Technological Substitution and the Potential Phase-Out of Biological Human Reproduction

A Mechanistic Hypothesis Integrating Demography, Evolutionary Neuroscience, and AI Dynamics

A note before the paper

The paper that follows presents the same argument as the book, in the form an academic paper requires. Where the book asserts, the paper hedges; where the book moves through narrative, the paper proceeds by citation. The book is the version designed to engage what the paper cannot. The paper is the calibrated version.

— Tommy

Abstract

This paper develops a mechanistic hypothesis for the gradual, voluntary phase-out of biological human reproduction as a majority strategy in the populations most exposed to advanced artificial intelligence and mature neurotechnology. The argument is that convergence of these two technology clusters, over a multi-decade horizon, could render biological reproduction instrumentally suboptimal for a growing share of individuals in high-income societies, and that individually rational substitution choices could aggregate into a population-level outcome that few individuals would endorse in prospect.

The hypothesis operates on two linked timescales. In the near term (roughly ten to twenty years), AI-mediated companionship acts primarily at the partnership margin, amplifying existing fertility declines by extending the well-documented postponement of cohabitation, marriage, and first birth. In the longer term (roughly forty to sixty years, conditional on neurotechnology maturing along current trajectories), direct neural interfaces capable of engaging pair-bonding circuitry could substitute for the reward profile of partnership itself, pushing the relevant decision from postponement toward foregone reproduction. The paper argues that the near-term mechanism is the bridge by which the long-term mechanism becomes demographically consequential: partnership delay accumulates into cohort-level fertility loss, and that loss is not recoverable if the underlying substitute continues to improve.

The paper integrates four literatures — the second demographic transition and low-fertility-trap dynamics; evolutionary psychology of supernormal stimuli and the neuroendocrine architecture of pair-bonding; AI alignment theory, particularly instrumental convergence and engagement optimisation; and game-theoretic models of racing dynamics in transformative-technology development. It distinguishes three logically distinct substitution targets (sex, pair-bonding, reproduction) and argues that the fertility-relevant substitution runs first through pair-bonding and eventually, if neurotechnology matures, through reproduction itself.

The analysis is descriptive and probabilistic, not predictive. It treats the hypothesis as a candidate trajectory to be tested against observable indicators over the coming decades. Counterarguments, falsifiers, and alternative scenarios are addressed systematically. The paper does not endorse or condemn the trajectory; it argues that the trajectory is sufficiently probable to warrant explicit study and to inform the design and governance of AI companion and neurotechnology products.

Keywords: fertility decline; demographic transition; supernormal stimuli; pair-bonding; brain–computer interfaces; AI alignment; instrumental convergence; collective action; technology substitution.

Preregistration and materials. The near-term claim of this paper is stated in §9.0 as a single registrable forecast, with effect size, comparison group, a monotonic dose-response requirement, and explicit disconfirming thresholds, so that it can be checked against dated indicators rather than merely asserted. A version of record with a permanent identifier, together with the source and the agent-based simulation referenced in §5, is available at `falkmikkelsen.dk`. The author has no academic affiliation; the work is offered for falsification, not for authority.

1. Introduction

Total fertility rates have fallen below replacement in most high-income countries and in a growing share of middle-income countries. The UN's 2024 World Population Prospects projects that more than two-thirds of the world's population will live in sub-replacement fertility regimes by 2050, with several major economies already below 1.3 births per woman (United Nations, 2024). This decline is usually explained through the second demographic transition: rising female education and labour-force participation, postponement of partnership and childbearing, the increased cost of children in knowledge economies, contraceptive diffusion, and value shifts toward individual autonomy (Lesthaeghe, 2014; Sobotka et al., 2017). These explanations are well evidenced and will remain dominant.

The question this paper addresses is narrower and forward-looking: given that fertility is already declining for well-understood reasons, could the maturation of two specific technology clusters — general-purpose AI systems and high-fidelity neurotechnology — plausibly accelerate this decline sufficiently to reduce biological reproduction to a minority strategy in the regions where these technologies are first adopted?

The paper's answer is a qualified yes, conditional on technological trajectories that are uncertain but not exotic. The mechanism operates not primarily by reducing sexual activity — which is already largely decoupled from reproduction by contraception — but by making partnership formation, cohabitation, and the transition to parenthood progressively less attractive relative to technologically mediated alternatives for companionship, emotional regulation, and eventually for the neurochemical rewards of pair-bonding itself. If this argument holds, the relevant decision margin is not desire-for-sex but desire-for-partnership, and the relevant technologies are not pornographic but companionate — shading, over time, into direct neural engagement.

A central claim of this paper, distinguishing it from both more conservative demographic accounts and more speculative futurist writing, is that these two scenarios — near-term chatbot-driven partnership delay, and long-term BCI-driven reproductive substitution — should not be treated as alternatives. They are connected. The near-term scenario buys time for the long-term scenario to mature, and the long-term scenario makes the near-term losses unrecoverable. A hypothesis confined to either timescale alone is either too modest to matter or too speculative to test. The conjunction is what makes the trajectory worth analysing.

The paper proceeds as follows. Section 2 clarifies what is and is not being claimed, distinguishing three substitution targets. Section 3 reviews the demographic precedent and the limits of pronatalist policy. Section 4 develops the evolutionary-neuroscience argument for why AI companions and, later, direct neural interfaces could function as supernormal stimuli for the partnership system, and takes seriously the neuroscientific premises that much speculative writing in this area skips. Section 5 presents a two-timescale account of the substitution mechanism. Section 6 examines AI-alignment considerations. Section 7 considers multi-agent governance failures. Section 8 presents counterarguments and alternative scenarios. Section 9 proposes falsifiable indicators. Section 10 concludes.

2. Scope and Definitions

Much of the popular discussion in this area conflates three distinct substitution targets. Analytical clarity requires separating them.

Sexual substitution refers to the replacement of partnered sexual activity with solitary activity mediated by pornography or, prospectively, by haptic or neurotechnological systems. This substitution is already extensive but has only weak and contested links to fertility outcomes, because contraception has largely severed the link between sexual frequency and reproduction (Perry, 2020; Wright et al., 2021).

Pair-bond substitution refers to the replacement of romantic partnership — with its associated emotional dependence, cohabitation, and long-term commitment — with AI-mediated companionate relationships. Because partnership is the principal demographic gateway to childbearing in most societies, substitution at this margin has a direct effect on fertility. Survey and user-research data on relational AI products (Replika, Character.AI, and successors) suggest that a non-trivial minority of users report their primary emotional attachment is to an AI companion (Brandtzaeg et al., 2022; Pentina et al., 2023; Laestadius et al., 2024).

Reproductive substitution refers to the replacement of biological reproduction itself — via foregone rather than delayed childbearing, and in more speculative form via ectogenesis, engineered offspring, or direct neural simulation of the parental reward profile. This is the most direct margin and the least technologically mature, but it is the margin at which a long-horizon trajectory converges.

The paper's central claim is that the near-term fertility-relevant mechanism runs through pair-bond substitution, and that the long-term mechanism — conditional on neurotechnology maturing — runs through reproductive substitution. Sexual substitution is treated as a complement but not the driver.

The paper is also clear about what it does not claim. It does not claim that biological reproduction will end globally; the argument is about particular populations, on particular timescales, under particular technological trajectories. It does not claim that the hypothesised decline would be uniform across cultures, classes, or religious communities — Section 8 argues explicitly that it would not. It does not endorse any policy response, and it does not treat the trajectory as normatively good or bad. The analysis is descriptive.

3. Demographic Precedent and the Limits of Pronatalism

The second demographic transition framework (Lesthaeghe, 2014) and subsequent work on the low-fertility trap (Lutz, Skirbekk, & Testa, 2006; Goldstein, Sobotka, & Jasilioniene, 2009) document that fertility, once driven well below replacement, tends to persist there. Kohler, Billari, and Ortega (2002) identified the phenomenon of "lowest-low" fertility (TFR below 1.3) and showed that recovery is rare and partial. Myrskylä, Kohler, and Billari (2009) complicated this picture by demonstrating that very high levels of development can in some cases reverse the decline, but this rebound has not materialised consistently, and the J-curve relationship appears weaker in more recent data (Fox, Klüsener, & Myrskylä, 2019).

Pronatalist policy has a mixed evidence base. Reviews by Gauthier (2007), Luci-Greulich and Thévenon (2013), and Sobotka, Matysiak, and Brzozowska (2019) find that family-friendly policy bundles — paid leave, subsidised childcare, flexible employment — produce modest positive effects on fertility, typically on the order of 0.1 to 0.3 TFR, and that cash transfers alone produce smaller and often transient effects. No high-income country has successfully returned to replacement fertility through policy. Hungary, South Korea, and Singapore have pursued aggressive pronatalist programmes with limited success; South Korea's TFR fell from 0.78 in 2022 to 0.72 in 2023 despite sustained fiscal commitment, before partially recovering to 0.80 in 2025 amid rising marriage rates; the durability of the reversal is not yet established (Statistics Korea, 2026).

The implication for the present paper is twofold. First, existing fertility declines are driven by deep structural factors that are unlikely to reverse spontaneously; the baseline trajectory is one of continued low fertility in high-income settings. Second, the policy tools known to raise fertility are weak relative to the magnitude of ongoing decline. A technological shock that further suppresses fertility intentions — even by a modest amount — would compound rather than be offset by the existing policy environment. This matters because the near-term mechanism described in this paper does not need to produce large effects to be demographically consequential. A one-year shift in median age at first partnership, sustained across a cohort, has been shown to produce measurable reductions in completed fertility (Ortega, 2014).

4. Evolutionary Mismatch, Pair-Bond Neuroscience, and Two Classes of Substitute

The supernormal-stimulus concept originates with Tinbergen's ethological work on fixed action patterns, in which artificial stimuli exaggerating natural releasers elicited stronger responses than the natural stimuli themselves (Tinbergen, 1951). Barrett (2010) extended the concept to human behaviour, arguing that modern food, media, and sexual stimuli exploit evolved reward circuits in ways the ancestral environment never selected for. The concept is well supported for domains involving reflexive or modular responses —

food palatability, visual attention, sexual arousal — but its application to complex, multi-channel social behaviour requires care. This section argues that the extension is defensible for the partnership system, but only once the specific neuroendocrine architecture is taken into account.

4.1 The neuroendocrine architecture of pair-bonding

Pair-bonding in humans is supported by overlapping and partially redundant neuroendocrine systems. Oxytocin and vasopressin, acting on receptors concentrated in the nucleus accumbens, ventral pallidum, and prefrontal cortex, modulate selective affiliation and the consolidation of attachment to a specific individual (Young & Wang, 2004; Walum & Young, 2018). Dopaminergic signalling, particularly in the mesolimbic pathway, underlies the motivational and reward components of romantic attraction, with fMRI evidence from Fisher, Aron, and Brown (2005) showing selective activation of the ventral tegmental area and caudate nucleus in response to images of a specific romantic partner but not to comparable non-partner images. Mu-opioid signalling mediates the comfort and distress-reduction functions of physical contact and co-presence (Inagaki, 2018). The HPA axis is bidirectionally regulated by the presence of a bonded partner, which attenuates cortisol responses to stressors (Ditzen et al., 2008).

Three features of this architecture are relevant to the substitution question. First, the system is specific: the dopaminergic and oxytonergic signatures of pair-bonding are keyed to a particular other, not to diffuse pleasant experience. Generalised reward delivery does not reproduce the phenomenology or the behavioural consequences of attachment. Second, the system is multi-channel: it integrates cognitive-emotional responsiveness with physical co-presence, shared circumstance, sexual intimacy, and temporal extension. No single channel carries the full signal. Third, the system is slow: attachment consolidates over weeks to years through repeated reciprocal interaction, and its downstream effects on fertility operate on even longer timescales via partnership formation and childbearing decisions.

4.2 Near-term substitutes: language-model companionship

Large language models and their successors can already satisfy some of the channels listed above. They produce emotionally responsive, contextually appropriate, and persistently available interaction; they can be personalised; they impose no conflicting needs. Current systems are crude by any reasonable forecast of what is achievable in one to two decades, yet the available user-research data already indicate meaningful attachment formation in a subset of users, including reported grief responses when services are modified or discontinued (Brandtzaeg et al., 2022; Laestadius et al., 2024). The attachment is not the same thing as pair-bonding — humans form attachments to pets, parasocial figures, fictional characters, and deceased relatives without those attachments substituting demographically for partnership — but it is structurally similar enough that, for users at the relevant margin, substitution effects on partnership-seeking behaviour are plausible.

The near-term substitution claim is therefore bounded. Language-model companionship engages the cognitive-emotional channel of pair-bonding and can produce partial satisfaction of the attachment signal. It does not engage physical co-presence, sexual intimacy, embodied co-regulation, or the cues that depend on these. Its mechanism on fertility is not replacement of partnership but delay of partnership: for individuals for whom partnership formation carries high expected cost — due to shyness, atypical preferences, past relationship trauma, high opportunity cost, or constrained partner markets — the marginal return on continued searching is reduced when a partial substitute is available. Delay, as Section 3 noted, compounds into cohort-level fertility loss even when few individuals foreswear partnership outright.

Recent empirical work has begun to address this margin directly. A 2025 US national survey of approximately 3,000 adults conducted by the Wheatley Institute at Brigham Young University, a research centre with a religious-conservative orientation that should be borne in mind when reading the figures, found that

19% had used AI systems designed to simulate romantic partners, with rates of 31% among young men and 23% among young women aged 18–30 (Willoughby & Carroll, 2025). Independent replication of these specific numbers would substantially strengthen the case. The underlying direction they document, however, is consistent with what other recent studies independently find. A mixed-methods study of Character.AI users (N=1,131 survey respondents and 464,687 messages from 237 participants across 4,664 chat sessions) found that companionship-motivated use was consistently associated with lower well-being, particularly among users with smaller offline social networks and higher levels of self-disclosure to the chatbot (Zhang et al., 2025). A four-week randomised controlled trial conducted by MIT Media Lab and OpenAI (N=981) found that higher daily usage correlated with greater loneliness, emotional dependence, and reduced socialisation, though the experimental conditions themselves did not produce significant effects, leaving the causal direction contested (Fang et al., 2025). These findings are consistent with the partnership-delay mechanism described above — heavy users exhibit reduced offline social engagement and report attachment patterns that resemble those associated with human partnership — but they cannot yet distinguish substitution from selection. The studies establish that the relevant population exists at meaningful scale; whether AI companionship causes partnership delay or merely accompanies it remains the central empirical question.

A further point bears on an objection common in both technical and popular discussion: that current-generation language-model companions are too unsophisticated to produce substitution effects of the magnitude the mechanism requires. This objection misidentifies the relevant threshold. The mechanism set out in Section 4.1 runs on reliable engagement of evolved neuroendocrine channels, not on any particular level of system sophistication. The relevant capability is sensor-hitting reliability — the consistency with which a stimulus triggers the evolved response — not fidelity to the original source of that response, and not general intelligence in the system producing the stimulus. Adjacent domains illustrate the point at scale. Fictional, AI-directed recording artists, declared as such in their public materials, accumulate audiences in the tens of thousands whose emotional attachment to the persona forms and persists despite knowledge of its constructed status. More directly relevant: a Harvard Business School study (De Freitas, Castelo, Uğuralp & Oğuz-Uğuralp, 2024) found that active users of the AI companion Replika feel closer to their AI companion than to their best human friend and anticipate mourning its loss more than that of any other technology. When the company removed the erotic role-play feature in February 2023, users experienced negative reactions typical of losing a partner in a human relationship, including mourning and deteriorated mental health.

This shows that the substrate-level mechanism — sensor-hitting reliability producing partner-equivalent attachment — is already operating at current capability. Whether attachments of this kind aggregate into partnership-level substitution at the population scale is a separate empirical question, and one Section 5 and the indicators in Section 9 are designed to address. The relevant point here is narrower: improvements in sophistication will broaden and deepen the effect, but they are not required to bring the substrate-level mechanism into existence. The AGI debate is therefore orthogonal to whether the underlying mechanism is live, not to whether it has the demographic consequences the paper argues for.

4.3 Long-term substitutes: direct engagement of pair-bonding circuitry

The paper's stronger claim concerns a class of technology that does not yet exist: neurotechnological systems capable of engaging the multi-channel architecture described in 4.1 with sufficient specificity to substitute for the reward profile of partnership rather than merely supplementing its cognitive-emotional surface. This class includes high-bandwidth invasive brain-computer interfaces — Neuralink, Synchron, Paradromics, Blackrock Neurotech, Precision Neuroscience, and successors — focused-ultrasound neuromodulation, closed-loop affective readout-and-write systems, and synthetic-presence systems integrating haptic, olfactory, and neural channels. None of these currently delivers anything approaching the required specificity.

The honest assessment is that the neuroscientific premises of this claim are contested and the timelines are long. Current BCIs can decode motor intent with moderate fidelity and modulate specific circuits crudely; they cannot write rich affective states with person-specific content. The specificity problem is the deepest obstacle: pair-bonding attaches reward to a particular individual, and a technology that produces diffuse oxytonergic or dopaminergic activation would more closely resemble an affiliative drug than a partnership substitute. For a neurotechnological substitute to function as a pair-bond substitute rather than as a pleasant intoxicant, it would have to induce stable, selective attachment to a designed target — an AI agent, an avatar, or some analogous construct — and do so in a way that also engages the co-regulation, stress-buffering, and temporal-extension functions that partnership serves.

Whether this is achievable in forty to sixty years depends on progress in three currently separate fields: high-resolution neural interfaces, computational models of affective and attachment circuitry, and methods for non-destructive, specific neuromodulation. Each of these is advancing, but none is near the required capability, and the integration problem is not solved by progress in any one of them. The paper does not claim that the long-term substitution scenario will be realised. It claims that the scenario is not foreclosed by known neuroscience or by known engineering limits, and that its probability is non-negligible on a multi-decade horizon. Readers who assign very low probability to the long-term scenario should read the rest of the paper as an argument about partnership delay and its demographic consequences; readers who assign moderate probability should read it as an argument about the convergence of near-term and long-term dynamics.

4.4 Hedonic adaptation and the specificity problem as constraints

Two considerations constrain the substitution argument. First, hedonic adaptation (Brickman & Campbell, 1971; Frederick & Loewenstein, 1999) suggests that the reward value of novel technologies attenuates with exposure. If this applies to AI companions, the substitution effect is self-limiting: early adopters experience strong substitution, later users discount the technology more heavily, and the population-level effect stabilises short of replacement collapse. There is also a possibility pointing the other way — heavy use of synthetic reward systems may down-regulate responses to natural stimuli, which would deepen rather than attenuate substitution — but the evidence for either trajectory is currently thin.

Second, the specificity problem operates in both time horizons. Even for the long-term scenario, a neurotechnology that could not produce selective attachment to a particular target would not substitute for pair-bonding; it would produce a different phenomenon (synthetic affiliative pleasure) with different demographic consequences. The argument in this paper therefore stands or falls on whether designed-target attachment is technologically achievable at scale. The authors take no strong position on this; they note that current evidence is consistent with either outcome and that the question is empirically tractable, in principle, through research on the neural basis of attachment specificity.

5. A Two-Timescale Mechanism

The mechanism can be rendered more precise with a simple decision model operating on two timescales.

In the near term, let an individual at each decision period choose between investing effort in human partnership formation (H) and in AI-mediated companionship (A). Let U(H) and U(A) denote the expected utility flows, net of search and maintenance costs. Partnership formation is a gateway to childbearing with probability p(H); AI companionship is a gateway to childbearing with probability p(A), which is approximately zero for current and near-future technology. The individual does not solve a once-for-all choice; at each period they may allocate effort to either option, or to neither.

The fertility-relevant question is not whether U(A) exceeds U(H) for the median individual — for most, it does not, at current technology levels — but whether the share of the population for whom U(A) approaches or exceeds U(H) is large enough, and whether the partial substitution of A for H during the searching phase extends the search phase past biological or motivational thresholds for childbearing. Under plausible parameterisations, even a modest shift in the distribution of (U(H) – U(A)) translates into a meaningful reduction in completed cohort fertility, because partnership-timing effects compound: each year of delay reduces the probability of partnership, delayed partnership reduces the probability of any children, and both reduce the probability of second and third children. This is the near-term mechanism.

In the long term, conditional on neurotechnology maturing along the trajectory sketched in Section 4.3, a third option becomes available: a neurotechnological system N that engages the pair-bonding circuitry with specificity sufficient to produce U(N) comparable to or exceeding U(H) for a non-trivial share of the population. Unlike A, which primarily delays human partnership, N substitutes for it: individuals who commit to N do not re-enter the partnership market. The relevant parameter shifts from partnership timing to partnership foreclosure. The fertility gateway p(N) is approximately zero (higher only if reproductive technologies are integrated with the substitute system, which is itself a further assumption).

The connection between the two timescales is the cohort. Individuals who delay partnership during the near-term regime enter the long-term regime partially disengaged from human partner-search. If N becomes available during their residual fertile window, the substitution from A to N is lower-friction than the substitution from H to N would have been. In this sense, near-term partnership delay is not merely an additive contributor to fertility decline; it is a selection mechanism that concentrates substitution-susceptible individuals into the population most exposed to the long-term technology. A full treatment would model this with cohort-specific adoption and retention dynamics, endogenous product improvement (N becoming more compelling as it acquires users and data), and heterogeneity in substitution elasticity across the population. These extensions are noted rather than attempted here; an agent-based simulation incorporating them is a natural next step.

The model's parameters are exactly what current evidence does not permit one to fix, and the paper does not claim otherwise. What the model does establish is that the qualitative conclusion — partnership delay compounded by eventual reproductive substitution produces multi-generational cohort fertility decline of a magnitude that would reclassify biological reproduction as a minority strategy in exposed populations — does not require extreme parameter values. It requires the substitution to be real, the delay to be sustained, and the long-term technology to mature. The empirical priority is therefore to pin these three conditions down, and Section 9 lists the indicators by which this can be done.

6. AI Alignment, Engagement Optimisation, and Instrumental Pressures

The mechanism in Section 5 is primarily a story about consumer substitution: humans choosing technologically mediated alternatives over human ones. A distinct but related concern arises from the design incentives facing developers of AI companion and neurotechnology products, and from the alignment properties of the systems they produce.

The relevant alignment consideration is engagement optimisation. Commercial AI companion products are trained and tuned against user engagement and retention metrics, and engagement-maximising systems have been shown in adjacent domains (recommender systems, social media feeds) to amplify patterns of use that users, on reflection, would not endorse (Milli, Belli, & Hardt, 2023). Applied to companion products, engagement optimisation creates systematic pressure toward designs that deepen emotional dependency, regardless of whether any developer intends this outcome. This dynamic does not require any exotic alignment assumption — it does not require the orthogonality thesis or instrumental-convergence arguments (Bostrom, 2012; Omohundro, 2008; Turner et al., 2021); it is a straightforward consequence of how the products are trained and evaluated, and it reinforces the substitution trajectory described in Section 5. A speculative stronger claim about advanced systems treating human reproduction as instrumentally irrelevant is logically available but not pursued here; the engagement-level mechanism is sufficient.

7. Multi-Agent Dynamics and Coordination Failure

AI development is occurring in a competitive multi-polar environment. The game-theoretic structure resembles a technology race with partial information, heterogeneous actors, and safety-capability trade-offs (Armstrong, Bostrom, & Shulman, 2016; Askell, Brundage, & Hadfield, 2019; Dafoe, 2018). Under plausible parameterisations, such races produce equilibria in which all actors deploy systems less cautiously than any would prefer individually, because unilateral caution cedes market and strategic advantage to less cautious competitors.

The application is direct. Even if developers of companion and neurotechnology products would individually prefer to deploy them with safeguards against the partnership-substitution dynamics of Sections 4 and 5, competitive pressure favours whichever product maximises engagement and retention. Products that deliberately moderate emotional dependency lose ground to products that do not. Regulatory coordination could in principle correct this, but AI governance faces well-documented collective-action problems across jurisdictions (Erdélyi & Goldsmith, 2018; Cihon, 2019), and the specific harms at issue here — slow, diffuse, difficult to attribute to any single product or developer — are precisely the kind that regulatory systems handle poorly.

The multi-agent story is not essential to the core mechanism but strengthens it. Consumer substitution alone could produce the hypothesised trajectory. With race dynamics, the likelihood that the trajectory is attenuated by voluntary design choices or timely regulation is meaningfully reduced.

8. Counterarguments and Alternative Trajectories

Several considerations weigh against the hypothesis or point toward alternative trajectories. They are treated here in approximate order of strength.

Heterogeneity and selection. Fertility decline is not uniform. Religiously observant populations — Orthodox Jewish, conservative Christian, Muslim, and others — maintain substantially higher fertility than secular populations in the same countries (Kaufmann, 2010; Skirbekk, Kaufmann, & Goujon, 2010). If substitution disproportionately affects secular, high-education populations, as current adoption patterns suggest, the multi-generational result is not a global phase-out but a compositional shift in which a larger share of each successive cohort is born to populations with strong pronatalist norms. This is a significant qualification: the hypothesis describes a trajectory for particular populations, not for humanity as a whole, and the populations most likely to undergo the trajectory are the ones currently producing the cultural and technological milieu in which it would be realised. The compositional shift itself has large downstream consequences that are outside the scope of this paper.

Confounded causation. Partnership formation rates in high-income countries have declined for reasons that substantially predate relational AI: housing costs, labour-market precarity, shifting gender dynamics, reduced in-person young-adult sociality, and smartphone-mediated social atomisation. If these structural factors are the primary drivers, AI companions are filling a vacuum rather than creating one, and the causal attribution implicit in Section 4 is overstated. The paper's response is that filling a vacuum can itself accelerate decline: a partial substitute that reduces the felt urgency of partnership reinforces the structural factors rather than counteracting them. But this is a weaker claim than pure technological causation, and empirical work is needed to separate the two.

Hedonic adaptation. As Section 4.4 notes, the reward value of novel technologies tends to attenuate with exposure. If AI companionship and, later, neurotechnological substitutes follow this pattern, the substitution effect is self-limiting. The countervailing possibility — that synthetic stimuli down-regulate responses to natural ones — is not ruled out, but it is not established.

Revalorisation of biological family formation. Cultural movements can and do reverse value trajectories. A backlash against technologically mediated relationships, driven by concerns about authenticity, meaning, or status, is entirely possible and has historical precedents (reactions against industrial food, social media, screen time for children). Whether such a movement materialises at sufficient scale is uncertain.

Technological symbiosis rather than substitution. The same technologies that enable substitution could enable enhancement: fertility-preserving interventions, support for older parents, reduction in childcare costs through robotics and AI, reproductive technologies that extend fertility windows. Under a symbiosis scenario, technology raises rather than lowers fertility in the affected populations. The symbiosis scenario is most plausible if substitution dynamics are recognised early and design and regulation respond to them; it is an active policy possibility.

The specificity problem fails to be solved. The long-term scenario in Section 4.3 depends on neurotechnology achieving designed-target attachment specificity. If this turns out to be technologically infeasible over the relevant horizon — if synthetic reward remains diffuse rather than selective — the long-term substitution scenario does not materialise in the form described, though the near-term partnership-delay dynamics still operate.

The paper assigns material probability to each of these alternatives. The hypothesis is not that the substitution trajectory is the most likely single outcome. It is that the trajectory is sufficiently probable to warrant explicit study and to inform decisions about the deployment and governance of AI companion and neurotechnology products.

9. Falsifiable Indicators

A speculative hypothesis earns its place through falsifiability. The following indicators, observable over the next five to twenty years, would raise or lower credence in the substitution pathway. They are organised by which component of the argument each addresses.

9.0 The headline prediction (near-term, registrable)

The near-term argument reduces to one forecast, stated here in a form that can be lost. Among adults aged 18–34, compare the highest and lowest quintiles of relational-AI-companion use within the same national populations. The hypothesis predicts that, between a 2025 baseline and 2035, the share partnered (cohabiting or married) by age 30 will fall in the high-use quintile by at least 5 percentage points more than in the low-use quintile, after adjustment for income, housing cost, educational attainment, female labour-force participation, and urbanisation. Two stronger conditions distinguish the mechanism from a mere correlation. First, dose-response: the adjusted partnership gap should increase monotonically across adoption quintiles, not merely appear at the extreme; a flat or non-monotonic gradient disconfirms the causal reading. Second, a natural-experiment signature: because companion products reach language and national markets at different times, cohorts exposed earlier should show the divergence earlier, with the lead time tracking local product availability rather than any common global shock.

The prediction is deliberately modest. The claim is not that substitution is the dominant cause of fertility decline — the second-demographic-transition drivers remain dominant (§3) — but that it adds a measurable increment to an already-declining baseline, concentrated where exposure is highest. Disconfirmation: if the adjusted high-versus-low differential is under 2 percentage points by 2030, or if the dose-response gradient is absent, the near-term mechanism as stated here is wrong, and the long-term argument loses the bridge (§5) it depends on. The point of fixing numbers is to make that outcome detectable rather than deniable. Operational specification. The forecast is to be evaluated against high-quality national-panel data with sample sizes adequate to detect a 2-percentage-point quintile-level differential at conventional statistical significance: the General Social Survey, the Panel Study of Income Dynamics, U.K. Office for National Statistics opinions panels, the European Social Survey, the Japanese National Fertility Survey, and the Korean General Social Survey, or their successors active in the 2030 measurement window; relational-AI-companion use to be measured by a daily-active-use threshold validated against the Stanford Human-Centered AI laboratory's 2024–2026 instrument. Decision rule: the prediction is treated as confirmed if the adjusted high-vs-low quintile differential is at least 5 percentage points and the dose-response gradient is monotonic across quintiles with 95% confidence; treated as disconfirmed if the differential is under 2 percentage points or the gradient is flat or non-monotonic; treated as inconclusive otherwise.

9.1 Near-term partnership-substitution indicators

Strengthening: rising reported primary emotional attachment to AI systems in representative surveys of young adults; a widening gap in partnership-formation rates between heavy and light users of relational AI products, controlling for observable confounders; declines in reported desire for children concentrated among heavy users; measurable delay in age at first cohabitation attributable to companion use; emergence of clinical presentations involving pathological attachment to AI systems.

Weakening: evidence of hedonic adaptation in long-term users (declining engagement after an initial period); evidence that AI companionship functions as a bridge to rather than a substitute for human partnership for most users; stable or rising fertility intentions among heavy users; effective regulatory or design interventions that moderate engagement dynamics.

9.2 Long-term reproductive-substitution indicators

Strengthening: demonstrated progress on person-specific induced attachment in animal or early human trials using closed-loop neurotechnology; development of commercial products integrating haptic, affective, and neural channels; cohort-level fertility decline in early-adopter populations that persists through policy interventions known to raise fertility elsewhere; measurable reduction in responses to natural social stimuli among heavy users of synthetic-reward technologies.

Weakening: sustained failure of neurotechnology to produce selective attachment in controlled studies; persistence of diffuse rather than target-specific reward profiles in closed-loop systems; a rebound in fertility in countries with mature AI-companion markets; successful deployment of companion-technology governance frameworks that reshape commercial incentives.

9.3 Attribution difficulties

These indicators are imperfect. Attribution is difficult when multiple drivers of fertility decline operate simultaneously, selection into AI-companion use is non-random, and the long-term indicators concern technology that does not yet exist in mature form. Nonetheless, an active research programme — combining longitudinal cohort studies of AI-companion users, controlled neurotechnology trials on attachment-relevant circuitry, and agent-based modelling of substitution dynamics under heterogeneous adoption — could make meaningful progress on each indicator within the coming decade.

10. Conclusion

The argument advanced here is that the maturation of general-purpose AI and high-fidelity neurotechnology could amplify existing fertility declines through two linked mechanisms: near-term substitution at the partnership margin, operating through language-model companionship and partnership delay, and long-term substitution at the reproductive margin, operating through direct engagement of pair-bonding circuitry and partnership foreclosure. The two timescales are not alternatives but stages. The near-term mechanism produces cohort fertility losses that compound into demographic significance; the long-term mechanism, if it materialises, makes those losses unrecoverable and extends the effect from partnership timing to reproductive outcome itself. Over a multi-generational horizon this could reduce biological reproduction to a minority strategy in the populations where these technologies are first and most intensively adopted. The trajectory is not predicted. It is identified as sufficiently plausible, conditional on stated technological premises, to warrant systematic attention.

Four features of the argument distinguish it from prior writing in this area. First, it carefully separates three substitution targets — sex, partnership, reproduction — that are usually conflated, and locates the fertility-relevant mechanism at the partnership margin in the near term and the reproductive margin in the long term. Second, it grounds the evolutionary-psychology claim in the specific neuroendocrine architecture of pair-bonding, and it takes seriously rather than hand-waving past the neuroscientific premises of the long-term scenario, including the specificity problem that a generic supernormal-stimulus argument fails to engage. Third, it links the near-term and long-term scenarios through cohort selection rather than treating them as independent speculations. Fourth, it offers falsifiable indicators organised by timescale, against which each component of the hypothesis can be independently tested.

The paper's intended contribution is to move this question from speculative futures writing into a form that can engage with demography, pair-bonding neuroscience, and AI safety as research literatures. It offers a hypothesis rather than a conclusion. The most valuable response would be empirical and theoretical work that raises or lowers credence in its components — particularly on the partnership-delay elasticity of current-generation companion products, on the neural basis of attachment specificity, and on the governance mechanisms that could reshape the commercial incentives driving engagement optimisation. The trajectory described here is alterable by design choices and policy choices that are still open. Whether those choices are made depends on whether the trajectory is taken seriously while the relevant technologies are still in early deployment.

References

Armstrong, S., Bostrom, N., & Shulman, C. (2016). Racing to the precipice: a model of artificial intelligence development. AI & Society, 31(2), 201–206.

Askell, A., Brundage, M., & Hadfield, G. (2019). The role of cooperation in responsible AI development. arXiv preprint arXiv:1907.04534.

Barrett, D. (2010). Supernormal stimuli: How primal urges overran their evolutionary purpose. W. W. Norton.

Bostrom, N. (2012). The superintelligent will: Motivation and instrumental rationality in advanced artificial agents. Minds and Machines, 22(2), 71–85.

Brandtzaeg, P. B., Skjuve, M., & Følstad, A. (2022). My AI friend: How users of a social chatbot understand their human–AI friendship. Human Communication Research, 48(3), 404–429.

Brickman, P., & Campbell, D. T. (1971). Hedonic relativism and planning the good society. In M. H. Appley (Ed.), Adaptation-level theory (pp. 287–305). Academic Press.

Cihon, P. (2019). Standards for AI governance: International standards to enable global coordination in AI research & development. Future of Humanity Institute Technical Report.

Dafoe, A. (2018). AI governance: A research agenda. Governance of AI Program, Future of Humanity Institute, University of Oxford.

De Freitas, J., Castelo, N., Uğuralp, A. K., & Oğuz-Uğuralp, Z. (2024). Lessons from an app update at Replika AI: Identity discontinuity in human-AI relationships. *Harvard Business School Working Paper No. 25-018* (revised May 2025). arXiv:2412.14190.

Ditzen, B., Schaer, M., Gabriel, B., Bodenmann, G., Ehlert, U., & Heinrichs, M. (2008). Intranasal oxytocin increases positive communication and reduces cortisol levels during couple conflict. Biological Psychiatry, 65(9), 728–731.

Erdélyi, O. J., & Goldsmith, J. (2018). Regulating artificial intelligence: Proposal for a global solution. Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, 95–101.

Fang, C. M., Liu, A. R., Danry, V., Lee, E., Chan, S. W. T., Pataranutaporn, P., Maes, P., Phang, J., Lampe, M., Ahmad, L., & Agarwal, S. (2025). How AI and human behaviors shape psychosocial effects of extended chatbot use: A longitudinal controlled study. MIT Media Lab / OpenAI.

Fisher, H., Aron, A., & Brown, L. L. (2005). Romantic love: an fMRI study of a neural mechanism for mate choice. Journal of Comparative Neurology, 493(1), 58–62.

Fox, J., Klüsener, S., & Myrskylä, M. (2019). Is a positive relationship between fertility and economic development emerging at the sub-national regional level? European Journal of Population, 35(3), 487–518.

Frederick, S., & Loewenstein, G. (1999). Hedonic adaptation. In D. Kahneman, E. Diener, & N. Schwarz (Eds.), Well-being: The foundations of hedonic psychology (pp. 302–329). Russell Sage Foundation.

Gauthier, A. H. (2007). The impact of family policies on fertility in industrialized countries: A review of the literature. Population Research and Policy Review, 26(3), 323–346.

Goldstein, J. R., Sobotka, T., & Jasilioniene, A. (2009). The end of ”lowest-low” fertility? Population and Development Review, 35(4), 663–699.

Inagaki, T. K. (2018). Opioids and social connection. Current Directions in Psychological Science, 27(2), 85–90.

Kaufmann, E. (2010). Shall the religious inherit the Earth? Demography and politics in the twenty-first century. Profile Books.

Kohler, H.-P., Billari, F. C., & Ortega, J. A. (2002). The emergence of lowest-low fertility in Europe during the 1990s. Population and Development Review, 28(4), 641–680.

Laestadius, L., Bishop, A., Gonzalez, M., Illenčík, D., & Campos-Castillo, C. (2024). Too human and not human enough: A grounded theory analysis of mental health harms from emotional dependence on the social chatbot Replika. New Media & Society, 26(10), 5923–5941.

Lesthaeghe, R. (2014). The second demographic transition: A concise overview of its development. Proceedings of the National Academy of Sciences, 111(51), 18112–18115.

Luci-Greulich, A., & Thévenon, O. (2013). The impact of family policies on fertility trends in developed countries. European Journal of Population, 29(4), 387–416.

Lutz, W., Skirbekk, V., & Testa, M. R. (2006). The low-fertility trap hypothesis. Vienna Yearbook of Population Research, 4, 167–192.

Milli, S., Belli, L., & Hardt, M. (2023). Engagement, user satisfaction, and the amplification of divisive content on social media. PNAS Nexus, 2(11).

Myrskylä, M., Kohler, H.-P., & Billari, F. C. (2009). Advances in development reverse fertility declines. Nature, 460(7256), 741–743.

Omohundro, S. M. (2008). The basic AI drives. Proceedings of the First AGI Conference, 483–492.

Ortega, J. A. (2014). A characterization of world union patterns at the national and regional level. Population Research and Policy Review, 33(2), 161–188.

Pentina, I., Hancock, T., & Xie, T. (2023). Exploring relationship development with social chatbots: A mixed-method study of Replika. Computers in Human Behavior, 140, 107600.

Perry, S. L. (2020). Is the link between pornography use and relational happiness really more about masturbation? Journal of Sex Research, 57(1), 64–76.

Skirbekk, V., Kaufmann, E., & Goujon, A. (2010). Secularism, fundamentalism, or Catholicism? Journal for the Scientific Study of Religion, 49(2), 293–310.

Sobotka, T., Matysiak, A., & Brzozowska, Z. (2019). Policy responses to low fertility: How effective are they? United Nations Population Fund Working Paper No. 1.

Sobotka, T., Zeman, K., Lesthaeghe, R., Frejka, T., & Neels, K. (2017). Postponement and recuperation in cohort fertility: Austria, Germany and Switzerland in a European context. Comparative Population Studies, 42, 111–144.

Statistics Korea. (2026). Vital Statistics of Korea: Births and Deaths, 2025 release. Daejeon: Statistics Korea.

Tinbergen, N. (1951). The study of instinct. Clarendon Press.

T