility is established.¹⁷ The Mauritian comparator is useful less for its substantive rule, which largely mirrors the French position, than for what it illustrates about transplantation: a civil-law presumed-fault regime can operate within a system that, like India's, also carries substantial common law

¹⁵Code civil (Fr.), art. 1242, al. 4 (formerly art. 1384, al. 4).

¹⁶Cass. civ. 2e, 19 February 1997, D. 1997, 265 (*Bertrand v. Domingues*); Cass. civ. 2e, 10 May 2001 (*Levert*). See also Cass. civ. 2e, 17 February 2011, No. 10-30.439. A 2024 decision of the Cour de cassation sitting en banc went further, removing cohabitation as a strict precondition for parental liability: Cass. ass. plén., 28 June 2024.

¹⁷Mauritius Civil Code, arts. 1382 to 1384. For judicial application of the presumption of liability under article 1384, see *Harvey v. Ramyad*, [2021] UKPC 23 (on appeal from the Supreme Court of Mauritius), discussing the presumption against a "gardien" and the narrow scope for exoneration.

inheritance, suggesting that the obstacle to adopting such a model in India is not structural incompatibility but a genuine policy choice about where the burden of proof should fall.

The *United Kingdom* represents the closest doctrinal relative of Indian law, since Indian tort principles here were received directly from English common law, including the *Moon v. Towers* rule discussed in Part III. English law has not adopted a fault-based or presumed-fault civil liability regime for parents of tortfeasor children; instead, it has developed a statutory quasi-liability mechanism outside tort law. Section 8 of the Crime and Disorder Act 1998, as amended by section 18 of the Anti-social Behaviour Act 2003, empowers a court to make a parenting order against a parent, requiring compliance with specified requirements and, where appropriate, attendance at a counselling programme, where desirable to prevent repetition of a child's harmful conduct.¹⁸ This mechanism imposes obligations on a parent in response to harmful conduct without reframing that conduct as a civil wrong for which the parent is liable in damages. The English comparator thus demonstrates a genuine third model, separating civil compensation, which continues to depend on proof of parental fault, from a distinct, non-compensatory mechanism aimed at correcting parental conduct going forward.

What can India draw from these three models. The French and Mauritian presumed-fault regimes offer administrability and victim protection but, applied to genuinely undetectable AI-generated harm, risk holding blameless parents liable in a manner this paper regards as normatively indefensible on the facts of Scenario E, and are a poor fit for a common law system whose negligence doctrine has never operated on a presumption of fault. The English quasi-liability model is comparatively attractive because it decouples victim compensation, which should remain fault-based, from a separate, forward-looking mechanism directed at improving parental engagement in cases short of civil fault, a distinction drawn on directly in Part IX. None of the three jurisdictions has yet confronted AI-generated content by minors in its parental liability doctrine, so the comparison here is one of structural analogy rather than a directly transplantable rule.

¹⁸Crime and Disorder Act 1998 (UK), s. 8, as amended by the Anti-social Behaviour Act 2003 (UK), s. 18.

VIII. CONSTITUTIONAL, CHILD-RIGHTS AND POLICY CONSIDERATIONS

Any reform of parental responsibility in this area sits at the intersection of three constitutionally grounded interests that do not point in the same direction. The victim's interest in an effective remedy for reputational harm is itself constitutionally rooted, since the Supreme Court has held that the right to reputation forms part of the right to life and personal liberty under Article 21.¹⁹ This interest counsels in favour of a liability rule generous enough to give a defamed victim a realistic prospect of compensation, particularly given the practical hollowness of a minor's own liability described in Part VI. Set against this is the minor's own developing autonomy and the legitimate interest, itself linked to Article 21 and to the constitutional value placed on access to education and technology, in not being subjected to disproportionate parental or state surveillance as the price of permitting a child to use tools that are, for the great majority of uses, entirely legitimate and educationally valuable. A framework that responds to the risk of rare AI-generated defamatory misuse by encouraging blanket parental surveillance of a child's private communications, or by discouraging parents from permitting independent technology use at all, would impose a real cost on the child's own developing autonomy and privacy interests, and, if it deterred legitimate educational use of generative tools, a real cost to the child's access to technology and, indirectly, to education itself.

This tension is not merely rhetorical. A presumed-fault or strict-liability rule of the French or Mauritian kind, examined in Part VII, creates exactly the incentive structure most likely to produce disproportionate surveillance, because the only way for a parent to reduce exposure under such a rule is to monitor the child's device use as closely and continuously as possible, an incentive a legal system genuinely committed to a child's privacy and developing autonomy should be reluctant to create, particularly for adolescents whose stage of development ordinarily calls for a gradual expansion, not a technologically enforced contraction, of private space. Article 14's guarantee of equality is also engaged at the margins, since a rule keyed to a household's capacity for constant device monitoring will be easier to satisfy for parents with the time, literacy, and resources to install parental-control software than for parents without them, a distributive concern any reform proposal should be candid about.

¹⁹*Subramanian Swamy, supra* note 8.

The correct resolution, this paper argues, is not to treat one of these interests as categorically prevailing over the others, but to use the fault-based structure of negligent supervision itself as the mechanism for balancing them, applied with the technologically literate content developed in Parts IV and V rather than an inherited, purely physical-world conception of what a reasonable parent can know and do. A correctly calibrated negligence standard already achieves much of this balance: it protects the victim's Article 21 interest wherever the parent genuinely had notice and failed to act, as in Scenario B, D, and F, while protecting the child's and parent's privacy and autonomy interests wherever the harm was genuinely undetectable, as in Scenario A, C, and E, by declining to impose liability rather than forcing continuous surveillance to avoid it. The remaining policy work, addressed in Part IX, is to make that calibration explicit, rather than abandon the fault-based structure for either automatic liability, which risks disproportionate surveillance, or blanket immunity, which would leave victims without a remedy in cases where a parent did have sufficient notice.

IX. TOWARDS A FRAMEWORK FOR PARENTAL RESPONSIBILITY IN INDIA

This Part evaluates five possible directions for Indian law, and proposes that Indian courts adopt the second of them, developed with explicit reference to digital conduct, rather than any of the other four.

Option 1, maintaining the existing personal-negligence standard without modification, has the advantage of doctrinal continuity and requires no legislative intervention, but leaves Indian courts without any guidance on how the "knew or ought to have known" inquiry should be applied to a fact pattern, generative AI misuse, that did not exist when the doctrine's content was last meaningfully elaborated. The risk is not that courts will reach the wrong result, since the six scenarios in Part IV suggest that a careful, unguided application of the existing standard is broadly capable of reaching defensible outcomes, but that courts will reach inconsistent results across similar fact patterns in the absence of any articulated framework for what "reasonable" supervision of AI tool use actually requires.

Option 2, judicial clarification of a negligent-supervision standard specifically applicable to digital conduct, is the option this paper recommends. It requires no new statute, only a considered appellate articulation of the factors relevant to reasonableness in this specific context, and it preserves the fault-based structure defended in Part VIII. This paper proposes that such a standard be organised around what may usefully be labelled a *reasonable digital supervision* inquiry, built from eleven factors: the age of the minor; the nature and severity of the conduct; the accessibility of the specific AI tool used; the foreseeability of misuse given the tool's known functions; any previous conduct of the minor involving similar tools or similar harms; the parent's actual knowledge, from whatever source; the parent's realistic opportunity to intervene, assessed against the genuinely private and rapid character of AI generation described in Part V; the parent's practical control over the device or account in question; the presence or absence of safeguards, whether parental controls or safeguards built into the tool itself, that were available to the parent but not used; whether reasonable supervision, realistically defined rather than idealised, could actually have prevented the harm, which requires courts to take Scenario E's lesson seriously rather than assuming supervision is always capable of preventing digital harm; and whether the parent acted reasonably after discovering the conduct, capturing the distinct post-discovery duty identified in

Scenario F. This framework does not lower the bar for claimants; it raises the quality of the inquiry, by directing courts away from an undifferentiated, retrospective question of whether the parent was a good parent in some general sense, and toward the specific, forward-looking question the doctrine has always asked but that Indian courts have not yet had occasion to answer in this technological context.

Option 3, a rebuttable presumption of parental responsibility confined to narrowly defined circumstances, such as cases involving a documented prior pattern of misuse, would in substance formalise the outcome already reached in Scenario B and Scenario D under ordinary negligence principles, and adds comparatively little beyond Option 2 while introducing the administrability costs, discussed in Part VII in connection with France and Mauritius, of a presumption that a parent must then affirmatively rebut. It is not rejected outright, but this paper regards it as an incremental refinement available to the legislature at a later stage, once Option 2 has generated a body of appellate guidance against which the narrow circumstances warranting a presumption could sensibly be defined.

Option 4, statutory parental duties concerning serious digital misconduct enacted independently of tort law, modelled loosely on the English parenting-order mechanism examined in Part VII, has genuine attractions as a complement to, rather than a substitute for, Option 2. Because it need not be compensatory, it could operate in cases falling short of provable negligence, directing a parent toward supervisory or educational intervention without exposing that parent to damages disproportionate to demonstrated fault. This paper regards Option 4 as a promising legislative supplement worth separate study, not as a substitute for clarifying the tort standard, since a purely regulatory mechanism does nothing to address the victim's need for a civil remedy where parental fault genuinely is established.

Option 5, a broader statutory framework distributing responsibility among parent, minor, platform, and AI developer, is the most structurally ambitious option and is, in the medium term, the direction in which this paper believes Indian law should eventually move, given the analysis in Part VI showing that no single actor can fairly absorb the whole of this problem. The 2026 Amendment Rules already represent a significant, if partial, step in this direction on the platform side. A comparable statutory articulation of developer responsibility does not yet exist in Indian law, and this paper does not attempt to draft one; it notes only that Option 5 is best matched to the

four-way allocation defended in Part VI, and that Option 2 should be understood as the parental component of that larger, still-developing framework rather than a complete answer on its own.

Weighed against the criteria of fairness, administrability, proof, protection of children's rights, parental autonomy, victim compensation, technological feasibility, and the risk of over-deterrence, Option 2 scores best among the five. It is fairer than Option 3 and the French and Mauritian models in Part VII, because it retains proof of fault rather than shifting the burden onto parents for harm they could not have prevented. It is more administrable than Option 5 in the near term, requiring judicial elaboration rather than new primary legislation. It protects children's rights and parental autonomy better than any presumed-fault model, for the reasons developed in Part VIII, because it does not incentivise disproportionate surveillance as the only route to reduced liability exposure. Its principal limitation, candidly stated, is that it leaves the digitally invisible category of harm identified in Part V without a civil remedy against the parent, which is precisely why this paper regards Option 2 as necessary but not sufficient, and situates it within a longer-term trajectory toward Option 5, in which the platform obligations already strengthened by the 2026 Amendment Rules, and a still-undeveloped Indian conception of AI-developer responsibility, would carry the weight parental liability cannot fairly carry in those cases.

A final word on limitations is warranted before the conclusion. This paper's account of Indian negligent-supervision doctrine is necessarily built on limited direct Indian appellate authority, and the absence of reported decisions squarely on point should not be read, and has not been read here, as proof that Indian law contains no such doctrine; it is read, more cautiously, as evidence that the doctrine remains underdeveloped in Indian courts, a limitation acknowledged in Part III. The regulatory landscape addressed in Part VI, particularly the 2026 Amendment Rules, is recent enough that its practical operation, including how frequently intermediaries in fact comply with the new synthetic-content obligations, has not yet been tested by litigation, and this paper's assessment of its effectiveness is accordingly provisional. Finally, the pace of change in both generative AI technology and Indian digital regulation is such that any framework proposed here should be understood as addressing the structural shape of the problem rather than its every technical detail, which will continue to evolve.

X. CONCLUSION

Indian tort law has never made parents automatically liable for the wrongs of their children, and this paper has found no principled reason to depart from that position merely because the instrument of harm is now a generative AI application rather than a rumour, a letter, or a fist. What generative AI has changed is not the legal test but the factual terrain on which it operates: a meaningful category of AI-generated defamatory conduct by minors is now genuinely, structurally invisible to even a conscientious parent until after the harm is done, in a way earlier forms of childhood misconduct rarely were. The doctrinally honest response is not to strain negligent-supervision doctrine into something closer to strict liability, which would punish diligent parents for harm they could not have prevented and create incentives toward disproportionate surveillance, but to state the "knew or ought to have known" standard with enough technological precision that courts can distinguish the parent who had real notice and failed to act from the parent who simply could not have known. Where genuine digital invisibility is present, responsibility is more fairly addressed through the intermediary obligations recently strengthened by the Information Technology Rules, and, in time, through a more developed Indian conception of the responsibility of those who design and release generative AI tools with foreseeable potential for misuse by minors, than through any expansion of parental fault. India does not need a new law of parental liability for the age of generative artificial intelligence. It needs a more precise application of the law it already has, coupled with a clearer allocation of responsibility to the platform and the developer, who are, in the cases that matter most, better placed than any parent to prevent the harm.