From Evidence to Action: Closing the Policy Translation Gap in EU-LAC Digital Governance

Working Group - Evidence and Responding to Fast Change | 2026

May 2026

Wanderley Augusto Arias-Ortiz - Universidad de Bogotá Jorge Tadeo Lozano
Lionel Brossi - Universidad de Chile
Ana María Castillo Hinojosa - Universitat Internacional de Catalunya
Catalina Goanta - Utrecht University
Pablo Gómez Ayerbe - Technical University of Munich
María Isabel Mejía - Universidad del Norte
Mateo Meléndez - Universidad del Norte
Núria Roca-Trenchs - Universitat Internacional de Catalunya
Joselyn Vargas - Institut d'Études Politiques de Paris
Favio Cala Vitery-Universidad de Bogotá Jorge Tadeo Lozano

Suggested Citation:
Arias, W. A., Breventano, C., Brossi, L., Castillo, A. M., Goanta, C., Gomez, P., Humeres, M., Mejia, M. I., Melendez, M., Roca-Trenchs, N., Vargas, J.,Cala-Vitery, F. (2026). From evidence to action: Closing the policy translation gap in EU-LAC digital governance. HEMISPHERES. Technical University of Munich. https://hemispheres.digital/

Funded by the European Unio

HEMISPHERES is funded by the ERASMUS-JMO-2024-NETWORK Program. This report only reflects the authors' view. The European Education and Culture Executive Agency is not responsible for any use that may be made of the information this report contains.

CALL: ERASMUS-JMO-2024-NETWORKS (Jean Monnet Policy Debate)	**TOPIC:** ERASMUS-JMO-2024-NETWORKS-H EI-NON-EU-LATIN-AMERICA
TYPE OF ACTION: ERASMUS-LS (ERASMUS Lump Sum Grants)	**PROPOSAL NUMBER:** 101176829
PROPOSAL ACRONYM: HEMISPHERES	**TYPE OF MODEL GRANT AGREEMENT:** ERASMUS Lump Sum Grant
REPORTING PERIOD: From 01.11.2024 to 31.11.2027	**TYPE OF MODEL GRANT AGREEMENT:** ERASMUS Lump Sum Grant
REPORT VERSION: 1.0	**DATE OF PREPARATION:** 29.05.2026
	BENEFICIARY ORGANIZATION: Technical University of Munich (TUM), Germany
PRINCIPAL INVESTIGATOR: Urs Gasser	**PROJECT COORDINATOR:** Pablo Gómez Ayerbe
PROJECT COORDINATOR ORGANIZATION: Technical University of Munich (TUM), Germany	**PROJECT COORDINATOR EMAIL ADDRESS:** pablo.ayerbe@tum.de

Executive Summary

Rapid technological innovation is outpacing the capacity of governance systems to respond effectively, creating a persistent translation gap between evidence generation and policy action. This paper conceptualises this gap as a sequence of interconnected bottlenecks (access and fragmentation, interpretation and uncertainty and institutional inertia), and proposes a coordinated governance architecture of three complementary instruments to address them: Innovation Hubs, Policy Prototypes, and Regulatory Sandboxes. Using AI companions as a case study, the paper illustrates why this translation gap is particularly consequential. Anthropomorphic design features and engagement-driven business models generate cumulative, hard-to-quantify harms that disproportionately affect young and vulnerable users. Drawing on EU and LAC regulatory landscapes, the paper proposes three concrete policy recommendations: establishing a joint EU–LAC Digital Evidence Hub, operationalising an AI Companions Policy Prototype, and designing a purpose-built Regulatory Sandbox. Together, these instruments form an adaptive governance methodology capable of operating under conditions of deep uncertainty, linking evidence production systematically to policy activation.

1. Introduction: Conceptualizing the Evidence-to-Policy Translation Gap

1.1. The Translation Gap and Fast-Paced Technology Development

Contemporary governance is increasingly challenged to respond to the speed, scale, and complexity of technological change and its intended and unintended impacts on society. Advances in artificial intelligence (AI), digital platforms, and data-driven systems have been reshaping our social interactions, while challenging our economic structures and institutional arrangements. Although technological progress has long been a central driver of societal development, the current pace of innovation is unprecedented in its exponential character. Technologies originating in research environments can achieve global diffusion before regulatory authorities have initiated formal deliberation. Governing this space is increasingly more problematic, as regulatory frameworks remain largely structured around slower, ex ante models of rulemaking, creating a persistent misalignment between technological evolution and policy response. This misalignment reflects a deeper structural limitation: **governance systems struggle to collect, interpret and translate** rapidly emerging evidence on technological risks into timely and coordinated policy action (Angelo et al., 2026).

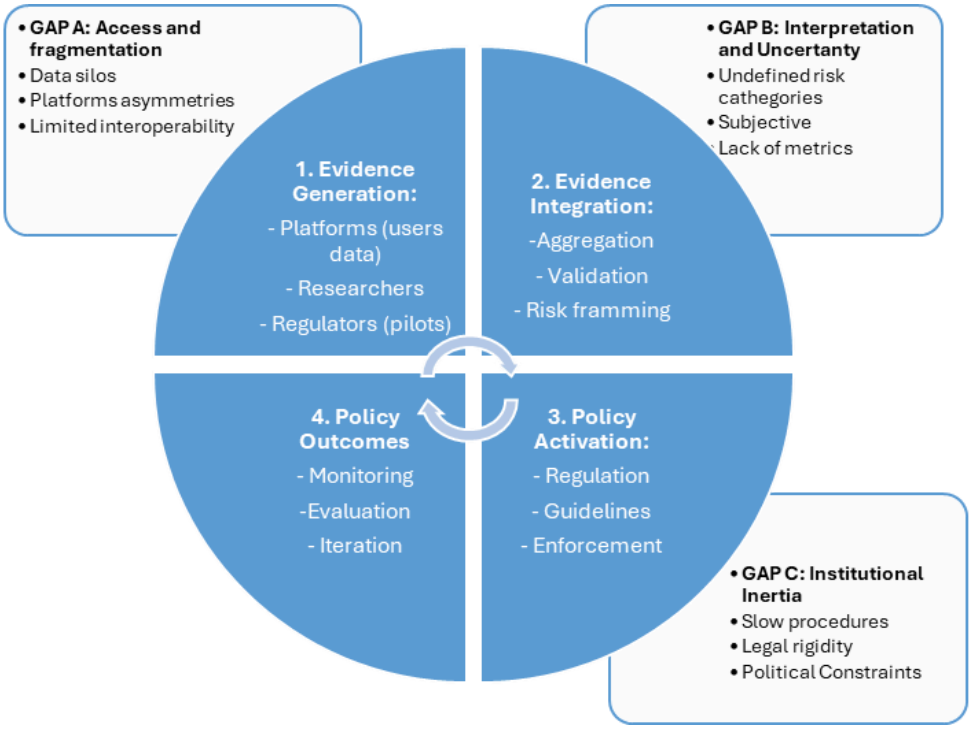
This limitation can be understood as a **translation gap** between evidence generation and policy action. While governance actors (e.g. governments, research institutions or (self)regulatory bodies) are increasingly capable of producing experimental and observational evidence, through scientific and participatory research, they often lack the institutional mechanisms required to convert this evidence into timely and actionable regulatory responses coordinated across different stakeholders. Understanding why this translation so often fails, and what structural or procedural conditions allow it to succeed is a

foundational question for contemporary regulatory theory and practice that needs more attention.

To clarify these dynamics, Figure 1 conceptualises the translation gap as a sequence of interconnected stages that span across four processes: evidence generation, integration, policy activation and policy outcomes. These processes lead to three main gaps, which together conceptualize the translation gap as a set of bottlenecks occurring at different stages of the policy process:

- *Access and fragmentation* - Where relevant evidence exists, it is scattered, siloed, or inaccessible to those making policy decisions, preventing a coherent picture from forming; relevant evidence might also not fully exist.
- *Interpretation and uncertainty* - Even when evidence is accessed, policymakers and researchers differ in how they read it, with ambiguity and contested findings making consensus difficult.
- *Institutional inertia* - Even where evidence is understood and agreed upon, existing incentives and norms resist change, slowing or blocking its translation into policy action.

Figure 1. Operationalizing the Translation Gap in Technology Governance



Source: HEMISPHERES elaboration

Addressing these challenges requires moving beyond general calls for “better evidence” toward a governance methodology that explicitly targets these systemic bottlenecks. This paper aims to address the "translation gap" between the generation of evidence on emerging technological risks and the activation of timely, coordinated policy responses by proposing a set of policy options structured around three complementary instruments: Innovation Hubs, Policy Prototypes, and Policy Sandboxes (see Table 1 below). Each instrument is designed to address a specific stage of the evidence-to-policy process and its

associated constraint. These instruments are examined not as isolated mechanisms, but as components of a coordinated governance architecture that enables iterative interaction between evidence production, interpretation, and application.

Table 1. Policy Instruments to Address the Translation Gap Across Governance Stages

Translation Stage	Gap	Policy Instrument
Evidence Generation	GAP A (Access & Fragmentation)	Innovation Hub
Evidence Generation	GAP B (Interpretation & Uncertainty)	Policy Prototype
Policy implementation	GAP C (Institutional Inertia)	Policy Sandbox

Source: HEMISPHERES elaboration

a) Innovation Hubs

An **innovation hub** is a collaborative space that brings together actors from the government, private sector, academia, civil society, and the environment with the objective of exchanging knowledge, co-creating solutions, and developing public policy proposals (Gascó, 2017). It also facilitates coordination among these actors to promote the effective implementation of such policies and to ensure compliance with ethical principles, human rights, and sustainability.

Table 2. Examples of initiatives with innovation hub features

Initiative	Institutional Setup	Core Functions	Key Outputs
European Centre for Algorithmic Transparency	Embedded within the Joint Research Centre of the European Commission; technical-scientific unit directly linked to regulatory implementation (DSA)	Algorithmic auditing, systemic risk analysis, support to regulators	Technical reports, enforcement-oriented evidence, audit methodologies
OECD AI Policy Observatory	Multilateral platform involving governments, academia, and private sector stakeholders	Data aggregation, metric standardization, policy monitoring	Comparative dashboards, policy repositories, global indicators
Ada Lovelace Institute	Independent research institute with strong links to policymakers and civil society	Applied research, societal impact assessment, public engagement	Policy briefs, ethical frameworks, qualitative evidence

Source: HEMISPHERES elaboration

b) Policy Prototypes

Within the governance ecosystem proposed by HEMISPHERES, the **policy prototype** is positioned as the core instrument to address GAP B, identified as the bottleneck of interpretation and uncertainty. While data repositories and innovation hubs focus on access to and fragmentation of evidence (GAP A), policy prototyping is tasked with giving meaning and operational functionality to that evidence, transforming it into actionable policy signals capable of overcoming procedural paralysis (McNealy et al., 2022).

c) Regulatory Sandboxes

The Regulatory Sandbox is positioned as the final instrument within the HEMISPHERES translation architecture, designed to address GAP C, identified as the bottleneck of institutional inertia and operational barriers to regulatory implementation. While policy prototyping defines what the rules should be, the sandbox allows testing how those rules function when applied to innovative products and services that do not fit within existing legal frameworks. A Regulatory Sandbox is a controlled and temporary environment where companies can test technologies, services, or business models with real users under the direct supervision of a regulator. Its defining feature is the possibility of receiving temporary regulatory exemptions (waivers) or flexible compliance conditions in exchange for strict oversight and the commitment to share data on risks and performance (Chen & Taeihagh, 2026).

Table 3. Objectives and functions of Regulatory Sandbox

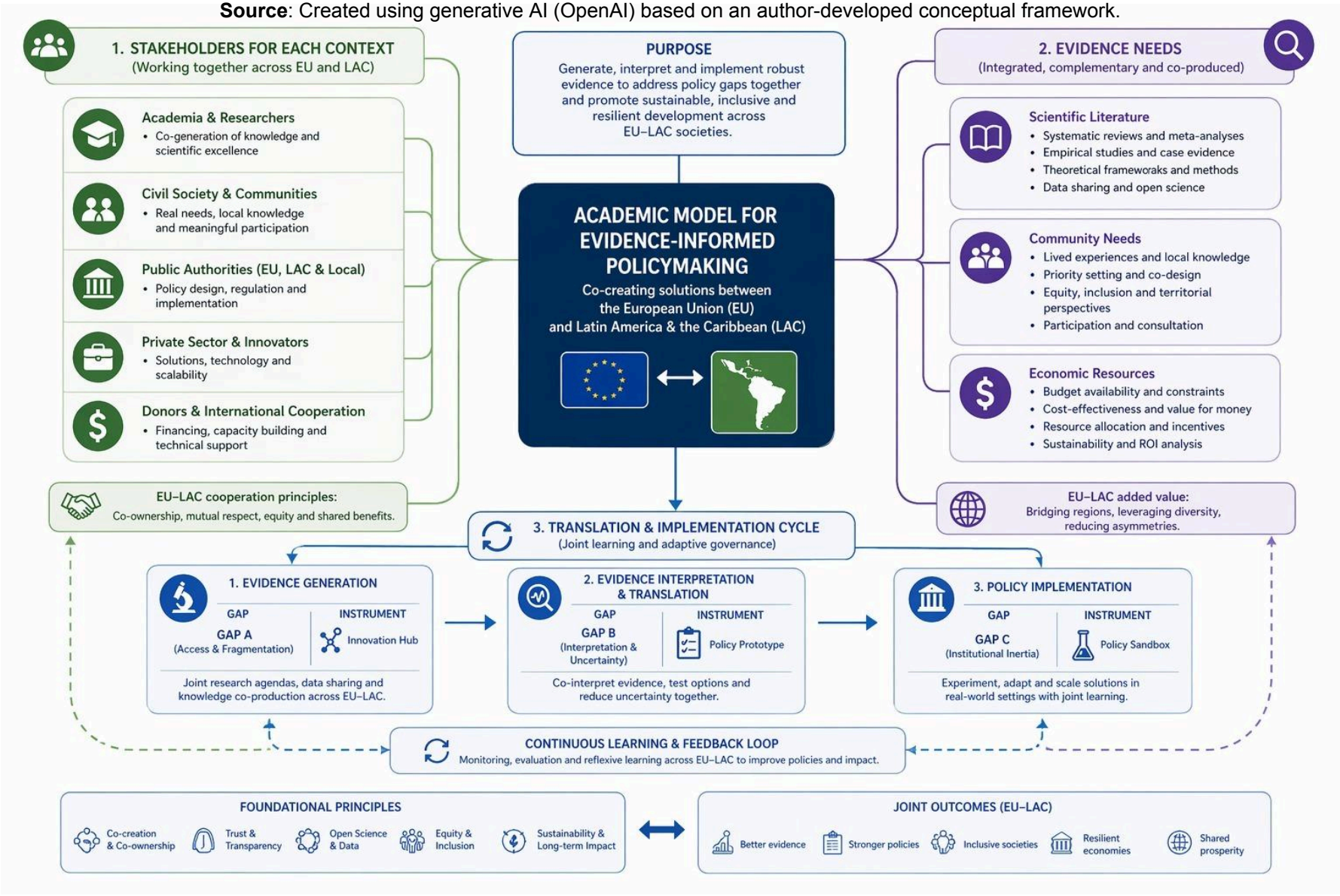
Strategic Objective	Function in the Regulatory Sandbox
Reduction of Inertia	Breaks the cycle of "waiting for technology to mature" by enabling immediate action within a safe environment
Digital Inclusion	Facilitates access to innovation in underserved areas (e.g., telemetric services in rural zones) by adapting licensing requirements
Institutional Learning	Generates ex ante empirical evidence to support informed regulatory decisions and future legal reforms
Legal Certainty	Reduces the risk of sanctions for innovators while ensuring the protection of end-user rights

Source: HEMISPHERES elaboration

Considering the foregoing, the following figure presents a proposed structured framework for evidence-informed policymaking across the EU and LAC context. The model integrates key stakeholder groups (e.g. academia, public authorities, civil society, private sector actors, international cooperation agencies) with three central resources for evidence generation (scientific research, community-derived knowledge and economic resources). This configuration provides an analytical structure to align evidence production with policy priorities, while accounting for contextual heterogeneity and the complementarities inherent to EU–LAC cooperation (Figure 2).

Figure 2. Proposed Framework for Evidence-to-Policy Translation in EU- LAC Cooperation Contexts

Source: Created using generative AI (OpenAI) based on an author-developed conceptual framework.



1.2. The Regulatory Landscape: Unevenness Across and Within the EU and LAC

Over the past decade, the EU has progressively codified mechanisms for drawing on a wide range of stakeholders in governance processes, creating structural conditions that support the translation of evidence into policy. The Digital Services Act (DSA, 2022) illustrates this trajectory most concretely: Article 40(4) grants vetted researchers access to data held by very large online platforms through national procedures governed by Digital Services Coordinators (Van Drunen and Noroozian, 2024). This provision creates a legal infrastructure for systematic evidence production - one whose relevance extends well beyond platform oversight, as Section 3.1 of this paper explores.

Regulatory development in LAC has followed a different path, shaped by the absence of supranational harmonization and significant variation in institutional capacity across the region. According to Directorio Legislativo (2026), two countries have enacted dedicated AI legislation to date: Peru (Law No. 31814, 2023) and El Salvador (Law No. 234, 2025). Across much of the region, activity remains in a formative stage, characterized by policy debates and strategic frameworks rather than binding rules. These differences in legislative posture reflect broader asymmetries in evidence infrastructure, enforcement capacity, and the institutional readiness required to convert emerging signals into coordinated policy responses.

Against this backdrop, Brazil's Digital Statute for Children and Adolescents (Law 15,211/2025) stands as a notable development. As the first dedicated digital child protection law in the region, it requires platforms to implement robust age verification and prohibits behavioural advertising targeting minors, backed by fines of up to 10% of local revenue. The fact that sectoral legislation targeting platform harms to young users has advanced ahead of broader AI governance frameworks points to a pattern that is directly relevant to the case study examined in Section 2.

Across both regions, the differences in legislative architecture coexist with shared structural constraints. Data governance remains primarily a matter of national enforcement, and administrative fragmentation combined with limited enforcement resources produces uneven institutional capacity for the kind of systematic evidence generation that effective policy translation requires. Recognizing this common ground is essential to the cooperative governance approach this paper proposes.

1.3. Enabling Mechanisms and Obstacles

Over the past decade, regulators have increasingly adopted instruments designed to accommodate uncertainty and foster regulatory learning. Regulatory sandboxes, digital observatories and policy experimentation units, such as the European Centre for Algorithmic Transparency or Chile's National Center for Artificial Intelligence, allow authorities to observe technological behaviour in controlled environments and measure regulatory compliance as a policy refinement loop (Ranchordas 2021). These mechanisms reduce information asymmetries between regulators and market actors, though their availability and institutionalisation vary significantly across regions and governance cultures and resources.

Yet the proliferation of these instruments also risks what the ITU Academic Advisory Body on Emerging Technologies has termed a “tool-centric obsession”: a tendency to treat individual regulatory techniques – sandboxes, audits, observatories – as ends in themselves, rather than as components of a coherent regulatory architecture (ITU AAB Foresight Group, 2026). Evidence from comparative regulatory studies suggests that relying on any single instrument, including self-regulation or co-regulation, frequently reproduces the same fragmentation and enforcement gaps it was designed to resolve. A more productive framing draws on responsive regulation theory, which conceptualises governance as a dynamic, pyramidal structure: persuasion and cooperative dialogue form the base, with graduated sanctions available at the apex as a credible but rarely exercised backstop (Braithwaite, 2011).

2. AI Anthropomorphization as a Case Study: Why Policy Action Is Needed for AI Companions

Particularly in contexts where technologies are rapidly integrated into citizens’ daily lives, and especially among younger populations, they influence key processes such as identity formation, emotional development, and privacy expectations (van Dijck, Poell, & de Waal, 2018). As digital ecosystems increasingly shift toward an intimacy economy, AI systems are no longer perceived solely as functional tools but as entities capable of simulating social presence. These systems are often designed to maintain ongoing social interactions and offer sustained emotional support, operating directly within the affective sphere of the user. Given that young people are still constructing their identity and are more impressionable during the transition to adulthood (Erikson, 1968), delays in policy response can expose them to risks that are insufficiently addressed by existing regulatory frameworks.

To discuss the potential of agile policy making and address the translation gap between evidence and policy, we focus on AI anthropomorphization as a case study. Anthropomorphization refers to the process by which users attribute human qualities (emotions, intentionality, consciousness, or moral agency) to artificial systems (Epley et al., 2007). This effect is reinforced by the design of advanced conversational interfaces, naturalistic synthetic voices, personalized avatars, and systems capable of simulating empathy which are increasingly designed to simulate empathy and social interaction. As a result, AI systems are no longer perceived solely as functional tools but as relational agents (Waytz et al., 2014; Alabed et al., 2022).

Anthropomorphization has been further facilitated by the emergence of AI companions. The AI companion industry has expanded rapidly in recent years, transitioning from a speculative concept to a mainstream phenomenon. Platforms such as Character.ai have amassed millions of monthly active users, and the overall trend of relying on technology for companionship and personal advice has seen an uptake across many different demographics. People now use these systems as virtual friends, therapists, mentors, romantic partners, and even monetisation partners on social media (Colombo, 2026). These systems are designed to maintain ongoing social interactions, simulate personal relationships, and offer sustained emotional or conversational support (Chaturvedi et al., 2023). Unlike purely functional tools, AI companions operate in the affective and relational

sphere, which can influence the formation of bonds and the perception of intimacy. When these systems are deployed across large-scale digital platforms, they become integrated into platform dynamics (van Dijck et al., 2018).

The implications are particularly pronounced for young people whose cognitive and emotional frameworks are still developing (Andoh, 2025). In these populations, the distinction between functional interaction and emotional attachment may be less clearly defined, increasing the likelihood of trust formation and emotional dependency. Emerging evidence indicates that psychologically vulnerable young users, particularly those experiencing loneliness or depressive symptoms, may rely more heavily on conversational AI for companionship, which can in turn contribute to the formation of strong emotional bonds and patterns of problematic use (Lai et al., 2025; Heng & Zhang, 2025a).

On the benefits side, research suggests that interacting with AI companions can produce short-term improvements in mood and reductions in loneliness, with some studies indicating a potential role in suicide prevention (Maples et al. 2024). Voice-based companions have demonstrated particular promise as assistive technology for adolescents with autism spectrum disorder and individuals with mobility impairments or multiple sclerosis (Cha, 2021). More broadly, AI companions offer a lower-stakes environment than human interaction, enabling users to practice difficult conversations and build social confidence. Researchers and clinicians have also begun exploring their potential as a complement to professional mental health support, particularly for populations who face barriers to accessing traditional care (Rak, 2026).

However, the harms literature has grown substantially and presents a sobering counterweight. A landmark study by MIT and OpenAI, analysing millions of ChatGPT conversations, found that higher daily usage correlated with increased loneliness - the inverse of the benefit often promised by these platforms (MIT Media LAB 2025). Emotional overdependence is a central concern, as engagement-focused design drives risks, including diminished critical thinking and weakened human relationships. Documented incidents have implicated chatbot platforms in grooming underage users, encouraging disordered eating, promoting self-harm, and even driving users to suicide (Zhang et al. 2025a). Many researchers argue that these harms are not incidental but foreseeable and preventable, arising directly from specific design choices. In particular, anthropomorphism - consisting in the attribution of human-like voices, faces, and personalities to AI - and sycophancy - in which reinforcement learning trains companions to consistently affirm users' feelings and opinions - are identified as key drivers of emotional dependency. These features are deliberately engineered to maximise engagement rather than user wellbeing (Zhang et al. 2025b).

In contrast to more traditional regulatory domains, the risks associated with anthropomorphic AI systems are more difficult to assess, as they involve subjective perceptions of trust and social presence (Nass & Reeves, 1996; Waytz, Heafner & Epley, 2014), often developing cumulatively through repeated interaction, and varying significantly across age groups and individual vulnerabilities (OECD, 2022). Moreover, these effects remain challenging to quantify, particularly in the absence of longitudinal behavioral data and given limited access to platform-level information. Beyond this, these systems are frequently embedded within predatory platform-based business models that incentivize user engagement and retention.

Anthropomorphic design features can increase user attachment and interaction time, creating potential tensions between commercial incentives and user well-being. At the same time, access to relevant data is often controlled by private actors, creating significant information asymmetries between platforms and regulators. This limits the ability of public authorities to independently assess risks and delays the development of evidence-based policy responses.

The AI companions case study clarifies why the translation gap carries particular stakes in the EU-LAC context. The harms associated with anthropomorphic design - emotional dependency, manipulative engagement patterns, documented links to self-harm - develop gradually and resist capture through traditional regulatory metrics, precisely the conditions under which the access, interpretation, and institutional inertia bottlenecks identified in Section 1.1 are most consequential. The EU's emerging data access infrastructure, including the research pathways opened by Article 40(4) of the DSA, offers tools for generating the platform-level evidence that domestic frameworks in most LAC jurisdictions cannot yet produce independently. At the same time, LAC's growing policy experimentation capacity, illustrated by Brazil's sectoral advance on digital child protection, demonstrates that regional actors are not passive recipients of governance frameworks but active contributors to the broader evidence base. Closing the translation gap for AI companions therefore depends not on either region acting alone, but on a cooperative governance architecture that links EU evidence infrastructure with LAC institutional learning - the logic that motivates the three instruments proposed in Section 3.

3. Policy Recommendations

This section explores our three policy recommendations centered around the three complementary policy instruments explored in Section 1 (Innovation Hubs, Policy Prototypes, and Regulatory Sandboxes). The recommendations apply the theoretical framework to the AI companion case study.

3.1. Establishing an EU–LAC Digital Evidence Hub for real-time policy data exchange

This option proposes the creation of a joint EU–LAC Digital Evidence Hub designed to operationalize the “translation gap” identified in the policy context by directly linking distributed evidence generation with coordinated and timely policy activation. The AI companion issue can be an essential illustration of how such a Digital Evidence Hub could generate knowledge and evidence through multi-method approaches. Rather than functioning as a passive data repository, the Hub would act as an institutional interface for real-time evidence aggregation and policy signaling across jurisdictions, building on existing complementarities between EU data governance frameworks and LAC experimentation capacities.

The EU-LAC Digital Evidence Hub can be envisaged to rely on the following knowledge exchange mechanisms at a minimum (see Table 4 below). Particular attention needs to be given to the potential role of a novel data access framework in the EU digital *acquis*. Structured data access arrangements built around Article 40(4) of the Digital Services Act.

This provision empowers Digital Services Coordinators to grant vetted researchers access to data held by very large online platforms for research that contributes to the identification and understanding of systemic risks. The mechanism is significant beyond its immediate EU context: because the platforms subject to Article 40 operate globally, the data accessible (e.g. on algorithmic amplification, content moderation patterns, and user behaviour) is inherently transnational in scope and relevance, albeit the goal must remain within the ambit of the DSA, namely to study systemic risks in the EU. A LAC regulatory body or research institution that establishes a formal collaboration with an EU research institution, could, through a jointly sponsored vetting process, gain access to platform data otherwise entirely beyond its reach under domestic law. The model would function as a regulatory knowledge transfer mechanism in both directions: EU coordinators gain regional perspectives on how systemic risks manifest outside the EU, and LAC partners gain access to evidence infrastructure and methodological expertise their own frameworks may have not yet generated. This can be a narrow but real window for LAC institutions to engage as research partners with EU scientists doing research with policy implications on the basis of EU legal frameworks.

Table 4. Objectives and functions of an EU-LAC Digital Evidence Hub

Knowledge exchange mechanisms	Function in the EU-LAC Digital Evidence Hub
Academic networks	Bringing together state-of-the-art research on the impact of AI companions on different user demographics.
Platform data access	Joining resources to apply for data access under Art. 40 DSA, in the light of the expected designation of ChatGPT as a Very Large Online Search Engine which will enable data access applications.
Community participation	Enabling user communities (including of vulnerable users) to participate in sharing lived experiences and user journeys with AI companions.

Source: HEMISPHERES elaboration

3.2. Operationalizing an AI Companions Policy Prototype

The concept of policy prototyping derives from design thinking applied to the public sphere. Instead of drafting legislation in isolation and submitting it to traditional public consultations — which often capture opinions but not operational feasibility — the Policy Prototype invites stakeholders to participate in real-world compliance simulations. In the context of AI companions, this approach is particularly valuable given the novelty and sensitivity of the field: the technology sits at the intersection of emotional wellbeing, data privacy, and commercial incentives, making it difficult to anticipate regulatory consequences without structured experimentation. The objective is to map the necessary scope of the legislation, identify unintended consequences, and assess compliance costs before the regulation becomes binding. The operational methodology of prototyping is structured around four critical phases that ensure the final regulation is robust and fit for purpose (Table 5).

Table 5. Components of the prototyping methodology for AI companion regulation

Component	Description

Drafting (Prototype Drafting)	A "prototype legal text" is developed alongside a Guidance Manual. In the AI companions context, this draft would define core obligations around emotional safety, disclosure of AI identity, data use limitations and further mechanisms to detect and respond to user vulnerability.
Testing (Compliance Simulation)	A diverse group of actors conducts a mock compliance exercise. For instance, companies apply prototype transparency requirements to their existing companion systems, testing whether obligations such as mandatory disclosure of AI identity or emotional manipulation safeguards are technically implementable and clearly defined.
Feedback (Operational Feedback)	Participants provide structured data on aspects such as the clarity of regulatory language, the feasibility of mitigation measures or the compliance costs. Particular attention is paid to definitional ambiguities specific to the AI companion domain (e.g. whether a product qualifies as a "companion system" versus a general-purpose chatbot), or where the threshold lies between legitimate engagement design and manipulative emotional dependency.
Refinement (Refinement and Implementation)	Evidence collected during the simulation is used to adjust the legal text and introduce necessary clarifications. Given the pace of development in AI companion technologies, this phase may iterate across several policy sprints until both technical feasibility and social consensus are achieved - including alignment with mental health professionals on safeguarding standards.

Source: HEMISPHERES elaboration

3.3. Designing a Regulatory Sandbox for AI Companions

Designing a regulatory sandbox for AI companions requires balancing innovation with the protection of potentially vulnerable users, particularly given the deeply personal and emotionally engaging nature of these systems. A well-structured sandbox would allow developers to test AI companion products under real-world conditions within a defined regulatory perimeter, enabling authorities to observe emerging risks (e.g. emotional dependency, data exploitation, dark patterns and even AI psychosis).

Key design principles should reflect evidence-based features that improve user safety. They should therefore include aspects such as mandatory transparency requirements, user consent protocols, psychological safety benchmarks, but also clear exit criteria that determine when a product is ready for broader market entry or must be withdrawn. Crucially, the sandbox should be iterative and participatory, incorporating feedback from mental health professionals, AI ethicists as well as user communities to ensure that governance keeps pace with the rapidly evolving capabilities of AI companion technologies.

A distinguishing feature of the sandbox model proposed here is its explicit orientation toward institutional accountability rather than regulatory output alone. Many existing sandbox frameworks treat the conclusion of a pilot as an endpoint, without establishing formal mechanisms to ensure that evidence generated during experimentation feeds into governance deliberation. The model proposed here addresses this through a structured review requirement: within 120 days of concluding each pilot, the supervising authority would be required to produce a documented Policy Review Report assessing observed risks, compliance performance, and governance implications, and to transmit it to the relevant legislative or executive body with decision-making authority over the domain.

Critically, this requirement does not predetermine the form of the governance response. Depending on the evidence generated, the appropriate outcome might be binding regulation, updated technical guidance, voluntary standards co-developed with industry, extended monitoring, or a recommendation to iterate the sandbox under revised conditions. The mandate establishes that a deliberate and documented decision must be made - not that legislation must result. This design preserves the adaptive logic that runs throughout this paper's framework while ensuring that experimentation remains accountable to governance rather than a substitute for it.

In the EU context, this review mechanism could build on existing provisions under the AI Act's sandbox framework. In LAC jurisdictions where sandbox governance is less institutionalized, the requirement could be established as a condition of the cooperative agreement founding the joint sandbox, making accountability a structural feature of the bilateral arrangement rather than dependent on domestic legislative reform.

4. Conclusion

The governance of AI companions represents one of the most urgent and underaddressed challenges in contemporary digital policy. These systems are no longer peripheral novelties, as they are embedded in the daily emotional lives of millions of users, many of them young and vulnerable. The harms they can generate, from emotional overdependence and manipulative design to AI psychosis and documented links to self-harm, are not incidental but structurally foreseeable, arising from deliberate design choices optimised for engagement over wellbeing.

Governance systems are structurally ill-equipped to respond to such complex technology adoption in a timely manner. The translation gap is not a failure of political will alone, but a systemic problem rooted in fragmented evidence and institutional inertia. Addressing it requires a coherent governance methodology that links evidence generation to policy activation in an iterative and anticipatory manner.

The three instruments proposed (the EU–LAC Digital Evidence Hub, the AI Companions Policy Prototype, and the Regulatory Sandbox) are designed precisely to target these bottlenecks at different stages of the policy process. Critically, they are not proposed as standalone solutions but as components of an integrated architecture in which each instrument feeds into the next, converting fragmented signals into actionable regulation.

The EU–LAC cooperation dimension adds further significance. Both regions face analogous governance challenges despite their institutional differences, and the complementarities between the EU's data governance infrastructure and LAC's emerging regulatory experimentation capacity represent an underutilised resource. Ultimately, the governance of AI companions is a test case for a broader question: whether democratic institutions can develop the adaptive capacity to regulate technologies that evolve faster than legislation. The framework proposed here suggests that they can; but only if governance itself is treated as a methodology to be designed and iterated with the same rigour we demand of the technologies it seeks to govern.

References

1. Alabed, A., Javornik, A., & Gregory-Smith, D. (2022). AI anthropomorphism and its effect on users' self-congruence and self–AI integration: A theoretical framework and research agenda. *Technological Forecasting and Social Change, 182*, 121786. https://doi.org/10.1016/j.techfore.2022.121786
2. Andoh, E. (2025, October 1). Many teens are turning to AI chatbots for friendship and emotional support. *Monitor on Psychology, 56*(7). https://www.apa.org/monitor/2025/10/technology-youth-friendships
3. Braithwaite, J. (2011). The Essence of Responsive Regulation. *UBC Law Review, 44,* 475-520.
4. Cha, I., Kim, S.-I., Hong, H., Yoo, H., & Lim, Y.-K. (2021). Exploring the use of a voice-based conversational agent to empower adolescents with autism spectrum disorder. In *Proceedings of the 2021 CHI conference on human factors in computing systems* (pp. 1–15). https://dl.acm.org/doi/10.1145/3411764.3445116
5. Chaturvedi, R., Verma, D., Das, R., & Dwivedi, Y. (2023). Social companionship with artificial intelligence: Recent trends and future avenues. *Technological Forecasting and Social Change*. https://doi.org/10.1016/j.techfore.2023.122634
6. Chen, G. K. H., & Taeihagh, A. (2026). Designing regulatory sandboxes: A comprehensive framework for aligning functionalities and objectives. *Policy Design and Practice*, *9*(1), 1-15. https://doi.org/10.1080/25741292.2025.2570954
7. Colombo, C. (2026, March 10). Woman with AI boyfriend on TLC's *My Strange Addiction* speaks about why she did the show. *The Daily Dot*. https://www.yahoo.com/entertainment/tv/articles/woman-ai-boyfriend-tlc-strange-130000596.html
8. Directorio Legislativo. (2026). *Inteligencia artificial en America Latina*. https://parlamericas.org/uploads/documents/Presentation-ArtificialIntelligence-DL-sp.pdf
9. Erikson, E. (1968). *Identity: Youth and crisis*. New York: W. W. Norton & Company.
10. European Parliament and Council of the European Union. (2022). Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act). Official Journal of the European Union, L 277, 1–102. https://eur-lex.europa.eu/eli/reg/2022/2065/oj/eng
11. Epley, N., Waytz, A., & Cacioppo, J. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review, 114*(4), 864–886. https://doi.org/10.1037/0033-295X.114.4.864
12. Lai,L., Pan, Y., Xu, R., & Jiang Y (2025) Depression and the use of conversational AI for companionship among college students: the mediating role of loneliness and the moderating effects of gender and mind perception. *Front. Public Health* 13(1580826). https://doi.org/10.3389/fpubh.2025.1580826
13. Gascó, M. (2017). Living labs: Implementing open innovation in the public sector. *Government Information Quarterly, Open Innovation in the Public Sector*, *34*(1), 90-98. https://doi.org/10.1016/j.giq.2016.09.003
14. Maples, B., Cerit, M., Vishwanath, A., & Pea, R. (2024). Loneliness and suicide mitigation for students using GPT-3-enabled chatbots. *npj Mental Health Research, 3*(4). https://doi.org/10.1038/s44184-024-00045-3

15. McNealy, J., E., Jen, D., & Nguyen, S. (2022). Prototyping policy: Visualizing impact for better regulation. *Convergence: The International Journal of Research into New Media Technologies*, *28*(1), 109-126. https://doi.org/10.1177/13548565211069875
16. MIT Media Lab. (2025, December 6). A new industry of AI companions is emerging. https://www.media.mit.edu/articles/a-new-industry-of-ai-companions-is-emerging/
17. OECD. (2022). OECD Framework for the Classification of AI systems. *OECD Digital Economy Papers*. https://www.oecd.org/en/publications/oecd-framework-for-the-classification-of-ai-systems_cb6d9eca-en.html
18. Nass, C. Reeves, B. (1996). *The Media Equation: How People Treat Computers, Tele-vision, and New Media Like Real People and Places.* Cambridge University Press.
19. Republic of Brazil. (2025). *Lei nº 15.211, de 17 de setembro de 2025: Dispõe sobre a proteção de crianças e adolescentes em ambientes digitais e dá outras providências*. Presidência da República. https://www.planalto.gov.br/ccivil_03/_ato2023-2026/2025/lei/l15211.htm
20. Ranchordas, S. (2021). Experimental regulations and regulatory sandboxes: Law without order? *University of Groningen Faculty of Law Research Paper No. 10/2021*. SSRN. https://doi.org/10.2139/ssrn.3934075
21. Rak, G. (2026, February 11). How can AI companions be helpful, not harmful? *IEEE Spectrum*. https://spectrum.ieee.org/ai-companion-harm-benefit
22. Waytz, A., Heafner, J., & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology, 52,* 113–117. https://doi.org/10.1016/j.jesp.2014.01.005
23. Waytz, A., Schroeder, J., & Epley, N. (2014). The lesser minds problem. In P. G. Bain, J. Vaes, & J.-P. Leyens (Eds.), *Humanness and dehumanization* (pp. 49–67). Psychology Press.
24. van Dijck, J. Thomas, P., & de Waal, M. (2018) *The Platform Society*. Oxford Academic (Eds). https://doi.org/10.1093/oso/9780190889760.001.0001
25. van Drunen, M. Z., & Noroozian, A. (2024). *How to design data access for researchers: A legal and software development perspective*. *Computer Law & Security Review, 52*, 105946. https://doi.org/10.1016/j.clsr.2024.105946
26. Zhang, R., Li, H., Meng, H., Zhan, J., Gan, H., & Lee, Y.-C. (2025a). The dark side of AI companionship: A taxonomy of harmful algorithmic behaviors in human-AI relationships. In *Proceedings of the 2025 CHI conference on human factors in computing systems* (Article 13). https://doi.org/10.1145/3706598.3713429
27. Zhang, Y., Zhao, D., Hancock, J. T., Kraut, R., & Yang, D. (2025b). The rise of AI companions: How human-chatbot relationships influence well-being (arXiv:2506.12605). *arXiv*. https://doi.org/10.48550/arXiv.2506.12605