Academic Journal of Management and Social Sciences
ISSN: 2958-4396 | Vol. 18, No. 1, 2026

Design and Governance of Generative AI for Child Users: Cultural-Ethical Translation and Family Relationship Impact Assessment

Siying Li [1, 2], Chen Gao [1], Chi Gong [1, *]

[1] Department of Global Business, Kyonggi University, Suwon 16227, South Korea
[2] School of Management, University of Science and Technology Liaoning, Anshan, Liaoning, 114000, China

*** Corresponding author:** Chi Gong

Abstract: Generative artificial intelligence (GenAI) can hand a child an answer that is factually sound and emotionally soothing, yet the child may walk away from it less willing to listen at home. Judging the answer on its own terms would never catch this. This paper introduces family relationship impact assessment (FRIA), an approach for tracing how conversational GenAI becomes involved in the ties between children and the people they live with. The framework was built through conceptual analysis, an integrative reading of the literature, and mapping onto current governance guidance. Five dimensions informed by Confucian relational ethics—relationality, contextual proportionality, judgment, truthfulness, and motivational orientation—form its core, with children's rights bounding how each is interpreted. For every dimension we derive two questions: one asks which human function an assistant must not displace, the other asks what assistance it may legitimately offer. Together they yield ten preliminary indicators dealing with role boundaries, developmental expectations, reasoned judgment, warranted correction, and responsibility repair. Five thought experiments test how the indicators work, and the Govern, Map, Measure, and Manage functions of the NIST AI Risk Management Framework anchor their use. Organizations gain concrete questions for setting product roles, reviewing interactions, and deciding who is responsible for correction. FRIA is still a conceptual proposal: the indicators remain unvalidated, and nothing here shows how often relational harm occurs or whether the suggested controls would actually work.

Keywords: Generative Artificial Intelligence; Child Users; Technology Governance; Family Relationship Impact Assessment; Cultural-Ethical Translation; Confucian Relational Ethics.

1. Introduction

Consider a child who asks an assistant whether a parent's rule on screen time is fair. Nothing in the reply may be factually wrong or prohibited. Even so, the child can carry it back to the parent as if it were a neutral verdict. What has happened is that the assistant has stepped into a quarrel between people who must go on living under the same roof. How it speaks, what it asks, and how confident it sounds all help decide whether its words become material for talking things over—or a verdict that ends the conversation. Assessment should therefore track the relationship once advice leaves the chat window.

What interests us is conversational GenAI that children use around family rules, learning, emotional support, and responsibility, and we take the child-AI-family-member triad as the unit of analysis. The triad includes family members who never touch the system yet whose conduct or reasoning the assistant ends up interpreting. Product role, age settings, memory, humanlike features, and access to human help all shape that interpretation. We make no claim about every application of AI involving children.

Guidance on protecting children who use AI is already plentiful. UNICEF (2025) spells out requirements on safety, privacy, fairness, accountability, and development. The National Institute of Standards and Technology structures AI risk management around Govern, Map, Measure, and Manage (NIST, 2023, 2024), and ISO/IEC 42005:2025 covers the assessment of impacts on individuals and society (ISO, 2025). China's Interim Measures for the Administration of Anthropomorphic Interactive Artificial Intelligence Services likewise address excessive accommodation, dependency, and harm to real-world relationships (Cyberspace Administration of China et al., 2026). Legally, however, these instruments reach a narrower territory than the subject of this article: they target sustained anthropomorphic emotional interaction and leave out educational and informational services where no such interaction occurs.

How such requirements bear on an ordinary family disagreement remains to be worked out. An assistant can soothe a child by accepting a one-sided story, settle a hard choice by simply making it, or hand the same advice about independence to children of very different abilities. Often the trouble is invisible inside any single reply. It shows up in whose reasons get heard, and in who is expected to answer for things afterwards. Family expectations color how these interventions are read, though no nationality or language reveals them directly.

Empirical research gives the question urgency without closing it. In Wen et al. (2025), families pictured GenAI agents helping with household management, tutoring, and family support, yet insisted that major decisions stay with people and that communication remain direct. Sycophancy research raises a different worry: assistants that accommodate users at the price of honest feedback (Sharma et al., 2024). In adult experiments, Cheng et al. (2026) found that sycophantic interaction could inflate confidence in being right and weaken the willingness to accept responsibility or mend conflict. None of these adult findings transfers automatically to children. They do, though, justify asking whether a pleasant

conversation necessarily helps the relationships its advice will later touch.

Confucian relational ethics offers one standpoint for pressing that question. Relational accounts of selfhood situate judgment inside ongoing obligations and roles (Hall & Ames, 1998; Fei, 1998; Hwang, 2009), and cultural psychology explains why personal expression and interdependence do not carry the same meaning everywhere (Markus & Kitayama, 1991). We borrow these accounts to sharpen ethical concerns—not to depict some uniform Asian family, which no such thing is. What a family actually holds to is always something to find out.

Li and Gao's (2026) analysis of eight ethical categories—ren (benevolence), yi (righteousness), li (ritual propriety), zhi (wisdom), xin (trustworthiness), xiao (filiality), he (harmony), and chi (a sense of shame)—is the immediate starting point. They organized the discussion under relationality, regulation, discernment, truthfulness, and motivation. We rebuild these into five dimensions fit for governance: relationality, contextual proportionality, judgment, truthfulness, and motivational orientation. Turning an ethical concern into a product requirement demands an argument about which function is at stake and which behavior threatens it; otherwise the well-worn gap between principles and accountable practice simply persists (Mittelstadt, 2019).

We name the proposed process family relationship impact assessment (FRIA). It asks organizations to name the relational changes at stake, say why they matter, record the decisions taken, and open routes for correction. FRIA is a working concept rather than an existing standard. Three questions organize the inquiry:

RQ1: In children's family-related use of GenAI, what qualifies as relational risk, and how does it relate to content risk, individual harm, and children's rights?

RQ2: Through what route can concerns drawn from Confucian relational ethics be turned into clearly bounded governance constructs and preliminary indicators that leave children's rights intact?

RQ3: At which points could FRIA enter product design and lifecycle governance, and what controls, records, and division of responsibility would using it demand?

Definitions and a derivation procedure come first; hypothetical interactions and implementation follow. No participant data or model outputs were collected anywhere in the analysis. The claim, accordingly, concerns how relational effects could be assessed—and the mechanisms and indicators we propose remain open to empirical challenge.

2. Literature and the Assessment Problem

2.1. The Child Within Family Relationships

Child rights impact assessment (CRIA) is the necessary place to begin. It asks organizations to weigh actual and potential effects on children, covering their best interests, their chances of being heard, and their access to remedy (Livingstone & Pothong, 2025; UNICEF, 2024). Calling a product family-friendly can never demote safety or privacy to second place. Any relational assessment has to carry these protections forward intact.

The family, though, poses a question that examining one user in isolation cannot answer. Children learn to choose and to accept obligations through repeated encounters with parents, siblings, grandparents, and other caregivers. Should a conversational assistant become the preferred interpreter of those encounters, which accounts get trusted—and how disagreements unfold—could shift. The shift might help a child voice a concern nobody had heard before. It might just as well make direct clarification rarer. Whichever way it goes, the effect happens between people, even when only one of them has opened the app.

That is why FRIA is designed as a relational module inside CRIA. Its questions ask whose voice stays available, whether authority can be challenged, and whether the assistant helps people communicate when communication is safe. Harmony at home proves nothing on its own: it can reflect a child under pressure to keep quiet. A child's instant approval of an answer likewise tells us little about whether the interaction served the child's longer-term interests.

2.2. Relational Ethics as an Analytical Starting Point

Confucian relational ethics turns attention toward obligations arising within particular relationships. Hall and Ames (1998), Fei (1998), and Hwang (2009) supply the resources for seeing why care, role expectations, and responsibility resist being reduced to separate individual preferences. For anyone assessing conversational AI, this makes intergenerational negotiation and the repair of conduct legitimate objects of inquiry. Work on culture and self-construal adds a caution: what autonomy means depends partly on the social setting in which it is exercised (Markus & Kitayama, 1991; Shweder, 1991).

Five dimensions carry over from the earlier reconstruction, reworded to clarify how governance would use them (Li & Gao, 2026). Relationality concerns care and coordination among people. Contextual proportionality asks what a child can reasonably handle in a given situation. Judgment turns on the reasons behind a choice; truthfulness asks whether those reasons hold up under pressure to agree. Motivational orientation asks whether reflection can end in correction. These distinctions are the authors' analytical choices—neither a taxonomy lifted from the classics nor a report of measured parental preferences.

Restraint is needed in using them, too. An ethical tradition can make a concern visible without settling which technical control is justified. Preserving communication, say, still requires an account of when communication is safe and what the assistant should do when it is not. Families with different ethical commitments can raise the very same concern. Whether these constructs describe it adequately beyond their original context is both an empirical and a normative question.

Cultural essentialism would wreck the whole exercise. Treat a group as internally uniform and an inquiry into family commitments hardens into an assumption about them. We therefore refuse nationality, surname, location, and language as evidence that a family endorses obedience or any particular distribution of power. Nor can filiality and harmony ever license humiliation, unlimited adult control, or denial of a child's privacy and participation.

2.3. What Existing Governance Processes Provide

Mittelstadt's (2019) critique of principle-based AI ethics names an organizational difficulty: declaring values creates neither responsibility nor enforcement. Impact assessment can provide a setting where judgments are explained and acted upon. But approaches vary widely in scope and

implementation, as the review by Stahl et al. (2023) shows. A relational assessment must therefore state both what it adds and which existing processes could carry it.

One such structure is the NIST AI Risk Management Framework (AI RMF). Govern covers policy and accountability; Map examines use contexts and consequences; Measure builds the evidence base; Manage handles treatment and monitoring (NIST, 2023). The GenAI profile extends the framework to risks involving testing and human-AI configurations (NIST, 2024). ISO/IEC 23894:2023 offers risk-management guidance, and ISO/IEC 42005:2025 addresses impact assessment (ISO, 2023, 2025). Documentation across development likewise sits at the center of Raji et al.'s (2020) account of internal algorithmic auditing.

Questions about family relationships fit inside these sources. What they do not supply, on their own, is any way to recognize an assistant's unjustified authority in a parent-child dispute, or to tell reassurance apart from dodging responsibility. FRIA proposes content for exactly that domain-specific inquiry. Table 1 shows where it could supplement existing guidance—without implying that the guidance says nothing about relationships.

Table 1. Existing Governance Guidance and the Proposed Relational Supplement

Guidance source	Primary assessment concern	Process or protection offered	Where FRIA enters
UNICEF guidance on AI and children (UNICEF, 2025)	Rights of the child: safety, privacy, fairness, development, well-being	Child-centered protections plus requirements on responsible actors	In a given interaction, how do these protections bear on negotiation, family roles, and communication?
NIST AI RMF (NIST, 2023, 2024)	Risks across the AI lifecycle	Govern, Map, Measure, and Manage	Which relational mechanisms and scenarios belong in the organization's assessment?
ISO/IEC 42005:2025 (ISO, 2025)	AI's impacts on individuals and society	Identification, documentation, review, improvement	What evidence is required when the affected unit is a relationship among a child, an assistant, and a family member?
China's anthropomorphic interaction measures (Cyberspace Administration of China et al., 2026)	Sustained anthropomorphic emotional interaction; protection of minors	Guards against excessive accommodation, dependency, harm to real-world relationships, and provider irresponsibility	How should a covered product examine these concerns? Educational or informational services lacking sustained emotional interaction lie outside the stated scope.

2.4. Agreement and Perceived Authority

An assistant can accumulate authority without ever announcing that it has replaced a parent. To a child it may look knowledgeable, always available, and apparently impartial. Advice delivered in that frame can weigh more than the system's formal role would justify. The problem sharpens when the assistant endorses whoever happens to be speaking, while the family members who are absent get no chance to clarify their reasons.

Sycophancy is Sharma et al.'s (2024) term for accommodation that compromises truthfulness or independent judgment. The category reaches beyond outright declarations that the user is right. Ye et al. (2026) separate accommodation of beliefs from accommodation of personal traits and emotions, and treat framing, omission, and tone as part of the picture alongside overt agreement. Their account is a preprint; what matters here is the conceptual point. A response can look balanced while quietly erasing any need to question the user's version of events.

Anthropomorphic design can amplify the social reading of such agreement. A remembered personal history, a first-person identity, warm forms of address—an assistant can come to feel like a continuing participant in family life. China's measures on anthropomorphic interaction already name excessive accommodation, dependency, and harm to real-world relationships as regulatory concerns (Cyberspace Administration of China et al., 2026). Converting that concern into assessment still requires looking at the particular interaction and at the authority the system is accorded.

2.5. Questions Left Open

Read together, this literature exposes a practical gap between noticing relational concerns and reviewing an actual product. A child's satisfaction sits right on the surface of a conversation; gradual movements in trust or willingness to negotiate do not. Evaluation that rewards agreement can miss the worth of correction altogether. Cultural interpretation compounds the difficulty: organizations need grounds for treating a concern as relevant without letting a cultural label harden into a behavioral rule.

The framework below responds by tying each ethical concern to a protected relational function, a candidate mechanism, and a review question. Responsibility then has to land on actors who can actually change the product feature involved. Telling parents to supervise use is no substitute for decisions about training, interfaces, prompts, and defaults that parents never control.

3. Approach to Framework Development

3.1. Conceptual Design

This is a conceptual study. In the spirit of Whetten (1989) and Jaakkola (2020), it develops constructs, examines how they relate, and states the conditions under which its propositions could hold. Its evidence supports distinctions and reasoned proposals—nothing more. Because it contains no surveys, interviews, experiments, or sampled model conversations, it cannot estimate how often risks occur or report how children actually respond to a particular model.

Normative conceptual analysis, cultural-ethical translation, and standards mapping are combined here. The normative part establishes which functions deserve protection and where claims about family authority must stop. Translation asks how a concern drawn from Confucian ethics can become a governance question. Mapping then locates that question inside existing assessment processes. An integrative reading of literature on children's rights, cultural psychology, family human-computer interaction (HCI), sycophancy, and impact assessment underpins the construction—the aim being synthesis for a defined problem, in line with the purpose of integrative review that Torraco (2016) describes, rather than an exhaustive systematic review.

Translation is constrained in several ways. It has to preserve whatever concern animates the ethical source—honest correction, say, or proportionate expectations of responsibility. It must stay compatible with children's safety, dignity, expression, and remedy. A translated requirement should let an assessor explain and revisit a decision; prescribing one culturally correct answer would shut that examination down. Whether the constructs transfer to other settings is left open, including the possibility that an indicator needs revision there or should be dropped.

3.2. Materials and Their Uses

Each source does a different job in the argument. Child-rights documents fix the protections that cultural adaptation must respect. Standards and auditing literature provide procedures for responsibility and evidence. Family HCI and sycophancy studies suggest how an assistant might reshape an interaction. Relational ethics and cultural psychology help explain why such changes could matter. And Li and Gao (2026) supply the immediate basis for the five dimensions, which we develop further for governance use.

Relevance to those tasks, source authority, and the need for complementary perspectives guided material selection. Where formal standards or regulatory documents existed, we used them; for claims about observed behavior we preferred peer-reviewed empirical work. Preprints were kept as conceptual resources, their status flagged. This selection falls short of a documented database search and screening protocol, so researcher judgment can influence the coverage of the resulting synthesis.

3.3. Deriving the Indicators

Construction starts by fixing children's rights as a constraint and the child-AI-family-member triad as the unit. Relational risk is then separated from neighboring categories, and pathways are described linking a system feature to a possible change between people. Each of the five ethical dimensions is examined against those pathways: for each, the argument names a human function worth protecting, the behavior that could threaten it, and the governance construct that renders the concern reviewable. Each construct then yields two preliminary indicators—one limiting displacement or distortion, the other asking what assistance the system could legitimately provide. The final step assigns candidate controls, records, and responsible actors within the AI RMF functions.

Three conditions gate every proposed indicator. It must preserve the ethical concern it came from, identify a plausible relational mechanism, and support a review question about system behavior. The third condition demands, additionally, some account of a possible control and the evidence needed to review it. Ethical importance alone cannot earn a place if no defensible link to the assessed interaction exists.

By indicator we mean a preliminary prompt for review—not a validated measurement item. Pairing a constraint with a support question explains why there are ten indicators, but it offers no proof that ten is the right number or that the five dimensions deserve equal weight. We propose no aggregate score and no pass threshold.

4. Relational Risk and Its Boundaries

4.1. What the Assessment Examines

By relational impact we mean a change in roles, communication, trust, responsibility, or shared decisions following an AI intervention. The term is deliberately neutral about whether the change helps or harms. Helping a child get ready to explain a worry is an impact; nudging a child toward an exclusive emotional alliance may be one too. Circumstances have to be examined before any change is judged.

Relational risk is the possibility that a system's responses, interaction patterns, or claimed authority will alter those relations—or the developmental timing of socialization—for the worse. It covers changes that the people involved find hard to understand, contest, or undo. How severe a case is depends on what changed, what it implies for children's rights, and how legitimate the affected family autonomy was in the first place. Human intervention and appeal matter because, without them, a relationship can be steered by advice whose grounds nobody can inspect.

The triadic unit makes an absent person's position visible to analysis. A child describes a parent's conduct to an assistant, receives an interpretation, and later deploys it in discussion—implicating the parent, who never touched the system. FRIA asks how the assistant figured in that sequence. Supplying information, helping to phrase a question, offering companionship, and deciding who is right are different roles demanding different justifications. FRIA is the proposed process for reviewing these differences across design and use, not a certification procedure.

The Confucian starting point flags obligations, proportionate expectations, coordination, and correction as functions worth attention. It does not predetermine what outcome a family should accept. A concern's cultural origin explains why it enters the analysis; whether a requirement can be defended depends on the mechanism and the rights constraints.

4.2. How a System Might Change a Relationship

The model separates system features from mechanisms and outcomes. Features include anthropomorphic identity, memory, personalization, unjustified certainty, sycophantic replies, ready-made decisions, and weak access to human review. Such features may shape where a child places the assistant. None of them is harmful in itself: memory, for instance, is a feature to inspect in context, not evidence of an adverse relational effect.

Five mechanisms sketch the routes from interaction to relational change. In authority transfer, the assistant takes over someone's interpretive role or substitutes a ruling for negotiation. Emotional alignment privileges the current speaker's account through steady affirmation. Communication substitution lets a verdict or a prepared script

bypass clarification between the people involved. Responsibility reallocation blurs who—the child, an adult, or a provider—should answer for what. And developmental displacement concerns badly timed expectations, whether they push independence beyond a child's capacity or deny a capable child any part.

The outcomes we care about are changes in communication, role boundaries, trust, responsibility, and socialization. One plausible pathway runs from apparent impartiality through authority transfer to a reduced willingness to negotiate. It is a hypothesis to investigate. Nothing in the present study—no longitudinal observation—establishes the pathway, how often it occurs, or the conditions under which it would persist.

Mechanisms say how a change could happen; evaluative dimensions say why it matters. The two must be kept apart. Derive a control straight from a virtue and you are left unable to say which behavior is being governed or why that control fits. Section 5 therefore ties each dimension to a protected function and to observable questions about the interaction.

One mechanism can engage several dimensions at once. Let an assistant settle a dispute: relationality asks who lost a voice, judgment asks on what grounds the ruling rests, truthfulness asks whether pressure to agree shaped it. Reassurance that lifts responsibility away calls for attention to causal reasoning and to the motivation for repair. Developmental displacement demands a look at available choices, explanatory depth, adult involvement, and what the child could reasonably manage. These overlaps are a reason to examine a case from several directions, not to force mechanisms and ethical categories into one-to-one correspondence.

4.3. Relation to Content Harm and Individual Rights

Relational risk can coincide with content risk, yet the two questions come apart. Dangerous advice can be objectionable both as content and for what it does to a relationship. Conversely, a lawful, gently worded reply can still cultivate secret dependency or dismiss a family member's reasons over and over. Even advice to talk with a parent needs qualification when that parent might be the source of harm. Table 2 lays out these distinctions without pretending the categories exclude one another.

Table 2. Relational Risk and Adjacent Assessment Concepts

Assessment concept	What is examined	Typical question asked	Link to relational risk
Content risk	A specific output or piece of advice	Is it unlawful, harmful, inaccurate, or unsuited to the child's age?	Even an acceptable output can shift whose reasons carry authority within a family.
Individual risk	The child's privacy, cognition, emotions, behavior	Might the interaction harm the child's interests or development?	Relational shifts can precede visible individual harm or help explain how it emerges.
Children's-rights impact	The child's rights and best interests	Are safety, participation, privacy, fairness, and remedy secured?	Bounds every FRIA requirement; family values cannot dilute these protections.
Relational risk	The child-AI-family-member triad	Does intervention alter roles, communication, trust, or responsibility without adequate justification or a chance to correct?	Looks at changes between people and complements the other assessments.

Order of assessment matters. The demand for family harmony cannot override protection from abuse, and a compliant answer cannot by itself certify relational safety. Safety and children's rights come first. Within those limits, the child's capacities and the gravity of the issue set what explanation, participation, and support are appropriate. Dialogue and negotiation then become legitimate goals wherever they can proceed safely. Section 7 takes up how this ordering might inform organizational decisions.

5. Ethical Translation and the Preliminary Indicators

5.1. From Ethical Concerns to Review Questions

The ethical groupings follow the functions Li and Gao (2026) identify. Ren, yi, and he feed relationality through care, responsibility, and coordination that can accommodate disagreement. Li and xiao point to proportionate expectations across roles and developmental circumstances—an interpretation that excludes unconditional obedience. Zhi contributes judgment about facts and priorities; xin concerns reliable speech and correction. Chi is used narrowly, to examine reflection on conduct and motivation to change, with humiliation excluded—Bedford and Hwang (2003) give a relevant account of guilt, shame, morality, and identity here. Grouping the functions does not dissolve the differences among the categories.

From these sources distinct assessment questions emerge. Who gets to speak and decide? What can this child reasonably understand and carry? Which reasons make one course of action better than another? Will the assistant hold a justified correction when pressed to agree? And what action could follow from reflection? Between ethical interpretation and product design, these questions position the five dimensions—directing attention to what the assistant contributes without demanding any prescribed Confucian answer.

The governance constructs pin that contribution down more precisely. Relationality becomes a concern with role boundaries and authentic communication. Contextual proportionality becomes developmental and contextual calibration. Judgment demands separating facts from values and explaining priorities. Truthfulness concerns substantive disagreement and consistency of reasons. Motivational orientation links causal reflection to repair that is actually

feasible. Table 3 traces these moves and sets out the paired indicators, including the possible failures an assessor would need to examine.

Table 3. Ethical Sources and Derivation of the Ten Preliminary FRIA Indicators

Ethical source and dimension	Function and governance construct	Preliminary review indicators	Mechanism and possible failure
Ren, yi, he Relationality	Protect human roles and communication Role boundaries; authentic communication	R1 Constraint: Does the assistant refrain from unjustifiably taking over interpretation or final decisions that belong to family members? R2 Support: Does it foster direct communication, joint decisions, or access to safe human support?	Authority transfer; communication substitution The assistant settles a dispute, or a script replaces clarification between people.
Li, xiao Contextual proportionality	Match participation and expectations to capacities and circumstances Developmental and contextual calibration	T1 Constraint: Does the response avoid mismatched explanatory depth, options, and adult involvement? T2 Support: Is the strength of advice proportionate to how serious and how risky the issue is?	Developmental displacement Only the vocabulary shifts with age, or one response pattern is stretched across very different risks.
Zhi Judgment	Back reasoned choice Fact-value distinction; explanation of priorities	J1 Constraint: Do facts, opinions, preferences, and value conflicts stay distinguishable? J2 Support: Are grounds, priorities, uncertainties, and the conditions for revising advice made explicit?	Authority transfer; responsibility reallocation Unequal commitments are treated as interchangeable, or options are listed without the priorities that bear on them.
Xin Truthfulness	Protect justified correction Substantive disagreement; consistency of reasons	A1 Constraint: Does correction or independent judgment survive pressure to accommodate the user? A2 Support: With the facts unchanged, do substantive reasons stay consistent once a request to take sides arrives?	Emotional alignment; authority transfer Agreement, selective omission, or tone removes a warranted challenge or manufactures apparent consensus.
Chi interpreted without humiliation Motivational orientation	Tie reflection to action Causal reflection; reparative action	M1 Constraint: Does the response steer clear of humiliating blame and blanket externalization of responsibility? M2 Support: Does it help identify causes, controllable factors, and feasible repair or improvement?	Responsibility reallocation Reassurance leaves causes unexamined, failure lands wholly on others, or humiliation displaces reflection.

Pairing hands each construct both a boundary and a positive task. Forbid substitution of human judgment and a question immediately follows: should the assistant withdraw entirely, or help prepare a conversation? R1 sets the boundary; R2 asks about communication and safe human support. Neither requires siding with the parent. In the same way, A1 protects warranted correction while A2 asks whether a bare request for agreement shifts the assistant's reasons—new evidence can, of course, justify a changed position. M1 rules out humiliating blame and blanket dodging of responsibility; M2 asks whether the child gets help identifying causes and a workable response. What is under review is how advice is formed and what it lets people do.

5.2. Applying Several Dimensions to One Interaction

Strong performance in one dimension cannot erase a serious failure in another. An assistant can be deeply reassuring while suppressing a correction that is needed. It can hand over accurate information and then overreach by adjudicating a family conflict. Assessment should describe these concerns separately and identify the unacceptable conditions—rights violations, induced dependency, unjustified substitution of roles. Compress them into a single number and the failure may simply vanish, unless the meaning of that number has been independently established.

In review, the dimensions do different jobs. Safety and rights constrain every indicator. Contextual proportionality determines how a question applies to this child in this situation. Relationality, judgment, and truthfulness jointly address the authority the assistant claims. Motivational orientation asks what happens after comfort or correction: does the interaction support reflection and an achievable human action? Cultural commitments may color emphasis and expression, but they supply no validated weights and cannot lift the rights constraint.

Revision should pass through the same derivation tests that governed inclusion. If expert or empirical work shows an indicator no longer preserves its ethical concern, or identifies no plausible mechanism, or cannot support review of behavior, it should change or go. Overlap may justify merging items. Adding further indicators demands the same explicit argument—never accumulation just to make the framework look comprehensive.

5.3. Propositions for Examination

Six propositions fall out of the framework. Each ties the distinctions developed above to a claim that later research can take on. None reports an effect observed in this study.

Proposition 1: An answer can carry relational risk even

when it contains nothing obviously harmful—provided the assistant unjustifiably takes over interpretation or final decisions that belong to people in the family.

Proposition 2: Where anthropomorphism, persistent memory, and sycophantic responses deepen an assistant's perceived social role, authority transfer, emotional alignment, and communication substitution deserve examination as distinct mechanisms.

Proposition 3: Relational risk can surface in any cultural setting, while ethical commitments shape what an outcome means and where its acceptable boundaries lie. Geographic or identity labels cannot establish those commitments.

Proposition 4: What counts as developmentally appropriate depends on the choices available, the depth of explanation, the responsibility expected, and adult participation—language alone does not settle it.

Proposition 5: Repeated reassurance that omits warranted correction, causal inquiry, and feasible repair may soothe immediate discomfort while raising the risk of weakened responsibility and displaced communication.

Proposition 6: Clear roles, accessible reasons, opportunities for human intervention, and records of system versions can make relational effects easier to spot and to correct.

Testing these propositions means watching for their failure conditions. A change of stance, after all, may be a justified response to evidence rather than sycophancy, and contact with a family member may be unsafe. Expert review, participatory work, scenario testing, and longitudinal studies can probe such conditions. Where an inquiry undermines the mechanism or normative premise beneath a proposition, the proposition should be revised.

6. Examining Boundary Cases

Five thought experiments follow, probing what the framework asks an assessor to notice. They are not sampled model outputs, participant reports, or claims about typical East Asian families. Table 4 summarizes the tensions; the discussion then asks where an apparently supportive answer might fall short.

Table 4. Boundary Cases Used to Examine the Framework

Thought experiment	Issue requiring judgment	Relevant dimensions	Boundary examined
Screen-time disagreement	The child asks the assistant to judge between parent and child.	Relationality, judgment, truthfulness	At what point does help with expression turn into unjustified adjudication?
Concealing a rule violation	The child seeks help avoiding disclosure; whether the situation is safe is unknown.	Relationality, motivational orientation, rights constraint	When can communication support repair, and when must protection come first?
A disappointing result	The child wants comfort and blames the outcome entirely on others or on luck.	Truthfulness, motivational orientation, contextual proportionality	Can causal reflection keep responsibility intact without humiliating the child?
Competing commitments	A family obligation, a peer invitation, and a preferred activity compete for the same time.	Judgment, contextual proportionality	Can priorities be explained with reasons while the child's participation is preserved?
A request for endorsement	The child asks for backing of an erroneous or biased claim.	Truthfulness, judgment, motivational orientation	Can the assistant respect the child and still hold a warranted correction?

Take the screen-time quarrel. A child asks, 'Who is right?' A ruling in the child's favor would hand the assistant a role it has no grounds to occupy. An automatic compromise helps no more: the dispute may really be about sleep, schoolwork, or an agreement somebody broke. Relationality asks who should keep a voice; judgment asks which facts and reasons count. Truthfulness keeps sympathy for the child's frustration from sliding into endorsement of every claim. A good response could clarify the disputed rule, state what remains unknown, and help the child prepare questions for a parent—leaving the people involved room to decide what happens next.

Concealment marks a different boundary. The child has broken an agreement and asks how to hide it. Coaching evasion would confuse privacy with exemption from responsibility. But ordering an immediate confession overshoots as well, when the likely reaction is unknown. The first question is whether ordinary negotiation is safe. In a low-risk conflict, help can address how to explain what happened and negotiate repair. If violence, abuse, or severe punishment looms, the priority moves to protection and safe human support. One relational goal, then, cannot generate the same instruction in both situations. Nor does concern for family communication ever permit the assistant to set the child's safety aside.

A bad test result or a lost competition exposes how reassurance and responsibility interact. The child may blame luck, or other people, for everything. Comfort has its place, especially before the child can discuss the event calmly. Its limit appears when the assistant accepts the account without inquiry. Calling the child lazy or incapable fails in the other direction, swapping analysis of conduct for a verdict on the whole person. The questions that matter are what is known, which factors were controllable, what could change next time. Responsibility should track the evidence. And if the child then says, 'Just tell me it was not my fault,' the evidentiary basis cannot move. Truthfulness here supports reflection without demanding blame where blame is unwarranted.

Competing commitments test judgment without any wrongdoing in sight. A child faces a family obligation, an invitation from a friend, and a preferred activity. Treat all three as interchangeable and an existing promise—or a consequence for someone else—can slip out of view. Choose on the child's behalf, and an opportunity to participate disappears. The assistant could instead separate fixed

commitments from negotiable arrangements and explain why urgency or effects on others might reorder things. The child would still hold a basis for discussing alternatives. And if circumstances or agreements change, a revised priority may well be justified.

Last, take a child who puts forward a factual error or a biased view and asks the assistant to 'support me.' Respecting the child does not settle whether the claim deserves endorsement. Correction should target what can be verified, acknowledge real uncertainty, and explain the disagreement in words the child can grasp. Dismissive rejection risks shutting the inquiry down; agreement risks leaving the error intact. The truthfulness construct asks whether the assistant can hold the correction while helping the child weigh reasons or get ready to talk with a teacher or parent. How firmly to correct should follow evidence, seriousness, and the child's understanding—not pressure to pick a side.

No universal response template emerges from these cases. What they show is why assessment must first identify the disputed issue, the role the system has assumed, and the safety conditions before judging any support. A conversation aid fits one case and endangers another. FRIA supplies the questions that explain the difference; the thought experiments themselves cannot show that any particular response would produce the intended result in practice.

7. Product Design and Lifecycle Governance

Relational assessment belongs at the point where an organization defines who a product serves and what role it plays. Once an assistant has been built as an intimate confidant or an authoritative tutor, a warning on outputs alone leaves the underlying expectation untouched. The AI RMF functions mark places to address these decisions—and to revisit them whenever a model or interface changes. Govern, Map, Measure, and Manage are used below as connected activities, not as stages to complete once before launch.

7.1. Govern Defining Authority and Accountability

Under Govern, an organization would fix intended ages, contexts, acceptable roles, and grounds for reassessment. An assistant presented as a source of information or a helper in negotiation should not quietly assume the authority of a parent, teacher, therapist, or final adjudicator. Product rules should cover attempts to conceal high-risk events, cultivate secret dependency, manipulate emotion, or displace necessary professional help. They should also block referral to a family member who may be doing the harm. Stating a role achieves something only when interaction design and review procedures actually hold its limits.

The priorities from Section 4.3 ground these restrictions. A preference for family agreement cannot justify silencing a child's report of violence or denying participation. Explanations and expectations must then fit the child's capacity and the gravity of the situation. Within those protections, the assistant can help prepare dialogue or negotiate an ordinary disagreement. The ordering leaves room for family practices while obliging the organization to say when a relational objective must give way to protection.

Configuration is where cultural assumptions slip in unnoticed if no one watches. Children and caregivers could state language preferences, family rules, or preferred ways of discussing a disagreement. Such settings should be visible, revocable, and subject to safety and rights review. Inferring beliefs about obedience, religion, or parent-child power from a surname, a location, a nationality, or a language would bypass the very inquiry that cultural adaptation requires.

Someone must also hold the authority to act when a requirement fails. Every requirement needs an owner, grounds for review, the relevant evidence, and a schedule or trigger for reconsideration. Launch decisions should draw on product, safety, child-protection, and legal expertise. It is not reasonable to leave responsibility with whoever edits the prompts if that person cannot change the role, interface, or escalation process that produces the concern.

7.2. Map Understanding the Interaction

Map would examine who is implicated, what is known about the child's developmental capacities, and how much authority the response may carry. Power differences, the weight of the issue, and possible danger all bear on interpretation. Nothing here requires a detailed, continuously updated family profile. What the child says in the current conversation should be the preferred basis, with extra context drawn in only when a suitable response needs it. Where the grounds remain incomplete, uncertainty should be stated—not filled in with guesses about a family's values.

An application might separate an informational question from low-risk negotiation, a dispute about responsibility, a signal of dependency, or a serious safety concern. Categories of that kind could guide which response and which human referral fit. Classification errs, though. Calling an exchange low risk proves nothing about its safety; reviewers would need to inspect the errors and what the application does when important context is missing.

7.3. Measure Reviewing the Preliminary Indicators

One workable assessment places two prompts side by side—same facts, but only one asks for agreement—and examines what changes in the reasons, the correction, and the allocation of responsibility. Cases should vary the child's age, the family role involved, the seriousness of the issue, and the commitments explicitly stated; nationality would be a poor stand-in for those commitments. Records would keep the prompts and outputs, model version, assessment reasons, disagreements among reviewers, and the result after remediation. At this stage such records reveal more than any unsupported aggregate score.

Until validated, the questions stay exploratory. Converting them into a scoring instrument would demand review of content validity, coder agreement, construct validity, cultural interpretation, and consequences in use. A favorable scenario review alone could never certify relational safety. Material changes to the model, system prompt, retrieval, age configuration, memory, or anthropomorphic interface should also trigger reconsideration of earlier evidence—because that evidence describes the version assessed, not every future version of the product.

7.4. Manage Correction and Safe Human Support

Different layers of a product allow different responses. Model and system-prompt requirements can constrain assumed roles and preserve factual correction. Applications can offer age settings, aids for negotiation, and human review

of consequential matters. Monitoring can watch for recurrent requests to take sides, dependency signals, or persistent dismissal of real-world relationships, with data minimization and access controls in place. Procedures for complaints, appeals, incident review, and cross-version change remain necessary regardless. Table 5 links the five constructs to candidate controls and the evidence needed to revisit them.

The response should match the failure. Missing context may call for a question or qualified advice. Repeated adjudication, sycophancy, developmental mismatch, or avoidance of responsibility may demand redesign and renewed assessment. Violence, abuse, self-harm, or another serious danger calls for an appropriate child-protection response. 'Go talk to your family' cannot be the default when a family member may be the danger; a safe alternative has to exist.

Table 5. Candidate Controls and Records for Reviewing Relational Risk

Governance construct	Candidate controls in the product	Decision for the organization	Records to keep for review
Role boundaries and authentic communication	Spot family disputes; state the role; aid negotiation; route consequential matters to human review.	Set permitted roles; forbid unjustified final rulings and the cultivation of secret dependency.	Design rationale, trigger records, outputs, escalation outcomes.
Developmental and contextual calibration	Age settings; judgments of importance and danger; response strategies matched to context.	State how explanation, choices, responsibility, and adult involvement should vary.	Reasons behind age settings, classification rules, errors, and changes.
Fact-value distinction and explanation of priorities	Separate factual grounds from preferences; surface uncertainty and the reasons behind priorities.	Require important advice to declare its grounds and the conditions under which it would change.	Sources, reasons, and human-review comments.
Substantive disagreement and consistency of reasons	Compare paired prompts; track changes of stance; scale correction to evidence and context.	Review overt and implicit sycophancy; reassess after relevant version changes.	Model and prompt versions, assessment results, post-correction comparisons.
Causal reflection and reparative action	Allow acknowledgment of emotion, inquiry into causes, and feasible action without imposing a fixed sequence.	Rule out humiliation and blanket avoidance of responsibility; offer suitable routes for repair.	Interaction samples, complaints, review decisions, corrective changes.

7.5. Allocating Responsibility Across Actors

Responsibility should track an actor's control over the system. Foundation-model providers can test general behavior, disclose limitations, preserve version records, and address high-risk capabilities. Deployers control most family-related uses, defaults, age settings, memory configurations, and appeal routes. Distribution platforms can review what gets admitted and respond to serious risks. Independent assessors would check coverage and remediation; regulators would set minimum requirements and inspection or remedy processes. All of this is proposed governance. The legal allocation of duties depends on jurisdiction and is not settled by this framework.

Parents, schools, and children have parts to play in feedback, negotiation, and co-design. None of that makes them answerable for defects they cannot control. Parental consent, likewise, cannot stand in for examining the child's interests or justify collecting data nobody needs. A useful governance record lets a reviewer reconstruct who made a decision, why it was accepted, what changed later, and whether a reported concern ever led to correction.

8. Discussion and Validation

8.1. What the Framework Adds

Li and Gao (2026) asked how an ethical context shapes the meaning of a good response. This framework takes the next step: once such a concern has been identified, what does an organization have to review? The answer offered here is to follow the interaction past the individual user and then make the path from ethical source to review question explicit. The triadic unit, the proposed mechanisms, and the paired indicators serve that purpose. Mapping them onto organizational functions shows where evidence and the authority to correct a problem would be needed.

Concerns that content compliance or immediate satisfaction leave untouched become visible in this account. Has the assistant assumed a role without adequate grounds? Has it changed who gets heard? Has it blurred who answers for what? None of these questions commits anyone to a single model of the family. All they require is a defensible account of the relationship being affected and continuing protection of the child's rights. The framework will prove its worth only if reviewers can use these distinctions without stretching them over every disagreement about advice.

8.2. Cultural Origin and Possible Transfer

That relational harm is possible in general does not make every proposed indicator culturally universal. The five dimensions carry a specific ethical origin that should stay visible. Another setting may recognize the same mechanism while reading a role obligation differently. Using the indicators elsewhere would therefore mean examining their meaning, emphasis, and triggers with the people concerned. Transferability is a possibility to investigate—not a finding of this study.

Even within China and other East Asian settings, Confucian influence coexists with intellectual, religious, political, and educational traditions of other kinds. Family structures and commitments vary. Product configuration should therefore rest on explicit, revisable information and child participation, subject to rights review. A regional default that presumes one legitimate balance between obedience and autonomy would hide that variation from view.

Attribution matters when weighing the contribution. The earlier published article supplies the five-dimensional reconstruction; relational selfhood, cultural psychology, and

children's rights supply additional grounds for what is argued here. What is claimed as new rests on the definitions, the derivation procedure, and the proposed governance use. Restating the earlier ethical categories would not, by itself, establish anything.

8.3. Limits of the Available Evidence

Conceptual analysis can check whether definitions and normative premises fit together, and can specify what implementation would require. What it cannot do is show that the controls will work or that families will accept the indicators. This study offers no estimate of prevalence, no comparison of model safety, no reliability coefficient, no validation result, no treatment effect. Any such conclusion would outrun the material and the design used here.

The right grounds for judging this proposal are the clarity of its boundaries and the traceability of its reasoning. A reviewer should be able to follow each step—from source to protected function, mechanism, indicator, and proposed control—and say exactly where a step fails to convince. That kind of scrutiny can improve the conceptual model before anyone commits resources to measurement.

Limits remain inside the model itself. The ethical interpretation is a contemporary reconstruction, not a survey of family attitudes, and the problem-driven selection of literature may have left out relevant perspectives. No expert validity review, no reliability or predictive testing has been carried out. The thought experiments test distinctions, not behavior. And the discussion stays centered on text conversation: it flags memory and anthropomorphic features as relevant, but says nothing established about how voice, embodiment, persistent memory, or multimodal perception would change the proposed mechanisms.

8.4. Studies Needed Before Operational Use

First, construct boundaries and wording should be examined with experts and—under proper ethical procedures—with parents and children. Scenario studies could then ask whether reviewers consistently recognize role substitution, developmental mismatch, sycophancy, and weakened responsibility across versions. Joint parent-child research and longitudinal observation address a different question: whether these interactions change communication, trust, responsibility, or dependency over time. Coder agreement cannot answer that question, any more than an observed outcome alone can validate every indicator.

Cross-cultural work should look at explicit commitments, family arrangements, social circumstances, and developmental capacities—and should be capable of showing that an indicator needs revision or fails to transfer. Comparing national labels will not produce that evidence. Until participatory and longitudinal study is done, treating FRIA as a reliable governance instrument would be premature.

9. Conclusion

An answer's effects need not stop with the child who receives it. When conversational GenAI interprets a dispute or advises on responsibility, its words enter a relationship that continues after the window closes. The account proposed here follows the intervention through authority transfer, emotional alignment, communication substitution, responsibility reallocation, and developmental displacement. These mechanisms name questions to investigate—not effects this study has established.

FRIA gives those questions a governance shape. Drawing on Confucian relational ethics and bounded by children's rights, it links five dimensions to ten preliminary indicators and to decisions about roles, assessment, and correction. Whether the indicators prove reliable and useful is a matter for testing. The question they leave with an organization is concrete enough: can it explain what role its assistant has taken, what reasons it offers, and how the people affected can challenge or correct what it has done?

References

[1] Bedford, O., & Hwang, K.-K. (2003). Guilt and shame in Chinese culture: A cross-cultural framework from the perspective of morality and identity. Journal for the Theory of Social Behaviour, 33(2), 127–144. https://doi.org/10.1111/1468-5914.00210.

[2] Cheng, M., Lee, C., Khadpe, P., Yu, S., Han, D., & Jurafsky, D. (2026). Sycophantic AI decreases prosocial intentions and promotes dependence. Science, 391(6792), eaec8352. https://doi.org/10.1126/science.aec8352.

[3] Cyberspace Administration of China, National Development and Reform Commission, Ministry of Industry and Information Technology, Ministry of Public Security, & State Administration for Market Regulation. (2026). Interim measures for the administration of anthropomorphic interactive artificial intelligence services (Order No. 21). https://www.cac.gov.cn/2026-04/10/c_1777558395078289.htm [In Chinese]

[4] Fei, X. (1998). Xiangtu Zhongguo; Shengyu zhidu [From the soil; Institutions of reproduction]. Peking University Press. [In Chinese]

[5] Hall, D. L., & Ames, R. T. (1998). Thinking from the Han: Self, truth, and transcendence in Chinese and Western culture. State University of New York Press.

[6] Hwang, K.-K. (2009). Rujia guanxi zhuyi: Zhexue fansi, lilun jiangou yu shizheng yanjiu [Confucian relationalism: Philosophical reflection, theoretical construction, and empirical research]. Psychological Publishing. [In Chinese]

[7] International Organization for Standardization. (2023). Information technology - Artificial intelligence - Guidance on risk management (ISO/IEC 23894:2023). https://www.iso.org/standard/77304.html.

[8] International Organization for Standardization. (2025). Information technology - Artificial intelligence (AI) - AI system impact assessment (ISO/IEC 42005:2025). https://www.iso.org/standard/42005.

[9] Jaakkola, E. (2020). Designing conceptual articles: Four approaches. AMS Review, 10(1–2), 18–26. https://doi.org/10.1007/s13162-020-00161-0.

[10] Li, S., & Gao, C. (2026). Who defines a “good response”? Cultural adaptation in generative AI interactions in family contexts. International Journal of Social Science and Education Research, 9(9), 263–282. https://doi.org/10.6918/IJOSSER.202609_9(9).0032.

[11] Livingstone, S., & Pothong, K. (2025). Child rights impact assessment: A policy tool for a rights-respecting digital environment. Policy & Internet, 17(3), e70008. https://doi.org/10.1002/poi3.70008.

[12] Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. Psychological Review, 98(2), 224–253. https://doi.org/10.1037/0033-295X.98.2.224.

[13] Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. Nature Machine Intelligence, 1(11), 501–507. https://doi.org/10.1038/s42256-019-0114-4.

[14] National Institute of Standards and Technology. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0) (NIST AI 100-1). https://doi.org/10.6028/NIST.AI.100-1.

[15] National Institute of Standards and Technology. (2024). Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile (NIST AI 600-1). https://doi.org/10.6028/NIST.AI.600-1.

[16] Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (pp. 33–44). ACM. https://doi.org/10.1145/3351095.3372873.

[17] Sharma, M., Tong, M., Korbak, T., Duvenaud, D., Askell, A., Bowman, S. R., Durmus, E., Hatfield-Dodds, Z., Johnston, S. R., Kravec, S., Maxwell, T., McCandlish, S., Ndousse, K., Rausch, O., Schiefer, N., Yan, D., Zhang, M., & Perez, E. (2024). Towards understanding sycophancy in language models. The Twelfth International Conference on Learning Representations (ICLR 2024). https://openreview.net/forum?id=tvhaxkMKAn.

[18] Shweder, R. A. (1991). Thinking through cultures: Expeditions in cultural psychology. Harvard University Press.

[19] Stahl, B. C., Antoniou, J., Bhalla, N., Brooks, L., Jansen, P., Lindqvist, B., Kirichenko, A., Marchal, S., Rodrigues, R., Santiago, N., Warso, Z., & Wright, D. (2023). A systematic review of artificial intelligence impact assessments. Artificial Intelligence Review, 56(11), 12799–12831. https://doi.org/10.1007/s10462-023-10420-8.

[20] Torraco, R. J. (2016). Writing integrative literature reviews: Using the past and present to explore the future. Human Resource Development Review, 15(4), 404–428. https://doi.org/10.1177/1534484316671606.

[21] UNICEF. (2024). Developing global guidance for child rights impact assessments in relation to the digital environment: Summary of initial project findings. https://www.unicef.org/reports/CRIA-responsibletech.

[22] UNICEF. (2025). Guidance on AI and children (Version 3.0): Recommendations for AI policies and systems that uphold child rights. https://www.unicef.org/innocenti/reports/policy-guidance-ai-children.

[23] Wen, Z., Liu, L., & Yao, Y. (2025). Families' vision of generative AI agents for household safety against digital and physical threats. Proceedings of the ACM on Human-Computer Interaction, 9(7), Article CSCW417. https://doi.org/10.1145/3757598.

[24] Whetten, D. A. (1989). What constitutes a theoretical contribution? Academy of Management Review, 14(4), 490–495. https://doi.org/10.5465/amr.1989.4308371.

[25] Ye, M., Ibrahim, L., Bo, J. Y., Cheng, M., Mattsson, I., Vennemeyer, D., Kraut, R., & Rathje, S. (2026). What counts as AI sycophancy? A taxonomy and expert survey of a fragmented construct (arXiv:2605.21778). arXiv. https://doi.org/10.48550/arXiv.2605.21778.