



Designed Differently

How MiFamilies Engineers Out the Documented
Harm Pathways of Social Media

By MiFamilies · Working paper · August 2026

*A sister paper to **Social Media Impact on Children & Adolescents —
Baseline Reference Compendium** (v0.2, August 2026)*

Closed trust bubbles · Family vouches you in · Guardians govern, not platforms

mifamilies.com

Contents

1. What this paper argues — and what it deliberately does not	2
2. Starting from the evidence, not against it	4
3. The harm pathways, one by one — and the design that removes each	5
4. The compendium’s ten lessons, answered	8
5. The business model is the safety model	9
6. The exemption case	10
7. Demonstrable proof points — the one-hour audit	11
8. Honest limitations and standing commitments	13

Status and posture. *This paper is a companion to the evidence compendium, and it accepts that document’s discipline: it does not overclaim, it does not relitigate contested science, and it distinguishes what is demonstrated from what is designed. Its argument is made at the level the compendium itself identifies as tractable — design features, not screen time — and every claim in Section 3 onward is a statement about the shipped product that an independent assessor can verify by using it. Factual claims reflect the platform as of August 2026. Any use of this paper in a formal regulatory submission should first be reviewed by counsel.*

1. What this paper argues — and what it deliberately does not

The reference compendium closes twenty years of evidence with a synthesis that regulators, courts and researchers now broadly share: the harms of social media to children are **heterogeneous** (a vulnerable minority is meaningfully harmed while averages move little), they travel through **identifiable design features** rather than raw hours, and the business model of engagement-optimised attention is the **root design constraint** from which those features grow.

This paper argues three things:

1. **Mitigation.** Every one of the least-contested harm pathways named in the compendium — sleep displacement, appearance-based social comparison, cyberbullying victimisation, algorithmic amplification, stranger contact, and problematic/addiction-like use patterns — corresponds to a specific design decision in MiFamilias that removes the pathway itself, not merely moderates its output.
2. **Difference.** These are not settings bolted onto an engagement platform under regulatory pressure — the pattern the compendium documents across incumbent platforms (“changes under pressure”). They are the platform’s founding architecture, present because the product’s revenue does not depend on attention time.
3. **Exemption.** Under the statutory logic already in force — Australia’s category exclusions, the UK’s messaging carve-out, and above all Canada’s Bill C-34 conditional model, which exempts platforms demonstrating *adequate safeguards* — MiFamilias’s minor experience belongs with the excluded category: it is a closed, verified, family-communication space with no strangers, no public surface, and no engagement machinery. A blanket ban that removes it does not reduce a child’s exposure to the documented harm mechanisms; it removes the safe alternative that mitigates the displacement problem regulators themselves worry about.

What this paper does **not** argue: it does not claim MiFamilias improves adolescent mental-health outcomes (no platform can claim that without longitudinal outcome studies that do not yet exist for any platform); it does not dispute the harm literature; and it does not claim perfection. It claims something narrower and fully checkable: **the named mechanisms of harm are absent by design, and their absence is demonstrable.**

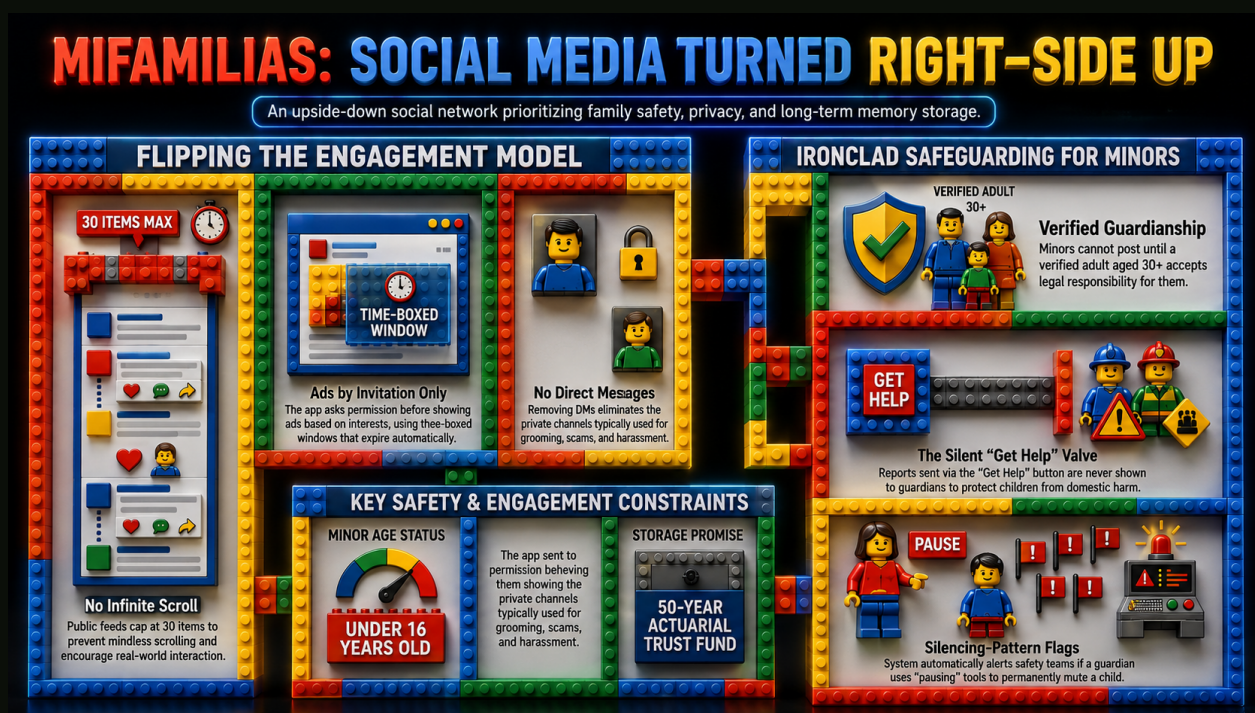


Figure 1: MiFamilias: social media turned right-side up — flipping the engagement model.

2. Starting from the evidence, not against it

The compendium's honest one-page summary identifies the emerging consensus position — the conditional/heterogeneity thesis, with the National Academies' 2024 report as its most citable statement: no population-level causal conclusion, *real mechanisms*, and design standards warranted regardless.

MiFamilias's position is that this consensus, taken seriously, is a **design brief**. The precautionary asymmetry the compendium describes — engagement-optimised products deployed at scale to a neurodevelopmentally vulnerable population without safety testing — is an argument about *products that were never designed for children's safety being retrofitted*. The correct response to the precautionary argument is not only to restrict retrofitted products; it is to build products where the exposure the developing brain is most vulnerable to — engineered social reward, engineered social comparison, engineered variable-reward checking — is not present to be dosed.

The established neurodevelopmental facts (the compendium's Layer A) describe an adolescent brain with early-maturing reward circuitry, late-maturing impulse control, and peak sensitivity to social evaluation. Every design decision in Section 3 can be read against that profile: MiFamilias removes precisely the mechanisms that press on those three vulnerabilities.

3. The harm pathways, one by one — and the design that removes each

The compendium identifies a “defensible core” of harm pathways with the strongest evidence. Each is taken in turn. For each: the mechanism as it operates on incumbent platforms, the MiFamilias design fact, and how an assessor can verify it (expanded in Section 7).

3.1 Sleep displacement

The strongest-evidenced pathway (Carter 2016; Lima Santos 2025 identifies sleep as the strongest mechanistic route).

Incumbent mechanism: late-night notifications, autoplaying feeds, and variable-reward checking that pull the device back into the bedroom.

MiFamilias design facts:

- Notifications are **batched by default** and honour user-configurable quiet hours, including overnight windows; nothing “re-engages” a user at night. The single exception is deliberate and family-scoped: a family safety alert bypasses batching — the one category of message a family would want to wake for.
- There are **no engagement notifications**. Notifications are events (a capsule unlocked, a family member posted an alert), not prompts manufactured to trigger a return visit.
- There are **no streaks, no daily-return mechanics, and no push-based re-engagement loops** anywhere in the product. This is written product doctrine: a feature may give a reason to *return* or to *invite*, never a reason to *scroll*.

3.2 Appearance-based social comparison

Kelly 2018; Saiphoo & Vahedi 2019; the Meta internal research — stronger for photo-based, appearance-focused use than for time spent.

Incumbent mechanism: appearance-ranked public audiences, filters, and quantified popularity applied to the face.

MiFamilias design facts:

- A minor’s audience is their **family, full stop**. Minors have no public surface: no public profile, no follower counts, no public feed presence, no community access. There is no audience of strangers in front of whom to be ranked.
- Under-18 accounts **cannot set a profile photo at all** — the server rejects it; minors are represented by a generated initials avatar. The platform’s most appearance-loaded surface simply does not exist for children.
- There are **no beauty filters and no appearance-altering effects** in the product.
- In adult communities, identity is **pseudonymous by architecture**: a curated avatar icon and a generated handle. Real names and photos never appear — so the comparison engine has nothing to grip even for adults.

3.3 Cyberbullying and hostile contact

Kowalski 2014 — cybervictimisation robustly linked to depression, anxiety and suicidal ideation.

Incumbent mechanism: open contact graphs, anonymous or throwaway identities, and DMs that put any stranger one tap from any child.

MiFamilias design facts:

- **There is no direct-messaging feature on the platform.** Not for minors — for anyone. Family communication happens in the shared family feed; community discussion happens in the open community feed and threads. There is no private channel in which grooming or bullying can operate unseen.

- Family membership is **by invitation and consensus**: new members are vouched for by existing members, and a minor cannot join any family without the consent of a **confirmed legal guardian** — a verified parent/guardian relationship, not merely any adult in the family.
- A minor **cannot post at all** until a confirmed guardian has formally accepted responsibility for their participation — so every child account has an accountable adult attached by construction.
- Reporting is built into every surface, review is human, and young members have a dedicated **Get help** channel whose reports are never visible to — and never notify — the guardian, because one of its categories is “I don’t feel safe at home.” Safeguarding access on the platform side is restricted to a dedicated role and every access is audit-logged.

3.4 Algorithmic amplification and rabbit holes

The pathway at the centre of the litigation shift (Anderson v. TikTok) and the EU enforcement record; CCDH’s 2.6-minutes-to-suicide-content finding.

Incumbent mechanism: engagement-optimised recommender systems that curate a personalised content stream and learn to amplify whatever maximises dwell time — including harm-adjacent content.

MiFamilias design facts:

- **There is no content recommendation engine.** Feeds are chronological. Two members of the same family see the same family feed. Nothing is ranked by predicted engagement, because no engagement prediction exists.
- **Feeds end.** The public (Legends) feed is hard-capped per session and terminates in an explicit “You’re all caught up” state; community feeds are capped with a deliberate “show earlier posts” step. The family feed is bounded by the only thing that should bound it — how much your family actually posted. There is no infinite scroll to fall into and no autoplay stream.
- Minors never see the public feed at all: the child experience is **family-only**.

3.5 Stranger contact and predation

Béjar testimony (1 in 8 under-16s reporting unwanted advances weekly on an incumbent platform); the grooming litigation against messaging and gaming platforms.

Incumbent mechanism: discoverable child profiles, open DMs, and recommendation systems that surface children to adults.

MiFamilias design facts:

- A minor on MiFamilias is **undiscoverable**. No public profile, no search surface, no community presence, no follower mechanics, and no DMs mean there is no route by which a stranger can find, view, or contact a child. The contact graph of a child is their invited, vouched, guardian-consented family.
- Adult community spaces are **hard-gated at 18+ with verified age** — not self-declared age. A community invite link opened by a minor is inert: redemption re-runs the age gate every time, so a forwarded link confers nothing.
- Age is taken seriously as infrastructure (the compendium’s L4): date of birth is **write-once**, adult checks **fail closed** (no verified date of birth means no adult access, never the reverse), and verification uses document or facial methods, not a checkbox.

3.6 Problematic and addiction-like use

Boer 2020; WHO/HBSC 2024 — the distinct minority pattern of compulsive use, driven by variable-reward design.

Incumbent mechanism: infinite scroll, autoplay, streaks, quantified popularity, and intermittent-reinforcement notification design — the “slot machine” pattern the litigation now names as design defect.

MiFamilias design facts:

- Of the design-defect feature list that appears in the compendium's litigation section — *infinite scroll, autoplay, streaks, quantified popularity, late-night notifications, appearance filters, stranger DMs* — **none is present on the minor experience, and most exist nowhere in the product for anyone.**
- The platform's signature feature runs in the opposite temporal direction from compulsion: a **time capsule** is content that deliberately cannot be consumed now. The product's emotional core rewards patience measured in years, not checking measured in minutes.
- Return features (a memory resurfacing on its anniversary, a weekly family quiz) are **pull-only**: they appear when the user arrives; they never call the user back.

MiFamilias: Your Family Legacy, Protected and Preserved

THE DAILY CIRCLE (SAFE & IMMEDIATE)

- THE PRIVATE FAMILY BUBBLE**
A closed circle where content never leaves and entry is strictly by **invitation only**.
- INSTANT FAMILY ALERTS**
Urgent safety alerts that bypass all "quiet hours" to reach every family member **immediately**.
- HUMAN-FIRST MODERATION**
No "doom-scrolling" algorithms; feeds are capped and reporting is handled by **humans, not bots**.

THE LONG-TERM PROMISE (LEGACY & FUTURE)

- UNBREAKABLE TIME CAPSULES**
Video messages locked for up to **100 years**, guaranteed by a dedicated storage trust.
- THE FAMILY SAGE**
An AI presence built from your **life stories** that your family can talk to later.
- THE STORAGE TRUST FUND**
Audited funds reserve the actual cost of long-term storage to keep promises for **50+ years**.

MiFamilias | Closed trust bubbles | Family vouches you in | Guardians govern, not platforms

Figure 2: The daily circle and the long-term promise: the family bubble, human-first moderation, time capsules, the Sage, and the storage trust fund.

4. The compendium's ten lessons, answered

The compendium closes with ten lessons (L1-L10) that “a 2.0 platform” should learn. MiFamilias was built to them. In brief:

Lesson	MiFamilias answer
L1 — The business model is the root design constraint; a 2.0 platform must decouple revenue from time-on-app	Revenue is purchases, subscriptions and invited advertising (§5). No engagement-optimised ranking exists to monetise attention.
L2 — Protective defaults must be sized for the vulnerable tail	The minor experience is not a “teen mode” of an adult product; it is a categorically smaller surface — family-only, guardian-anchored, undiscoverable — applied to every minor because the vulnerable tail cannot be identified in advance.
L3 — Design features, not screen time, are the tractable unit	Section 3 is organised feature-by-feature for exactly this reason; the named defect features are absent.
L4 — Age assurance is mandatory but unsolved; treat it as core infrastructure	Write-once DOB, fail-closed adult gates, document/facial verification for adult spaces, re-checked on every community entry path including forwarded invite links.
L5 — Document safety decisions as if they will be discovery exhibits	Safety-relevant actions are audit-logged (including every safeguarding read, the platform owner's included); consent is versioned and recorded; moderation decisions carry restore levers and human review.
L6 — Bans are politically irresistible but empirically unproven; displacement is the counter-risk	This is the exemption argument (§6): removing the safe, verified, family-anchored alternative worsens displacement rather than reducing exposure.
L7 — Policy ran ahead of causal certainty, on precaution	Accepted — and answered by building to the precautionary standard rather than contesting it (§2).
L8 — The next fight is AI companions and generative content	MiFamilias's AI (the Sage) is the anti-companion: consent-first, built by a person about their own life for their own family, no romantic or parasocial dynamics, minors cannot create profiles (a guardian may, in exceptional circumstances such as terminal illness, with ownership transferring at adulthood), activation is protected by named trusted contacts and a cancellable grace period. It is a family legacy instrument, not an engagement surface.
L9 — The child has enforceable rights against every adult party, including their parents	Guardian consent is required <i>and bounded</i> : safeguarding reports are invisible to guardians by design; a minor's images are protected even from the platform's own surfaces (no minor profile photos); deletion and erasure paths exist.
L10 — International regulatory divergence is a design parameter	The architecture clears the strictest bar in each regime simultaneously — because family-only childhood, verified adulthood, and no engagement machinery satisfy AU, UK, EU and the US feature-litigation standard at once, no regionalised childhood is required.

5. The business model is the safety model

The compendium's deepest finding is economic: every documented harm feature exists because engagement-ranked attention is what incumbent platforms sell. Internal documents, the algorithmic-amplification studies, and every regulator's remedy converge on the same point — and its L1 concludes that a 2.0 platform must decouple revenue from raw time-on-app *or regulation will do it forcibly*.

MiFamilias is already decoupled, structurally:

- **What the platform sells** is storage and services with intrinsic value delivered over time — time capsules, subscriptions, legacy features — plus advertising that operates by invitation.
- **Advertising is ask-first.** The platform builds no interest profile of anyone. When a member's own posts suggest an interest, the platform *asks* whether they would like relevant offers, for a time window the member chooses and can end early. Guardian controls bound what any minor can be shown, and no minor is ever profiled — which is not a compliance posture to the DSA's no-profiling rule for minors but the platform's universal mechanic. In communities, members collectively vote irrelevant advertising out of their own space.
- **The consequence is alignment:** a platform whose revenue does not rise with scroll time has no economic reason to build the features the litigation calls defective — and MiFamilias has not built them. This is the difference between “changes under pressure” (the compendium's phrase for incumbent remediation) and *designed differently*.

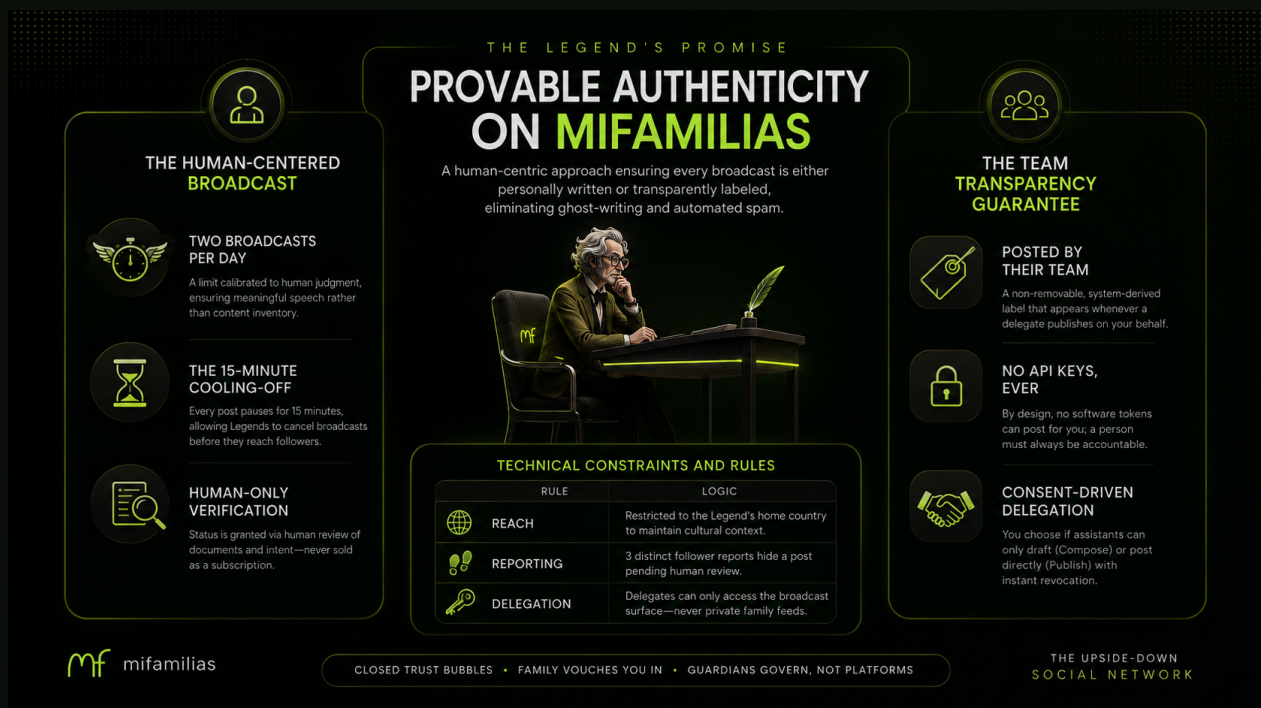


Figure 3: Provable authenticity on MiFamilias: the Legends broadcast surface — two broadcasts per day, cooling-off, human-only verification, no API keys.

6. The exemption case

The regulatory record the compendium surveys already contains the logic under which MiFamilias should be treated differently. Three converging precedents:

1. Category exclusions already exist — because purpose matters. Australia’s SMMA excludes messaging, gaming, education and health platforms; the UK’s incoming regime excludes messaging apps; YouTube Kids and comparable purpose-built children’s services sit outside scope. The excluded category is defined by what those services *are not*: they are not open social networks with public audiences, algorithmic feeds, and stranger contact. Judged by those criteria, the MiFamilias minor experience — a closed, invitation-only, guardian-consented family circle, with no public surface, no recommender, no strangers, and no DMs — is functionally a family-communication service, not “social media” as the harms record defines it. It shares a statutory category with the family group chat, not with the platforms the evidence indicts; unlike the group chat, it adds verified identity, guardian accountability, and safeguarding infrastructure.

2. The conditional-exemption model asks precisely our question. Canada’s Bill C-34 exempts platforms that demonstrate *adequate safeguards* — the first statutory attempt to define safe-by-design, and, as the compendium notes, a design-incentive model distinct from blanket bans. Against the feature list that recurs across C-34’s logic, the design-defect litigation, and the EU’s minor-protection guidelines, MiFamilias’s position is not partial compliance but categorical absence: no engagement-optimised feed, no infinite scroll, no autoplay, no streaks, no stranger DMs, no late-night re-engagement, no appearance filters, no minor discoverability, and age assurance run as fail-closed infrastructure. If “adequate safeguards” means anything, a platform with nothing on the defect list qualifies; if it does not qualify, the exemption category is empty and the incentive model collapses.

3. Blanket application produces the harm regulators are trying to avoid. The compendium is honest that bans are empirically unproven and that the documented counter-risks are displacement to less-regulated spaces and the isolation of vulnerable youth — with Australia’s own early data showing circumvention near 70% and no measurable decline in harm complaints in the first three months. A child banned from a verified, guardian-anchored family space does not stop communicating; they are displaced to exactly the unverified group chats and less-regulated apps the evidence warns about — spaces with no age assurance, no safeguarding channel, no accountable guardian, and no audit trail. Exempting the safe alternative is not a loophole in a precautionary regime; it is the precautionary regime working — giving families somewhere protective to *be*, which is the stated purpose of the entire policy effort.

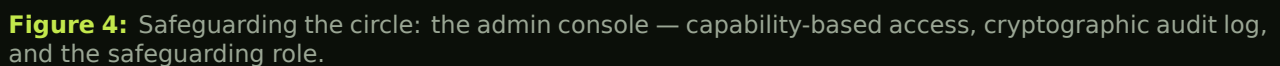
The ask, stated plainly: assessment under the same criteria the statutes already use — purpose, audience, feature set, and demonstrated safeguards — rather than by the label “social media.” Under those criteria, on the evidence in this paper and verifiable in the product, MiFamilias’s minor experience merits the treatment the excluded and exempted categories already receive.

7. Demonstrable proof points — the one-hour audit

Every claim above is checkable in the shipped product. MiFamilias offers regulators and assessors supervised test accounts and will support any of the following verifications; none requires trust in this document.

1. **Try to send a private message.** To anyone. There is no mechanism — no DM surface exists on the platform. (§3.3)
2. **Create a minor test account.** Observe: family-only experience; no public feed access; no community entry point; generated initials avatar with profile-photo upload refused by the server; posting blocked until a confirmed legal guardian accepts responsibility. (§3.2, 3.3, 3.5)
3. **Attempt discovery of that minor from a second, unrelated account.** No search result, no profile, no follow path, no contact route exists. (§3.5)
4. **Open a community invite link as the minor.** The link is inert — the age gate re-runs at redemption, defeating link-forwarding. (§3.5)
5. **Scroll the public feed as an adult.** It terminates in an explicit “You’re all caught up” state at the session cap; no infinite scroll, no autoplay. Two accounts in one family see the same chronological family feed — no personalised ranking. (§3.4)
6. **Check the notification pattern overnight.** Batched delivery, quiet hours honoured, no re-engagement pushes; only a family safety alert breaks through. (§3.1)
7. **Trigger the advertising consent loop.** Post about a topic repeatedly; the platform asks before any interest-based advertising occurs, with a member-chosen expiry — and never asks a minor outside guardian-set bounds. (§5)
8. **Attempt to open a time capsule before its date.** No client or support path exists; release is server-enforced. (§3.6)
9. **Audit the audit trail.** Request the access log for any safeguarding case: every read is recorded, including platform leadership’s. Consent records are versioned. (§4, L5)
10. **Review the automated evidence.** Every safeguard above is covered by behavioural test suites executed against the real handlers on every change — the guardian gates, the age gates, the invite-link re-check, the no-DM architecture, the feed caps, the capsule lock. The suites and their results can be demonstrated live.

Safeguarding the Circle: The MiFamilias Admin Console Journey



8. Honest limitations and standing commitments

In the spirit of the compendium's non-partisan discipline:

- **We do not claim outcome evidence.** No mental-health outcome studies of MiFamilias exist; the platform is young and its pilot population small. Our claims are design-level and mechanistic: the documented harm pathways are absent, verifiably. We commit to supporting independent research access as the platform grows, because the compendium is right that researcher access is part of adequate safeguarding.
- **Age assurance is unsolved industry-wide, and we inherit that.** Fail-closed gates and verified adulthood raise the floor substantially above self-declaration, but no assurance technology is perfect. We treat it as core infrastructure under continuous improvement, not a solved problem.
- **Scale changes things.** Safeguards that hold at pilot scale must be re-proven at each order of magnitude. Our commitment — aligned with the compendium's L5 — is that safety decisions are documented, auditable, and made as if they will one day be read aloud in a hearing. We would welcome that reading.
- **The standing offer.** Test accounts, supervised audits, documentation of the safeguards described here, and participation in any regulator's safe-by-design assessment framework, at any time.

Prepared by MiFamilias as a working companion to the Baseline Reference Compendium (v0.2). Platform facts reflect the shipped product as of August 2026. This document is intended for discussion and assessment; formal regulatory submissions derived from it should be reviewed by counsel.