When Stability Is Not Generativity: AI-Mediated Intimacy as a Coupling Sector of an Effective Relational Field Theory

Wanhong HUANG
huangwanhong@serendip.ngo

Abstract

This position paper develops AI-mediated intimacy inside an effective relational field theory derived from a background-independent substrate. Its central distinction concerns short-horizon stabilization and longer-horizon relational capacity, and the field-theoretic register makes that distinction structural: the quadratic part of the effective action fixes the relaxation spectrum, while routing, reachability, and generative capacity live in the vertices, and these are independent data. Four conditional results follow. A strictly triangular cross-sector self-energy leaves the propagator poles exactly invariant, so a factorial family with identical decay spectra can route responses substitutively, complementarily, or neutrally according to the sign of one off-diagonal component, and the factorial contrast is identified as a chain diagram joining a source vertex to that component. Adding a return channel makes the self-energy non-nilpotent, and the local stability boundary becomes the frequency at which a pole crosses into the unstable half-plane, which is the same condition as loop-gain unity. A Gaussian-sector underdetermination result exhibits effective actions with identical propagators, identical linear response, and identical stability whose cubic vertex controls whether a declared relational sector is reachable at all; the earlier controllability countermodel is recovered as its bilinear special case. Persistence of any of these effects requires a soft direction, multistability, or sustained driving. Symbolic configurations such as platform, persona, and accumulated memory enter as a background field whose barrier height formalizes switching cost. These results provide no population-level verdict. The proposed empirical program uses randomized feature policies, sector-specific coupling outcomes, consented availability-resilience contrasts, network-aware longitudinal designs, and multi-horizon capability profiles. Relational generativity remains a revisable attribution.

Keywords: AI-mediated intimacy; relational generativity; dynamical stability; self-energy and vertex structure; longitudinal causal design.

Statements

Status and correspondence

This paper is a working discussion document. Its formal claims are conditional on declared assumptions, and several of its bridges to empirical and philosophical questions remain open. Objections, counterexamples, corrections, alternative formulations, and pointers to relevant literature are all welcome and may be sent to `huangwanhong@serendip.ngo`.

Generative AI use

Generative AI systems were used in the preparation of this work. Anthropic's Claude and OpenAI's ChatGPT supported exploratory discussion of the conceptual framework, development of the formal construction and its notation, identification of candidate literature for subsequent checking, and drafting and revision of the manuscript in LaTeX. The research questions, theoretical commitments, formal claims, and the epistemic status assigned to each were determined and approved by the author, who bears sole responsibility for the content of this paper, including any errors it contains. Neither system is an author of this work and neither holds authorship credit, in accordance with the position that authorship carries responsibilities a generative AI system cannot assume.

License

1 Research Scope and Position

This section establishes the paper's role, target, and sequence. It defines the focal phenomenon, states the limited position, and separates mathematical possibility, empirical mechanism, and normative attribution.

AI companions can provide recognition, continuity, and immediate emotional support. They can also be embedded in commercial systems whose personalization, memory, availability, and exit conditions differ from many human relations. The relevant comparison therefore concerns a coupling architecture; an intrinsic opposition between human and artificial nodes remains outside the model. Human relations may scaffold exploration, exclude participants, reproduce domination, or fail to provide accessible support. AI-mediated relations may bridge toward human and institutional resources, substitute for them, or combine these effects at different horizons.

Definition 1.1 (AI-mediated intimacy)**.** AI-mediated intimacy denotes sustained interaction in which a person uses or experiences an AI system for emotional disclosure, recognition, companionship, or relational support. The definition leaves consciousness, personhood, authenticity, and moral standing open.

An earlier version of this paper made its argument with a stochastic adaptive network and a deterministic local skeleton, and established the stability distinction through a controllability countermodel. The present version derives the same distinctions inside an effective field theory obtained by coarse-graining a background-independent relational substrate. The reason for the change is that the field-theoretic register states the paper's thesis as a fact about the structure of the effective action rather than as a pair of examples.

Claim 1.2 (Structural separation)**.** In an effective action, the quadratic part determines the propagator, hence the relaxation spectrum and every stability property defined from unforced decay. The source-coupling and self-interaction vertices are separate data. Routing, reachability, and generative capacity are properties of the vertices. Consequently no stability measurement constrains them.

Claim 1.2 is the paper's organizing statement, and Section 5 supplies its formal content in three forms. The practical consequence is that a system can be measured to recover quickly and completely from every declared perturbation while the sector into which its recovery is routed, and the set of relational configurations it can subsequently reach, remain entirely undetermined by that measurement.

Claim 1.3 (Layered evaluation)**.** AI-intimacy research should estimate local stabilization, relational routing, functional reliance, and future relational capacity as separate, potentially co-occurring dimensions. A generativity attribution requires additional empirical and normative justification.

The argument proceeds through four stages. Sections 2 and 3 locate the model and define the effective action and its sectors. Sections 4 and 5 define the observables and derive the formal separation results. Sections 6–8 translate the results into mechanism-specific causal contrasts and an ethical longitudinal program. Sections 9 and 10 state the open obligations and the revisable position.

2 Conceptual and Empirical Lineage

This section locates the framework within stability theory, network dynamics, control, viability, field theory, intimacy research, and relational autonomy. The discussion assigns each source a bounded role and uses differences in empirical horizon as evidence for measurement separation.

Holling's distinction between equilibrium stability and ecological resilience provides a major conceptual ancestor for separating return behavior from response under disturbance [12]. Basin stability later made finite-perturbation recovery measurable relative to a selected attractor and perturbation distribution [17]. Neither concept carries an intrinsic judgment about relational value. Kalman's state-space formulation separates stability, controllability, and observable behavior [13]; viability theory adds state constraints and target-capture structure [1]. These formal resources are used as modeling tools whose social interpretation requires an additional empirical and normative bridge.

The field-theoretic register supplies the paper's structural argument. Response functions, self-energies, and vertex expansions are standard objects of nonequilibrium field theory [14], and resummation of repeated interaction into a dressed propagator follows the pattern established by Dyson [5]. The substrate from which the effective action is obtained is the group-field construction developed elsewhere in this research program and summarized in Section 3; its mathematical antecedents are group field theory and spin-foam models [2, 9, 19, 22], and its relational treatment of change follows the internal-clock tradition [20, 23]. None of these sources supplies evidence that relational or affective systems are group-field systems. The substrate is declared, and every result below is proved in the effective sector.

The network architecture joins node dynamics to changing relational structure. Adaptive-network theory studies coevolution between node states and topology [10], while multilayer-network theory supplies a vocabulary for several simultaneous kinds of connection [15]. The present human, AI, and institutional sectors are experimenter-defined projections within that general architecture. Their labels establish neither ontological kinds nor natural boundaries.

Direct empirical studies currently address different portions of the proposed trajectory. Replika research supplies exploratory evidence about perceived authenticity, social motives, interaction, and attachment [21]. Experiments and a repeated-use study report momentary reductions in state loneliness from AI companionship [4]. A four-week randomized study found outcomes that varied with modality, conversational condition, user characteristics, and realized use; associations involving realized usage intensity remain nonrandomized [6]. A two-week randomized comparison among first-year students found a loneliness benefit from human-peer contact and no statistically significant difference between the chatbot and journaling conditions [16]. A four-wave observational study reports small reciprocal within-person paths for a single-item emotional-isolation measure. For the broader social-connection measure, lower connection predicted later chatbot use, while chatbot use did not predict later connection; both findings carry substantial design qualifications [8].

These studies concern distinct estimands, systems, populations, and timescales. Momentary relief, multiweek change, longer-horizon association, and expansion of future relational capability form a sequence of open transitions. Human relationship research also supplies a counterweight to a simple friction-versus-relief story: responsive, nonintrusive support can facilitate exploration and goal pursuit [7]. AI accommodation may therefore scaffold later activity under some architectures.

Relational autonomy supplies the normative orientation for evaluating those architectures. Autonomy can be constituted and shaped through relations and institutions [18]; it need not coincide with maximum contact or solitary independence. Governance analysis of mental-health chatbots further identifies evidence, privacy, transparency, safety, accountability, vulnerable-user, and human-service-displacement obligations [3]. These sources discipline the inquiry while leaving the paper's generativity construct provisional.

3 Effective Action for Coupled Relational Sectors

This section defines the model in dependency order. It records the substrate briefly, introduces the sector decomposition, fixes the effective action and its truncation, and identifies where each dimension of a coupling profile appears in that action.

3.1 Substrate and Sectors

The substrate is a group field Ψ over a declared group G whose perturbative expansion generates combinatorial relational histories $\mathcal{C}$ with factorized configuration amplitudes $\mathcal{A}$. A participant, an AI system, and an institution are subcomplexes $\mathcal{C}_i \subseteq \mathcal{C}$; a relation is a pattern of shared boundary structure; and change is described relative to an internal relational reference χ rather than an external clock. This construction is developed in the companion framework and is used here only to justify three features of the effective description: the sectors are not separate substances, the coupling structure is derived rather than posited on a chosen domain, and symbolic configurations such as a platform or a persona are ordinary configurations of the same substrate.

Under a declared coarse-graining, retain order parameters

$$\phi = (\phi_H, \phi_A, \phi_I), \tag{1}$$

for the human, AI, and institutional sectors, together with declared relation layers indexed by ℓ. Sector labels are experimenter-defined projections of the retained field content.

3.2 Effective Action and Truncation

Write the effective action for the retained content as

$$\Gamma[\phi; \sigma, r] = \frac{1}{2} \int \phi_a \, D^{-1}_{ab} \, \phi_b - \int J_a \phi_a + \sum_{n \geq 3} \frac{1}{n!} \int \lambda_{a_1 \cdots a_n} \phi_{a_1} \cdots \phi_{a_n}, \tag{2}$$

where D is the retarded propagator matrix on the declared relational horizon, J collects declared sources including perturbations and available response actions, λ collects the vertices, r denotes the AI design condition, and σ denotes a symbolic background configuration in the sense of Section 5.7. Repeated indices include integration over the retained coordinates and the relational reference. The self-energy is defined by $D^{-1} = D_0^{-1} - \Pi$.

Assumption 3.1 (Local truncation regularity)**.** On the declared finite relational horizon the retained operator set is finite, D_0 is bounded with bounded inverse on the declared response space, the coupling operators are bounded, and neglected operators contribute below a stated tolerance. The sector construction, projection layers, and intervention meanings remain empirical hypotheses rather than consequences of the formalism.

All results below concern the retarded response problem generated by Equation (2) under Assumption 3.1. Fluctuation contributions and a stochastic stability notion remain separate extensions.

3.3 Where Coupling Dimensions Live

Table 1 disaggregates the coupling profile and names the term of Equation (2) in which each dimension appears. A single edge weight cannot represent these dimensions faithfully, and several may vary independently precisely because they occupy different parts of the action.

Table 1: Coupling dimensions, their location in the effective action, and their experimental roles.

Coupling dimension	Location in Γ	Candidate indicator	Interpretive limit
Bandwidth	source coupling J_A	intensity, duration, modality, disclosed content class	volume supplies no measure of quality or value
Cross-layer reach	off-diagonal self-energy Π_{ka} and cross-sector vertices	completed paths into peers, family, care, community, work, or law	potential paths may remain unused or coercive
Plasticity	curvature of Γ in the coupling directions	change after matched events or design assignments	fast change permits several response signs
Persistence	pole location and presence of soft directions	survival across time, reset, migration, and provider change	finite persistence and permanent retention require different pole structures
Switching cost	barrier height between minima of $\Gamma[\,\cdot\,;\sigma]$	exit effort, data loss, emotional cost, and substitute availability	attachment and harmful lock-in require separate evaluation
Reciprocity	symmetry of the cross-sector self-energy	mutual state change, constraint, initiative, exposure to consequences	behavioral symmetry does not settle phenomenology or moral status
Involuntariness	constraints on the admissible source family	obligations and consequences beyond immediate preference	imposed friction can support or damage capacity
Control	design parameter r entering vertices and sources	ownership of objectives, memory, access, moderation, intervention	formal control can diverge from effective user agency

4 Stability, Routing, and Relational Capacity

This section defines the three formal objects the paper compares. The definitions proceed from unforced recovery, through signed causal routing, to viability-constrained target access, keeping mathematical and normative content visible.

Definition 4.1 (Local stability)**.** The retained sector is locally stable on the declared horizon when the dressed retarded propagator $D(\omega) = (D_0^{-1}(\omega) - \Pi(\omega))^{-1}$ is analytic in the upper half plane, equivalently when every root λ of $\det(D_0^{-1} - \Pi) = 0$ satisfies $\operatorname{Re}\lambda < 0$. Exponential decay of the unforced response follows, with rate the spectral gap.

This property concerns the unforced retained sector and the selected norm. Finite-event recovery, convergence to a shifted minimum, basin size, and the sector in which repair occurs require additional observables.

For a binary AI design condition $r \in \{0, 1\}$ and a matched perturbation $s \in \{0, 1\}$, let $Y_{rs}(\chi)$ be a signed scalar outcome under the corresponding potential trajectory. Define the perturbation-specific AI interaction

$$\mathfrak{I}_\chi[Y] = Y_{11}(\chi) - Y_{10}(\chi) - Y_{01}(\chi) + Y_{00}(\chi). \tag{3}$$

Consistency, well-defined interventions, and an identification design are required before Equation (3) becomes an estimand. For sector outputs Y_k, a signed routing profile retains the trajectories $\{\mathfrak{I}_\chi[Y_k] : k \in \{HH, HA, HI\}\}$ over the declared horizon. Increased HA response alone describes uptake; substitution additionally requires a negative effect in a declared quality-adjusted human or institutional sector, and complementarity requires a positive effect there.

Definition 4.2 (Relational target-capability profile)**.** Let $\mathcal{V}$ be a declared viability set, $\mathcal{U}$ an admissible source family, $\mathcal{K}_j$ a declared relational target set, and $\ell_{\mathrm{act}} \geq 0$ a declared action cost density. For horizon H, define

$$\mathcal{E}_j(\phi_0; H) = \inf_{J \in \mathcal{U}} \int_0^H \ell_{\mathrm{act}}(J(\chi), \phi(\chi), \chi)\, \mathrm{d}\chi, \tag{4}$$

subject to stationarity of Equation (2), $\phi(\chi) \in \mathcal{V}$ throughout, and $L_j\phi(H) \in \mathcal{K}_j$. Set $\mathcal{E}_j = +\infty$ when the target is unavailable. For a declared decreasing map ψ with $\psi(+\infty) = 0$, the profile is $\boldsymbol{g}_H(\phi_0) = (\psi(\mathcal{E}_1), \ldots, \psi(\mathcal{E}_J))$.

The profile retains target-specific information. Any scalar aggregation adds weights whose normative role must be reported. Field content determines which capabilities become visible; $\mathcal{U}$ encodes practical resources; $\mathcal{V}$ encodes admissible trajectories; H privileges a timescale; target sets encode valued ends; and population aggregation distributes benefits and burdens. Equation (4) therefore has the status of an instrumental capacity object, and relational generativity remains a further attribution.

5 Formal Separation Results

This section establishes the paper's mathematical core. It proves that a strictly triangular cross-sector self-energy leaves the spectrum invariant, derives the factorial routing contrast as a chain diagram, locates the stability boundary as a pole crossing, and then proves the general underdetermination of the interacting sector by the Gaussian sector.

5.1 Spectral Invisibility of Routing

Lemma 5.1 (Nilpotent self-energy preserves the spectrum). *Let D_0^{-1} be triangular with respect to an ordering of the retained sectors and let Π be strictly triangular with respect to the same ordering, so that Π is nilpotent. Then*

$$\det(D_0^{-1} - \Pi) = \det(D_0^{-1}), \tag{5}$$

and the dressed propagator has exactly the poles of the undressed one.

Proof. A triangular matrix minus a strictly triangular matrix with the same orientation is triangular with unchanged diagonal. The determinant of a triangular matrix is the product of its diagonal entries, which gives Equation (5). Poles of the dressed propagator are the zeros of that determinant. □

Lemma 5.1 is the sharpest available statement of the paper's thesis. A cross-sector influence that never returns contributes a nilpotent self-energy, and a nilpotent self-energy is invisible to every observable defined from the pole structure. Stability measurements are therefore not insensitive to routing by accident of modeling; they are insensitive to it by the algebra of the response problem.

5.2 Factorial Routing as a Chain Diagram

This subsection derives perturbation-specific accommodation and routing from a minimal triangular sector system, so that every sign and stability conclusion traces to a declared coefficient.

Let $q \in L^1_{\text{loc}}$ with $q \geq 0$ be a declared perturbation profile. For AI condition r and perturbation s, let a_{rs} be AI-mediated uptake, d_{rs} a local distress deviation, and h_{rs} a declared human or institutional pathway deviation. With zero initial deviations, take

$$\partial_\chi a_{rs} = -\kappa_A a_{rs} + rs\,\pi q(\chi), \tag{6}$$

$$\partial_\chi d_{rs} = -\kappa_D d_{rs} + sq(\chi) - \theta a_{rs}, \tag{7}$$

$$\partial_\chi h_{rs} = -\kappa_H h_{rs} - s\chi_{\text{c}} q(\chi) + b\, a_{rs}, \tag{8}$$

with $\kappa_A, \kappa_D, \kappa_H, \pi, \theta > 0$, $\chi_{\text{c}} \geq 0$, and $b \in \mathbb{R}$. In the language of Equation (2), the coefficient π is a three-point source vertex $\lambda_{a\,r\,s}$ coupling the two assignments to the AI sector; θ and b are components Π_{da} and Π_{ha} of a strictly triangular cross-sector self-energy; and $\kappa_A, \kappa_D, \kappa_H$ are the diagonal inverse propagators. A useful decomposition is $b = b_{\text{bridge}} - b_{\text{displace}}$ with both terms nonnegative; neither term follows from AI identity.

The retarded propagator of a single sector is the positive kernel

$$D_\kappa(\chi) = \Theta(\chi)\, e^{-\kappa\chi}, \qquad (D_\kappa \star f)(\chi) = \int_0^\chi e^{-\kappa(\chi-\tau)} f(\tau)\, \mathrm{d}\tau. \tag{9}$$

Solving in dependency order gives $a_{rs} = rs\,\pi\, D_A \star q$, $d_{rs} = s\, D_D \star q - \theta\, D_D \star a_{rs}$, and $h_{rs} = -s\chi_{\mathrm{c}}\, D_H \star q + b\, D_H \star a_{rs}$. The factorial operation of Equation (3) cancels every term linear in a single assignment, leaving

$$\mathfrak{I}_\chi[a] = \pi\,(D_A \star q)(\chi), \tag{10}$$

$$\mathfrak{I}_\chi[d] = -\theta\pi\,(D_D \star D_A \star q)(\chi), \tag{11}$$

$$\mathfrak{I}_\chi[h] = b\,\pi\,(D_H \star D_A \star q)(\chi). \tag{12}$$

Proposition 5.2 (Stable accommodation with divergent routing)**.** *Under the stated assumptions, $\mathfrak{I}_\chi[a] \geq 0$ and $\mathfrak{I}_\chi[d] \leq 0$, and the sign of $\mathfrak{I}_\chi[h]$ equals the sign of the off-diagonal self-energy component b. The retained sector is locally stable in the sense of Definition 4.1 for every $b \in \mathbb{R}$, with poles $-\kappa_A, -\kappa_D, -\kappa_H$. Equal local accommodation and identical decay spectra therefore coexist with substitutive routing for $b < 0$, complementary routing for $b > 0$, and neutral routing for $b = 0$.*

Proof. The kernel in Equation (9) maps nonnegative functions to nonnegative functions, and convolution preserves that property, so Equations (10)–(12) give the three sign conclusions. In the ordering (a, d, h) the self-energy has nonzero entries only at positions (d, a) and (h, a) and is therefore strictly triangular; Lemma 5.1 gives the invariance of the pole set, and the diagonal entries supply the stated poles, all with negative real part. Off-diagonal components can still generate transient amplification, which the spectral conclusion does not bound uniformly over arbitrary parameter families. □

Equation (12) identifies the routing contrast as a chain: the source pair enters through the vertex π, propagates through the AI sector by D_A, crosses through the self-energy component b, and propagates to the observable by D_H. Reading the contrast this way explains why it carries sign information that the spectrum cannot: the sign lives on the crossing, and the spectrum sees only the endpoints.

Corollary 5.3 (Accommodation–routing separation)**.** *A reduction in the declared distress response supplies no sign information about a human or institutional routing response. Weak cross-layer reach corresponds locally to $b \approx 0$; crowd-out requires the additional empirical condition $b < 0$.*

5.3 Return Channels and the Pole-Crossing Threshold

This subsection tests the stronger claim that responsive AI coupling can destabilize relational dynamics. Adding one return channel makes the self-energy non-nilpotent, and Lemma 5.1 no longer applies.

Consider deviations governed by

$$\partial_\chi a = -\kappa_A a - \gamma h, \qquad \partial_\chi h = -\kappa_H h + ba, \qquad \gamma > 0, \tag{13}$$

where a negative h deviation activates AI uptake. Eliminating h gives the sector self-energy

$$\Pi_{aa}(\omega) = \frac{-\gamma b}{-i\omega + \kappa_H}, \tag{14}$$

so that the dressed inverse propagator is $D_{aa}^{-1} = (-i\omega + \kappa_A) - \Pi_{aa}$ and, writing $\lambda = -i\omega$, the pole condition is

$$P(\lambda) = \lambda^2 + (\kappa_A + \kappa_H)\lambda + \kappa_A\kappa_H + \gamma b = 0. \tag{15}$$

Proposition 5.4 (Feedback-gain stability region)**.** *The retained sector of Equation* (13) *is locally stable exactly when*

$$\kappa_A\kappa_H + \gamma b > 0. \tag{16}$$

For $b < 0$, stability is equivalent to $\gamma|b| < \kappa_A\kappa_H$. Equivalently, instability occurs exactly when the zero-frequency loop gain $|D_{aa}^{(0)}(0)\,\Pi_{aa}(0)| = \gamma|b|/(\kappa_A\kappa_H)$ reaches unity with $b < 0$, at which point a pole crosses into the closed upper half plane.

Proof. For a real quadratic with positive leading coefficient, both roots have negative real part exactly when the remaining coefficients are positive. The coefficient $\kappa_A + \kappa_H$ is positive by assumption, and the constant term is $\kappa_A\kappa_H + \gamma b$, which gives Equation (16). Equality creates a root at the origin and a negative constant term creates roots of opposite sign. Evaluating Equation (14) and $D_{aa}^{(0)} = (\kappa_A)^{-1}$ at $\omega = 0$ gives the stated loop gain, whose unit value is exactly the vanishing of the constant term. □

The threshold makes a previously intuitive transition conditional. Strong accommodation or rapid uptake alone creates no unstable pole. Instability appears only when a substitutive return loop exceeds the product of the two recovery rates. The reformulation also unifies two results across this research program: Equation (16) is the same condition as the convergence boundary of a reciprocal-response resummation, so the destabilizing regime here and the collective-mode regime in the companion developmental framework are one phenomenon in two sectors. Nonlinear saturation, delay, fluctuation forcing, and reference-dependent parameters can modify this local criterion.

5.4 Gaussian-Sector Underdetermination

This subsection proves the paper's principal new result. It exhibits a family of effective actions with identical quadratic sectors whose cubic vertex alone determines whether a declared relational sector can be reached at all.

Consider two retained sectors with a block-diagonal, ρ-independent quadratic part and one cubic vertex,

$$\Gamma_\rho[\phi] = \frac{1}{2}\int \phi_1 D_{11}^{-1}\phi_1 + \frac{1}{2}\int \phi_2 D_{22}^{-1}\phi_2 + \frac{\rho}{2}\int \phi_1^2\phi_2 - \int J\phi_1, \qquad \rho \geq 0, \tag{17}$$

with the source supported on sector 1. The first coordinate can represent an immediately available support response and the second a declared later relational target channel; the proof uses only Equation (17).

Proposition 5.5 (Gaussian-sector underdetermination)**.** *For the family in Equation* (17)*:*

1. *the propagator, its poles, the relaxation spectrum, and the entire linear response to J are independent of ρ, so every system in the family has identical stability in the sense of Definition 4.1;*

2. *the stationary configuration of the second sector satisfies*

$$\phi_2 = -\frac{\rho}{2} D_{22}\,\phi_1^2 + O(\rho^2), \qquad \phi_1 = D_{11}J + O(\rho), \tag{18}$$

so the reached amplitude in the second sector is first order in ρ;

3. *at $\rho = 0$ the second sector is exactly unreachable at every order for every source supported on the first sector;*

4. *the minimal source magnitude required to reach a target $L_2\phi(H) = \xi \neq 0$ obeys*

$$\|J\|^2 \gtrsim \frac{2|\xi|}{\rho\,\|D_{22}\|\,\|D_{11}\|^2}, \tag{19}$$

which diverges as $\rho \downarrow 0$.

Proof. Part (i) holds because ρ multiplies only a cubic term, and the propagator is the inverse of the second functional derivative evaluated at the reference configuration, where the cubic term contributes a factor linear in the fields and therefore vanishes. Linear response is $\phi = DJ$, independent of ρ. For part (ii), stationarity $\delta\Gamma_\rho/\delta\phi_2 = 0$ gives $D_{22}^{-1}\phi_2 + \frac{\rho}{2}\phi_1^2 = 0$, hence the first relation of Equation (18), and $\delta\Gamma_\rho/\delta\phi_1 = 0$ gives $D_{11}^{-1}\phi_1 + \rho\phi_1\phi_2 = J$, whose leading solution is $\phi_1 = D_{11}J$. Part (iii) follows because at $\rho = 0$ the action separates into two independent quadratic pieces and the second carries no source, so ϕ_2 remains at its reference value under every source and to every order. Part (iv) follows by inverting Equation (18) for the source magnitude and bounding the operator norms. □

Claim 5.6 (Stability data and generativity data are disjoint)**.** Propositions 5.2 and 5.5 isolate two independent failures of inference from stability. The first shows that a cross-sector influence can be invisible to the spectrum; the second shows that the entire quadratic sector, including every stability property and the full linear response, is compatible with any value of a vertex that governs whether a relational sector is reachable at all. An empirical program that measures only recovery therefore measures D, and the objects of relational evaluation are λ.

5.5 Bilinear Special Case and the Reachable Set

The earlier controllability countermodel is recovered as the bilinear case of Proposition 5.5, in which the ρ-dependence sits in a source-coupling vertex rather than a self-interaction. Consider

$$\partial_\chi y = -y + B_\rho v, \qquad B_\rho = \mathrm{diag}(1, \rho), \qquad y(0) = 0, \tag{20}$$

over $0 \le \chi \le H$ with $\rho \ge 0$ and $v \in L^2$. The finite-horizon controllability Gramian is

$$W_H(\rho) = \int_0^H e^{-(H-\chi)} B_\rho B_\rho^\mathsf{T} e^{-(H-\chi)}\,\mathrm{d}\chi = q_H\,\mathrm{diag}(1, \rho^2), \qquad q_H = \frac{1 - e^{-2H}}{2}. \tag{21}$$

For $\rho > 0$ the unit-energy reachable set is the ellipsoid $\{y : y^\mathsf{T} W_H(\rho)^{-1} y \le 1\}$ with area $\pi\sqrt{\det W_H(\rho)} = \pi q_H \rho$, and for a target $\xi = (0, \varrho)^\mathsf{T}$ with $\varrho \neq 0$ the minimum energy is

$$E_{\min}(\xi; \rho) = \xi^\mathsf{T} W_H(\rho)^{-1}\xi = \frac{\varrho^2}{q_H \rho^2}. \tag{22}$$

Proposition 5.7 (Equal stability with unequal target access)**.** *Every system in Equation* (20) *has unforced eigenvalues* $(-1, -1)$ *and identical exponential decay. Its reachable area varies linearly with ρ, the minimum energy in Equation* (22) *diverges as $\rho \downarrow 0$, and at $\rho = 0$ the target is unreachable under every finite-energy source. Local stability therefore determines neither reachable volume nor target-access cost.*

Proof. The unforced matrix is $-\mathbb{1}_2$ for every ρ, which gives the spectral statement. Equations (21) and (22) give the capacity and cost statements. At $\rho = 0$ the second row of the input matrix vanishes, so every reachable state has vanishing second coordinate. □

Reachable area depends on coordinates, source budget, and measure; it provides no direct measure of relational value. Viability constraints may remove paths contained in the ellipsoid, and a harmful or unwanted target deserves no positive weight. A transparent bridge from the response model to the capacity proxy is

$$\rho_H = \rho_0 + \lambda_H h(H) + \lambda_A a(H), \qquad \mathfrak{I}_H[\rho] = \lambda_H \mathfrak{I}_H[h] + \lambda_A \mathfrak{I}_H[a]. \tag{23}$$

Even with $b < 0$, a capacity contraction follows only when the right-hand side of Equation (23) is negative and ρ_H remains within the model's domain. AI-created pathways can offset human-pathway substitution, and the coefficients λ_H, λ_A require independent operational and normative justification.

5.6 Persistence Boundary

This subsection states the temporal limit of the stable response model and identifies the structures required before a finite-horizon routing effect can represent durable relational change.

Lemma 5.8 (Transient response under gapped propagators)**.** *If the declared perturbation has compact support and every pole of the dressed retarded propagator lies strictly in the lower half plane, then every connected response function built from finitely many propagators converges to zero. In particular each contrast in Equations* (10)–(12) *vanishes as* $\chi \to \infty$*.*

Proof. A gapped retarded propagator has an exponentially decaying kernel. After the support of the perturbation, each convolution of exponentially decaying kernels decays exponentially, and a finite product of such factors decays. □

Corollary 5.9 (Requirements for persistence)**.** *Persistent relational restructuring after a bounded perturbation requires at least one of: a soft direction, that is a pole at the origin corresponding to a flat direction of* Γ*; multistability, that is a second local minimum of* Γ *reachable from the first; or sustained driving. A soft direction removes the spectral gap of the full retained sector, and multistability supplies stable pre-event and post-event configurations.*

5.7 Symbolic Background and Switching Cost

Platform, persona, accumulated memory, conversational history, and social identification are symbolic configurations σ: ordinary configurations of the same substrate whose function is to condition later relational dynamics. They enter Equation (2) as a background field, and the conditioned action $\Gamma[\phi; \sigma]$ shifts minima, changes curvature, and raises or lowers barriers.

Definition 5.10 (Switching barrier and lock-in)**.** For symbolic configurations σ and σ' with corresponding minima ϕ_σ and $\phi_{\sigma'}$, define the switching barrier

$$\mathcal{B}(\sigma \to \sigma') = \inf_{\text{admissible paths}} \sup_{\chi} \Big(\Gamma[\phi(\chi);\, \cdot\,] - \Gamma[\phi_\sigma; \sigma]\Big), \tag{24}$$

the minimal peak cost along any admissible path from one configuration to the other. Lock-in obtains when $\mathcal{B}$ exceeds the fluctuation and resource scale available to the participant.

Claim 5.11 (Reliance, lock-in, and capacity are distinct)**.** Functional reliance is a large availability contrast; lock-in additionally requires a large switching barrier in the sense of Equation (24); and capacity contraction additionally requires a reduction in the number of low-curvature directions of $\Gamma[\,\cdot\,; \sigma]$. A system can exhibit any one of these without the others, and design choices affect them separately.

Claim 5.11 supplies the formal home for the switching-cost row of Table 1 and connects this paper to the candidate architecture of symbolic alienation developed in the companion developmental framework, in which a configuration generated by a system's own relational activity subsequently constrains that activity.

6 AI-Intimacy Mechanism Profiles

This section translates the formal distinctions into empirical constructs. It defines each construct through a causal contrast and a counterpattern, thereby preventing use intensity or immediate relief from carrying several meanings at once.

For quality-adjusted coupling sector $k \in \{HH, HA, HI\}$, treatment policies r, r', perturbation reference χ_0, and horizon H, define

$$\begin{aligned}\tau_k(H; r, r') = {}& \mathbb{E}\Big[C_i^k(\chi_0 + H) \mid do(R = r, S = 1)\Big] - \mathbb{E}\Big[C_i^k(\chi_0 + H) \mid do(R = r, S = 0)\Big] \\ & - \mathbb{E}\Big[C_i^k(\chi_0 + H) \mid do(R = r', S = 1)\Big] + \mathbb{E}\Big[C_i^k(\chi_0 + H) \mid do(R = r', S = 0)\Big] .\end{aligned} \tag{25}$$

Joint randomized assignment of R and an ethically benign standardized S identifies this design-by-perturbation policy interaction under the trial's conditions. When $S = 1$ records a naturally occurring event, randomized response policy identifies the history-conditional contrast

$$\begin{aligned}\tau_k^{\text{event}}(H; r, r' \mid \mathcal{H}) = {}& \mathbb{E}\Big[C_i^k(\chi_0 + H) \mid do(R = r), S = 1, H_{\chi_0} \in \mathcal{H}\Big] \\ & - \mathbb{E}\Big[C_i^k(\chi_0 + H) \mid do(R = r'), S = 1, H_{\chi_0} \in \mathcal{H}\Big] ,\end{aligned} \tag{26}$$

for a preregistered history class $\mathcal{H}$; it supplies no intervention contrast on event occurrence. Event-selection assumptions are required to transport this conditional effect beyond the observed event histories. Either design leaves the effect of realized AI use unidentified without further assumptions, because the assigned feature can act through several channels and realized use is selected dynamically.

Table 2 records the mechanism-specific estimands together with the object in the effective action that each targets. The rows can coexist: a system may stabilize distress, substitute for one pathway, bridge into another, and create functional reliance over different horizons.

Functional reliance can be represented through the bounded availability contrast

$$\begin{aligned}D_i(\Delta) = {}& \mathbb{E}[W_i(\chi + \Delta) \mid do(AI \text{ available})] \\ & - \mathbb{E}[W_i(\chi + \Delta) \mid do(AI \text{ unavailable, alternatives supplied})].\end{aligned} \tag{27}$$

A large D_i describes functional reliance. Harmful dependency adds evidence of impaired agency, endorsed-goal conflict, escalating switching cost, coercive design, or net welfare loss. High use alone supplies neither contrast.

Claim 6.1 (Architecture-conditional mechanism)**.** Rapid local accommodation, negative net routing $b < 0$, a large switching barrier, and reduced low-curvature directions can jointly produce strong short-run recovery and contraction in selected later relational capacities. Bridge components of the cross-sector self-energy, AI-created target channels, accessibility gains, and autonomy-support features can reverse or offset that result.

This claim is an empirical conjecture. The formal results show available configurations and required premises; they do not establish their prevalence, causes, or value.

Table 2: Mechanism-specific estimands, their formal targets, and discriminating counterpatterns.

Construct	Formal target	Primary causal contrast	Observable signature	Discriminating counterpattern
Stabilization	pole locations of D	assigned support design after a matched perturbation	faster affective or functional recovery and reduced volatility	recovery unchanged, fades immediately, or reflects reporting demand
Substitution	sign of Π_{HA}	$\tau_{HA} > 0$ jointly with sector effects	AI coupling rises while quality-adjusted human or institutional pathways decline	human decline precedes AI engagement, or randomized engagement leaves those paths unchanged
Complementarity	sign of Π_{IA}	joint AI support and human-resource access	AI coupling and human or institutional participation both improve	apparent joint change explained by baseline access or bridge assignment
Functional reliance	availability contrast	brief consented availability or migration contrast with alternatives supplied	disproportionate functional loss, failed migration, rising switching cost	functioning and agency remain robust during migration or planned unavailability
Lock-in	barrier $\mathcal{B}$	consented migration with alternatives and support supplied	exit attempted and abandoned at measurable cost	exit completed at declared cost with retained functioning
Network bridge	cross-sector vertex	enhanced opt-in warm handoff versus standard information	a new reciprocal tie or completed institutional connection persists beyond a declared duration	recommendations increase clicks without sustained ties or service uptake
Relational capacity	vertex set and reachable configurations	feature-policy contrast across target, viability, and horizon specifications	robust improvement in feasible, valued, reversible, nonredundant pathways	direction changes across reasonable specifications or fails held-out opportunity tasks

7 Longitudinal Causal Program

This section specifies a staged empirical design capable of discriminating the mechanism profiles. The sequence begins with construct validation, introduces safe feature-policy assignments, and extends measurement across individual and network horizons.

Table 3 summarizes the design. A single AI-versus-no-AI trial would confound accommodation, bridge, autonomy, memory, and governance features. A factorial policy trial can vary bounded personalization, opt-in bridging, and autonomy support while preserving core safety protections in every arm.

The standardized-perturbation analysis targets Equation (25); the natural-event analysis targets Equation (26). Actual-use effects require a longitudinal model for time-varying need, prior use, personalization, network response, and subsequent use. Conventional cross-lagged models can mix stable between-person differences with within-person dynamics [11]; a random intercept improves that decomposition and still leaves causal identification open. Longitudinal weighting or g-formula analyses require explicit sequential exchangeability, positivity, consistency, and correct-model assumptions.

Two design elements follow directly from the formal results and are new in this version. Because Proposition 5.5 shows that recovery measurements cannot constrain reachability, a study that reports only stabilization outcomes should be described as having measured the propagator and nothing else. Because Proposition 5.4 identifies the stability boundary with unit loop gain, a graded responsiveness assignment can estimate the margin to that boundary directly, and the estimated margin is a safety-relevant quantity that should be monitored rather than approached.

Table 3: Staged longitudinal design and ethical boundary conditions.

Study stage	Assignment structure	Horizon	Principal output	Ethical boundary
Construct validation	interviews, diaries, and participatory workshops	pretrial	valid measures of coupling quality, agency, viability, and reliance	usage intensity receives no clinical or moral label
Baseline panel	observational within-person panel	six to eight weeks	need, network, use, and pre-event trajectories	intimate content and third-party network data are minimized
Feature factorial	accommodation × bridge × autonomy support	hours to four months	policy effects on recovery, routing, and reliance	every arm retains safety, exit, portability, and crisis protections
Standardized perturbation	benign task assignment S × feature policy R	minutes to days	factorial interaction in (25) and the sign of b	social deprivation, intimate loss, and acute distress remain outside assignment
Event-triggered trial	low-risk randomized response after naturally occurring events	hours to weeks	history-conditional policy effect in (26)	grief, abuse, severe crisis, and acute risk remain outside randomization
Loop-gain estimation	graded responsiveness assignment within safe bounds	weeks	estimates of γ, b, and the margin in (16)	responsiveness levels stay well inside the stable region and are monitored
Network trial	clustered enhanced bridge access	three to twelve months	durable human and institutional ties plus spillovers	community consent and protection from coerced disclosure
Resilience assessment	consented migration or planned brief unavailability with alternatives	days to months	functional reliance, portability, and an estimate of $\mathcal{B}$	immediate restoration and human alternatives remain available
Long follow-up	continued observation after assignment	six to twelve months or longer	persistence, horizon reversal, and capacity profiles	consent is renewable and withdrawal carries no penalty

Network spillovers require cluster assignment or a declared exposure mapping. Logged assignment probabilities are essential for adaptive features. Pre-event trends, platform updates, migration, deletion, survey missingness, and informative attrition require direct modeling. Negative controls and quantitative sensitivity analyses should accompany the primary estimates. Reachability predictions should be tested against held-out action-opportunity tasks alongside their fit to observed trajectories.

The substitution conjecture loses support when randomized accommodation improves recovery while quality-adjusted human, institutional, and capacity outcomes show no decline across preregistered horizons. It also loses support when human-network decline clearly precedes AI engagement or disappears under randomized assignment and credible adjustment. Bridge claims require completed and sustained connections; a referral display, click, or stated intention provides weak routing evidence.

8 Ethical and Distributional Constraints

This section constrains the source, target, and data spaces used by the formal and empirical program. It treats participant welfare, relational autonomy, third-party privacy, population heterogeneity, and platform power as parts of model validity.

The components of the capability profile encode value choices. The retained field content selects visible relations; $\mathcal{U}$ encodes resources and practical agency; $\mathcal{V}$ encodes a conception of viable conduct; $\mathcal{K}_j$ selects valued ends; the horizon privileges immediate recovery or development; and aggregation can hide exclusion. Participant endorsement helps specify these objects and can itself be shaped by adaptation, manipulation, or constrained alternatives. Participatory construction should therefore be combined with rights, safety, third-party-impact, and anti-discrimination constraints.

The empirical record should report a vector containing feasible options, realized transition success, nonredundancy, reversibility, participant endorsement, cost, risk, robustness, and distribution. Several defensible target, viability, and horizon specifications should be preregistered. A reduction in tie count can expand capacity when harmful relations end; increased connectivity can reproduce coercion or exclusion. Heterogeneous effects should be reported by baseline isolation, disability, rural access, caregiving burden, age, and available support wherever sample size and privacy permit.

Experimental authorization has a narrower boundary than observation. Studies should avoid inducing loneliness, grief, relationship rupture, exclusivity, dependency, or high switching costs, and the barrier of Equation (24) should be estimated rather than manipulated upward. Crisis response remains outside feature manipulation. Every arm preserves transparent exit, deletion, portability, and alternative support. Planned unavailability requires consent, brief duration, substitutes, monitoring, and immediate restoration. Data collection minimizes intimate content and third-party representations. Independent safety monitoring uses preregistered stopping conditions for distress, isolation, coercion, discrimination, or functional decline.

Commercial engagement objectives can conflict with participant welfare and causal interpretation. Assignment, adverse-event review, and analysis require independence from engagement optimization. Deceptive anthropomorphism and concealed attachment optimization require explicit scrutiny. Future evidence of morally relevant AI welfare would add another affected population and require revision of the protocol.

9 Formal and Empirical Research Program

This section identifies the open obligations generated by the framework. The agenda separates formal extension, measurement validation, causal inference, and normative specification so progress in one area cannot silently certify the others.

Open Problem 9.1 (Vertex estimation)**.** Develop estimators for the cross-sector self-energy components and the leading vertices of Equation (2) from feasible relational data, with uncertainty quantification. Proposition 5.5 makes these, rather than the propagator, the objects of relational evaluation, and no current measurement program targets them.

Open Problem 9.2 (Persistent adaptive coupling)**.** Construct an identifiable extension in which a bounded perturbation changes a later minimum of Γ. Compare soft directions, multistability, and sustained driving as mechanisms while preserving distinct decay, plasticity, routing, and barrier parameters, as Corollary 5.9 requires.

Open Problem 9.3 (Target-capability validation)**.** Develop target families and action-cost models whose predictions transfer to held-out relational opportunities. Audit coordinate dependence, latent reparameterization, accessibility, cultural variation, and unequal resource constraints.

Open Problem 9.4 (Network causal identification)**.** Estimate feature-policy effects with treatment–confounder feedback, interference, evolving networks, platform updates, migration, and informative attrition. State the estimand preserved by every approximation.

Open Problem 9.5 (Relational value and AI moral status)**.** Specify conditions under which a robust capacity change warrants a generativity attribution. Keep reciprocity, authenticity, dependence, participant endorsement, third-party effects, and possible AI welfare as distinct considerations.

Several attractive claims remain outside the demonstrated framework. The model supplies no universal heat-sink mechanism, inevitable human-network decay, monadic-star topology, topological phase transition, or general critical AI threshold. Human-only networks possess no formal guarantee of repair. High cross-sector reach may also propagate misinformation, domination, overload, or exclusion. Empirical comparisons must preserve these countermodels.

10 Revisable Position

This section consolidates the defensible conclusions and states their revision conditions. The position remains deliberately narrower than a diagnosis of AI companionship as a social pathology or a celebration of it as a universal support technology.

The formal results establish one structural claim in four forms. A cross-sector influence that never returns contributes a nilpotent self-energy and is exactly invisible to the relaxation spectrum, so systems with identical decay can route substitutively, complementarily, or neutrally according to one off-diagonal sign. Adding a return channel makes the self-energy non-nilpotent and produces a sharp threshold at unit loop gain, which is the same condition that bounds a reciprocal-response resummation in the companion developmental framework. The entire quadratic sector, including every stability property and the full linear response, is compatible with any value of a vertex that determines whether a declared relational sector is reachable at all, and the earlier controllability countermodel is the bilinear case of that statement. Persistence of any of these effects requires a soft direction, multistability, or sustained driving.

The empirical implication is methodological and now sharper than in the earlier version. Measuring stabilization measures the propagator. Routing, reachability, switching barriers, and capacity are vertex data, and they require their own contrasts, their own horizons, and in most cases their own study designs. Current studies illuminate several earlier portions of this sequence and leave the longer-horizon capacity construct largely unmeasured.

The resulting position is modest: AI intimacy should be evaluated by the relational pathways its concrete architectures make feasible, costly, persistent, reversible, and distributively available. Stability supplies one important property within that evaluation, and the formal results locate exactly how little it settles. Generativity remains a further, revisable attribution.

References

[1] Jean-Pierre Aubin. *Viability Theory.* Birkhäuser, Boston, 2009. Reprint of the 1991 edition.

[2] D. V. Boulatov. A model of three-dimensional lattice gravity. *Modern Physics Letters A*, 7(18):1629–1646, 1992.

[3] Simon Coghlan, Kobi Leins, Susie Sheldrick, Marc Cheong, Piers Gooding, and Simon D'Alfonso. To chat or bot to chat: Ethical issues with using chatbots in mental health. *Digital Health*, 9:20552076231183542, 2023.

[4] Julian De Freitas, Zeliha Oğuz-Uğuralp, Ahmet Kaan Uğuralp, and Stefano Puntoni. AI companions reduce loneliness. *Journal of Consumer Research*, 52(6):1126–1148, 2026. Advance publication June 25, 2025.

[5] Freeman J. Dyson. The S matrix in quantum electrodynamics. *Physical Review*, 75(11):1736–1755, 1949.

[6] Cathy Mengying Fang, Auren R. Liu, Valdemar Danry, Eunhae Lee, Samantha W. T. Chan, Pat Pataranutaporn, Pattie Maes, Jason Phang, Michael Lampe, Lama Ahmad, and Sandhini Agarwal. How AI and human behaviors shape psychosocial effects of chatbot use: A longitudinal randomized controlled study. *arXiv preprint arXiv:2503.17473*, 2025.

[7] Brooke C. Feeney. A secure base: Responsive support of goal strivings and exploration in adult intimate relationships. *Journal of Personality and Social Psychology*, 87(5):631–648, 2004.

[8] Dunigan Folk and Elizabeth Dunn. How does turning to AI for companionship predict loneliness and vice versa? *Psychological Science*, 37(4):276–286, 2026.

[9] Laurent Freidel. Group field theory: An overview. *International Journal of Theoretical Physics*, 44(10):1769–1783, 2005.

[10] Thilo Gross and Bernd Blasius. Adaptive coevolutionary networks: A review. *Journal of the Royal Society Interface*, 5(20):259–271, 2008.

[11] Ellen L. Hamaker, Rebecca M. Kuiper, and Raoul P. P. P. Grasman. A critique of the cross-lagged panel model. *Psychological Methods*, 20(1):102–116, 2015.

[12] C. S. Holling. Resilience and stability of ecological systems. *Annual Review of Ecology and Systematics*, 4:1–23, 1973.

[13] Rudolf E. Kalman. Mathematical description of linear dynamical systems. *Journal of the Society for Industrial and Applied Mathematics, Series A: Control*, 1(2):152–192, 1963.

[14] Alex Kamenev. *Field Theory of Non-Equilibrium Systems*. Cambridge University Press, Cambridge, 2011.

[15] Mikko Kivelä, Alexandre Arenas, Marc Barthelemy, James P. Gleeson, Yamir Moreno, and Mason A. Porter. Multilayer networks. *Journal of Complex Networks*, 2(3):203–271, 2014.

[16] Ruo-Ning Li, Dunigan Folk, Abhay Singh, Lyle Ungar, and Elizabeth Dunn. Is a random human peer better than a highly supportive chatbot in reducing loneliness over time? *Journal of Experimental Social Psychology*, 125:104911, 2026.

[17] Peter J. Menck, Jobst Heitzig, Norbert Marwan, and Jürgen Kurths. How basin stability complements the linear-stability paradigm. *Nature Physics*, 9:89–92, 2013.

[18] Jennifer Nedelsky. Reconceiving autonomy: Sources, thoughts and possibilities. *Yale Journal of Law & Feminism*, 1(1):7–36, 1989.

[19] Daniele Oriti. Group field theory as the second quantization of loop quantum gravity. *Classical and Quantum Gravity*, 33(8):085005, 2016.

[20] Don N. Page and William K. Wootters. Evolution without evolution: Dynamics described by stationary observables. *Physical Review D*, 27(12):2885–2892, 1983.

[21] Iryna Pentina, Tyler Hancock, and Tianling Xie. Exploring relationship development with social chatbots: A mixed-method study of Replika. *Computers in Human Behavior*, 140:107600, 2023.

[22] Alejandro Perez. The spin-foam approach to quantum gravity. *Living Reviews in Relativity*, 16:3, 2013.

[23] Carlo Rovelli. *Quantum Gravity*. Cambridge University Press, Cambridge, 2004.