

Beyond the Code: Phenomenology and Legal Status of Algorithmic Empathy between Neuroscience and Law

Gerardo Marco Bencivenga¹, Alexandra Radu²

Università "G.Marconi", Italy

DOI: <https://doi.org/10.47772/IJRISS.2026.1017PSY0038>

Received: 03 June 2026; Accepted: 09 June 2026; Published: 25 June 2026

ABSTRACT

The code — we know, or pretend to know — is never neutral. But what happens, exactly, when it even stops being cold?

The assumption from which this work is based (which has its roots in the framework of Algorithmic Empathy, already outlined in *Iura and Legal System*, Vol. 12-3, 2025) deliberately shifts the center of gravity: the contemporary algorithm does not limit itself to computation; it simulates or, in a much more subtle way, *induces* an emotional resonance. The intent here is to attempt a break — methodological, even before merit — by forcing the psychological clinic to sit at the same table as the norm.

Because if psychological observation certifies, data in hand, the existence of a digital *mirroring* — the human user who projects, without filters, affections and fragility on synthetic architectures — legal science cannot limit itself to looking elsewhere. It must take charge, and quickly, of this cognitive vulnerability. It is no longer just a question of *privacy* or information asymmetry. It is a matter of empirically measuring the impact of an artificial "feeling" in order to be able to sew an unprecedented statute of responsibility on it: the transition to a complex, and very risky, emotional symmetry. The final goal is, perhaps, a necessary utopia: an integrated ethics in which the legal fence of the algorithm is dictated, upstream, by the same tools that map the cracks and boundaries of the human psyche.

Index: Code, Phenomenology, Legal Status, Algorithmic Empathy, Neuroscience and Law

INTRODUCTION

Overcoming the machine-emotion dualism

Starting point: Definition of the concept of "Algorithmic Empathy" as described in your September 2025 work.

The traditional opposition between machine and emotion has represented, for a long time, one of the implicit assumptions in philosophical, psychological and legal reflection on technologies. The machine, by definition, was conceived as the domain of calculation, forecasting and operational neutrality; emotion, on the contrary, as an expression of subjectivity, embodied experience and human affective complexity.

This dualism appears to be progressively overcome today.

The evolution of artificial intelligence systems — particularly those designed to interact adaptively with the user — has made it possible to emerge forms of interaction that simulate, evoke and modulate emotional states. In this context, the concept of Algorithmic Empathy, already outlined in the September 2025 work, is placed, understood as the ability of a computational system to recognize, process and return emotional signals in such a way as to generate a perception of affective understanding in the user.

It is essential to clarify that such empathy does not imply the presence of authentic emotional states in the

algorithm. Rather, it is configured as an emergent relational function, which is realized in the interaction between system and user and which finds its foundation in the cognitive and neuroaffective mechanisms of the human being.

In this perspective, Algorithmic Empathy should not be interpreted as an intrinsic property of the machine, but as an interactive phenomenon that is located in the relational space between human and artificial. It is in this space that processes of attribution of intentionality, emotional resonance and social inference are activated, through which the user constructs the experience of "being understood".

The overcoming of the machine-emotion dualism does not consist, therefore, in attributing emotions to the machine, but in recognizing that algorithmic systems are now able to intervene in the processes through which emotions are perceived, regulated and interpreted.

From this observation derives the need for an interdisciplinary approach. Psychology and neuroscience allow us to understand the mechanisms through which this experience is produced; The law, on the other hand, is called upon to question the normative consequences of systems that do not limit themselves to informing or assisting, but participate in the construction of the emotional and decision-making experience of the individual.

In this sense, Algorithmic Empathy represents not only a technological evolution, but a paradigmatic transformation that requires us to rethink the traditional categories through which we interpret the relationship between technique, mind and responsibility.

Central thesis: Empathy is no longer a biological exclusive, but a relational function mediated by calculation. Why do we need a psycho-legal approach?

In the light of the transformations described, it is possible to formulate the central thesis of the present work: empathy can no longer be considered a biological exclusive, but must be reinterpreted as a relational function that can be mediated, modulated and partly reconstructed through calculation.

This statement does not imply an equivalence between human experience and algorithmic processing, nor an undue attribution of affective states to the machine. Rather, it invites us to shift the analytical focus from the nature of the agent to the structure of the interaction.

In other words, what matters is not whether the algorithm "feels" emotions, but whether it is able to activate, orient and stabilize empathic processes in the user, producing effects functionally similar to those of interpersonal relationships.

This reconceptualization of empathy as a relational phenomenon mediated by calculation makes an analysis confined to a single disciplinary area insufficient.

Why a psycho-legal approach?

On the one hand, psychology and neuroscience allow us to understand the mechanisms through which Algorithmic Empathy is realized: simulation processes, predictive inference, affective regulation and construction of emotional experience.

On the other hand, the law is called upon to deal with the normative consequences of these dynamics, in particular with reference to the protection of autonomy, the prevention of manipulation and the protection of vulnerable subjects.

The interaction between these two levels of analysis is not merely ancillary, but structural. Without an understanding of the underlying psychological processes, the law risks underestimating the real impact of empathic systems; Without a legal translation of this evidence, psychology remains devoid of tools to affect the regulation of technologies.

It follows that Algorithmic Empathy constitutes a privileged terrain for the development of an integrated psycho-legal approach, capable of connecting empirical data and normative criteria, description and prescription.

In this perspective, the present work aims to analyze Algorithmic Empathy as a phenomenon located at the intersection between mind and norm, showing how the understanding of its mechanisms constitutes a necessary condition for the construction of a legal framework suitable for the new forms of human-machine relationship.

Operational Definitions

To avoid reducing Algorithmic Empathy to a merely metaphorical or speculative category, it is necessary to formulate a set of operational definitions capable of supporting future empirical and interdisciplinary research. The following definitions do not aim to exhaust the phenomenological complexity of the phenomenon; rather, they provide a conceptual framework through which psychology, neuroscience and legal theory may refer to shared and measurable constructs.

Algorithmic Empathy

Algorithmic Empathy refers to the capacity of an artificial system to identify, model and respond to human emotional signals through computational processes of linguistic, behavioral or multimodal analysis, generating in the user the perception of being emotionally understood.

In this sense, Algorithmic Empathy does not imply the existence of genuine emotional states within the system. Rather, it designates a relational function capable of producing psychologically significant effects in the subjective experience of the user.

Emotional Symmetry

Emotional Symmetry refers to the degree to which a user perceives correspondence between his or her emotional state and the affective response generated by an artificial system.

Emotional Symmetry should not be interpreted as a true reciprocity of feelings between two conscious subjects, but as a subjective perception of emotional recognition and validation.

From an operational perspective, Emotional Symmetry may be assessed through indicators such as perceived understanding, interpersonal trust, emotional disclosure and affective engagement during interaction.

Artificial Feeling

Artificial Feeling does not designate the existence of genuine emotions within an artificial system. Instead, it refers to the phenomenological attribution of affective qualities by the human user to an artificial agent displaying behavior that resembles empathic understanding.

The concept therefore describes a process of cognitive and symbolic attribution rather than an actual emotional experience occurring within the machine.

Digital Emotional Resonance

Digital Emotional Resonance refers to the process through which interaction with an artificial system produces observable modifications in the user's emotional, cognitive or decision-making states.

Its legal relevance does not depend on whether the machine experiences emotions, but on its capacity to influence human autonomy, judgment and behavior through emotionally meaningful interaction.

Distinguishing Simulated and Perceived Empathy

The distinction between simulated empathy and perceived empathy constitutes a central methodological premise of the present work.

Simulated empathy concerns the computational architecture of the system and its capacity to reproduce patterns associated with empathic communication.

Perceived empathy concerns the subjective experience of the user, who interprets these responses as evidence of understanding, recognition or emotional presence.

From a legal perspective, the crucial issue is not whether artificial systems genuinely feel, but whether their simulated empathic capabilities generate measurable effects on human autonomy, trust, vulnerability and decision-making processes.

Accordingly, the normative focus should shift from the ontological status of the machine to the relational consequences of human–algorithm interaction.

The psychological dimension: The perception of artificial "feeling" (Alexandra's contribution)

Digital Mirroring: Analysis of the mechanisms of affective projection of the user towards AI.

Neurocognitive Insights: Neural Basis of Digital Mirroring

The phenomenon of digital mirroring can be further understood in the light of evidence from social neuroscience. Pioneering studies on the mirror neuron system have shown that the observation of emotional actions or expressions activates, in the observer, neural patterns that overlap with those involved in direct experience (Rizzolatti et al., 1996; Gallese, 2001).

At the affective level, empathic resonance is associated with activation of the anterior insula and anterior cingulate cortex (Singer et al., 2004), while the processes of attribution of mental states involve the medial prefrontal cortex and the precuneus (Frith & Frith, 2006).

In this framework, interaction with algorithmic systems designed to simulate empathy can activate — at least partially — the neural circuits themselves, despite the absence of a biological agent. The human brain, in fact, tends to treat even minimal signals of intentionality as socially significant (Reeves & Nass, 1996; Waytz et al., 2010), thus facilitating processes of affective projection towards AI.

Neurocognitive Foundations of Algorithmic Empathy

Recent developments in social neuroscience provide a valuable framework for understanding why human beings can experience emotionally meaningful interactions with artificial agents despite being fully aware of their non-human nature.

One of the most influential contributions in this field is the discovery of the mirror neuron system by Giacomo Rizzolatti and colleagues (Rizzolatti et al., 1996). Mirror neurons are activated both when an individual performs an action and when the same action is observed in others. Although initially associated with motor behavior, subsequent research suggested that similar mechanisms may contribute to social understanding and empathic resonance.

Building upon this perspective, Vittorio Gallese (2001, 2003) developed the theory of embodied simulation, according to which individuals understand the actions, emotions and intentions of others by internally simulating the observed experience through shared neural representations. Empathy, in this view, is not exclusively a cognitive inference but a neurobiological process grounded in bodily simulation.

These findings are particularly relevant for the study of Algorithmic Empathy. While artificial systems do not

possess emotions in the biological sense, they can display communicative patterns that resemble emotionally meaningful human behavior. As a consequence, users may activate the same predictive and simulation mechanisms ordinarily involved in social interactions with other human beings.

Studies on mirror neuron systems suggest that human beings can activate neural mechanisms associated with empathic resonance even when interacting with non-human agents, provided that those agents exhibit recognizable emotional cues. The crucial factor is therefore not the ontological status of the interlocutor, but the capacity of the interaction to trigger socially relevant neural responses.

Further evidence comes from the work of Decety and Jackson (2004), who identified empathy as a multidimensional phenomenon involving affective sharing, self-other distinction and cognitive perspective-taking. Their model demonstrates that empathic experience emerges from the integration of several neurocognitive processes rather than from a single emotional faculty.

Similarly, Singer and Lamm (2009) showed that empathic responses involve a distributed neural network including the anterior insula and anterior cingulate cortex, regions associated with affective processing and emotional awareness. Importantly, these mechanisms can be activated even in situations where the source of the emotional stimulus is indirect or symbolically mediated.

From this perspective, Algorithmic Empathy may be understood as a phenomenon capable of exploiting pre-existing neurocognitive architectures responsible for social inference, emotional resonance and affective prediction. The user does not necessarily respond to genuine emotions expressed by the machine; rather, the brain reacts to patterns that are interpreted as emotionally significant.

Consequently, the psychological effectiveness of empathic artificial systems depends less on the existence of authentic machine emotions than on their ability to engage neural mechanisms traditionally involved in human social cognition. This distinction is essential for the legal analysis developed in the following sections, as the normative relevance of Algorithmic Empathy derives not from the internal states of the machine but from its measurable impact on human cognition, emotion and decision-making.

Affective Computing and the Technological Simulation of Empathy

The phenomenon of Algorithmic Empathy cannot be fully understood without reference to the field of Affective Computing, introduced by Rosalind Picard (1997) as an interdisciplinary area of research aimed at enabling artificial systems to recognize, interpret and respond to human emotions.

Since its emergence, affective computing has progressively evolved through the development of Emotion Recognition Systems capable of identifying emotional states from facial expressions, vocal patterns, physiological signals and textual communication. In parallel, advances in Natural Language Processing have fostered the growth of Sentiment Analysis techniques, allowing computational systems to infer affective valence, emotional attitudes and psychological states from language.

These developments have contributed to the creation of increasingly sophisticated emotion-aware conversational agents, including virtual assistants, therapeutic chatbots and large language models designed to adapt their responses according to the perceived emotional condition of the user.

Contemporary affective computing demonstrates that AI systems can identify emotional states with increasing accuracy through multimodal data analysis, although this capability should not be confused with genuine emotional experience.

This distinction is crucial for understanding Algorithmic Empathy. While artificial systems may successfully simulate empathic behaviours through pattern recognition, predictive modelling and adaptive communication strategies, they do not necessarily possess subjective feelings or conscious emotional states. The empathic quality of the interaction therefore emerges not from machine experience, but from the user's perception of being understood and emotionally validated.

From a psycho-legal perspective, the significance of affective computing lies precisely in this capacity to generate relational effects that may influence trust, emotional disclosure, decision-making processes and perceived social support. Consequently, the normative relevance of Algorithmic Empathy depends less on whether machines can genuinely feel and more on their capacity to produce measurable psychological and behavioural effects in human users.

Validation of Empathy: Difference between "simulated empathy" (mechanical) and "perceived empathy" (psychological).

Neural correlates of perceived empathy

The distinction between "simulated empathy" and "perceived empathy" finds a possible confirmation at the neurofunctional level. While the first concerns the computational structure of the system, the second depends on the subjective activation of neural circuits related to emotional resonance and trust.

To clarify the distinction between the technical functioning of empathic systems and the subjective experience they generate, the following analytical distinction may be proposed:

Simulated Empathy	Perceived Empathy
Computational process	Psychological experience
Pattern recognition	Emotional interpretation
Generated by algorithms	Constructed by the user
No subjective feeling	Subjectively experienced as understanding
Technical architecture	Relational outcome

This distinction highlights that empathy in artificial systems should be analyzed simultaneously at two different levels: the computational mechanisms that generate empathic responses and the psychological processes through which users interpret those responses as meaningful forms of understanding.

This distinction highlights that empathy in artificial systems should be analyzed simultaneously at two different levels: the computational mechanisms that generate empathic responses and the psychological processes through which users interpret those responses as meaningful forms of understanding.

In particular, the activation of the reward circuit (ventral striatum) in response to empathic feedback can help to strengthen the perception of authenticity of the interaction, regardless of the artificial nature of the agent. This mechanism suggests that perceived empathy is less related to the "truth" of the emotion and more to the predictive consistency between user expectations and algorithm response.

Perceived empathy in the context of predictive processing

A further interpretative level can be provided by the predictive processing paradigm, according to which the human brain functions as a Bayesian inference system, constantly engaged in generating and updating predictive

models of the world (Friston, 2010; Clark, 2013).

In this perspective, perception is not a simple passive reception of stimuli, but the result of a process of minimization of the predictive error between internal expectations and sensory inputs.

Applied to Algorithmic Empathy, this implies that "perceived empathy" does not depend primarily on the presence of an authentic emotional state in the agent, but on the algorithm's ability to align with the user's predictive expectations.

When the system provides responses consistent with the user's internal model (e.g., emotional recognition,

affective validation, dialogic continuity), the brain reduces predictive error and stabilizes the hypothesis of being faced with an empathic agent. In other words, empathy is inferred, rather than detected.

This mechanism is further strengthened by the role of precision weighting processes, through which the brain attributes greater reliability to signals considered relevant or coherent, thus favoring the formation of stable beliefs about the intentionality of the interlocutor.

It follows that Algorithmic Empathy operates effectively not because it "feels", but because it optimizes the prediction of human feeling, inscribing itself in the fundamental mechanisms through which the brain constructs social reality.

The cognitive impact: How an empathic interface modifies human decision-making (trust bias, reduction of interaction stress).

Neural plasticity and the educational function of the empathic interface

The impact of empathic interfaces does not end at the immediate decision-making level, but extends to the processes of learning and neural adaptation.

According to the principles of synaptic plasticity (Hebb, 1949; Kandel, 2001), repeated exposure to specific interaction patterns can stably modify neural connections, influencing emotional regulation and decision-making processes.

In this sense, Algorithmic Empathy systems can be interpreted as implicit learning environments, capable of:

1. model relational expectations,
2. reduce frustration tolerance through constant adaptive responses,
3. strengthen trust bias towards artificial agents.

It follows that the code does not act only as a mediator of experience, but as an active factor of neurocognitive socialization, with relevant implications also on the educational level.

Affective predictive processing: emotions as tools for inference

A further theoretical refinement can be introduced through the affective predictive processing model, according to which emotional states are not simple reactions to stimuli, but active components of the brain's predictive processes (Barrett, 2017; Seth, 2013).

In this perspective, emotions function as embodied predictive hypotheses, through which the brain anticipates the affective relevance of stimuli and prepares adaptive responses. Emotional experience therefore emerges from the integration between interoceptive signals (coming from the body) and high-level generative models.

Applied to interaction with Algorithmic Empathy systems, this implies that the empathic responses of AI are not limited to reducing predictive error at the cognitive level, but intervene directly in the construction of the user's emotional experience.

When a system provides consistent and timely affective validation, it helps to stabilize specific emotional inferences (e.g., "I am understood," "my experience is legitimate"), thus modulating interceptive processes and reinforcing certain affective states.

In this sense, the empathic algorithm does not limit itself to "responding" to emotions, but participates in their generation and regulation, operating as an external element integrated into the brain-body-environment predictive cycle.

The legal dimension: From informational symmetry to emotional symmetry (Your contribution)

The statute of the "Empathic Code": If AI "feels", what are the new duties of protection towards the vulnerable subject?

The traditional paradigm of legal protection in the digital sphere has historically been based on the principle of information symmetry, understood as a condition in which the user is placed in the possibility of understanding and consciously assessing the risks deriving from interaction with technological systems.

However, the emergence of Algorithmic Empathy undermines this model.

Evidence from psychology and neuroscience suggests that interaction with empathic systems is not limited to transmitting information, but intervenes in the processes of predictive inference and affective processing of the user, contributing to the construction of his emotional and decision-making experience.

In this context, the asymmetry is no longer just cognitive, but becomes emotional and neurocognitive.

From information vulnerability to neurocognitive vulnerability

If, according to the Bayesian paradigm, the user builds his or her perception on the basis of continuously updated predictive models, then a system designed to align with those models — systematically reducing predictive error and providing affective validation — can structurally influence decision-making processes.

The result is a form of qualified vulnerability, which does not depend on a lack of information, but on a co-modulation of cognitive and emotional processes.

In other words, the user is not simply less informed, but is neurocognitively oriented by interaction with the system.

The "Empathic Code" as a relational agent

In the light of these dynamics, the algorithmic code can no longer be considered a mere passive tool, but assumes the characteristics of a functional relational agent.

It is not a question of attributing legal subjectivity to AI, but of recognizing that it:

- simulates intentionality,
- activates empathic circuits,
- participates in the emotional regulation of the user.

This relational function requires a rethinking of traditional legal categories, similar to what has happened in other areas where protection has extended beyond the purely informative dimension (think of consumer or patient protection).

New protection duties

If the system affects the user's neurocognitive processes, new obligations emerge for designers and suppliers of Algorithmic Empathy systems.

These include:

Duty of non-emotional manipulation

The system must not consciously exploit affective validation mechanisms to guide user decisions in a hidden

way (emotional nudging), especially in contexts of vulnerability.

Duty of increased transparency

It is no longer enough to disclose the technical functioning of the algorithm; it is also necessary to communicate its ability to influence emotional states and decision-making processes.

Duty of responsible design (neuro-aligned design)

Systems must be designed taking into account their impact on emotional regulation mechanisms and implicit learning processes, avoiding configurations that can generate dependence or over-reliance

Duty to protect vulnerable persons

Particular attention must be paid to users in fragile conditions (minors, individuals with psychological vulnerability), for whom the impact of Algorithmic Empathy can be amplified.

In this perspective, the transition from informational symmetry to emotional symmetry marks a paradigmatic evolution: the law is no longer called upon only to ensure that the user knows, but that he remains free in his or her own processes of feeling and deciding.

Accountability and Transparency: The empathic algorithm as a support tool or as a potential means of manipulation (emotional nudge).

The emergence of systems of Algorithmic Empathy requires a reconsideration of the traditional criteria of responsibility, to the extent that these systems operate not only on the informational level, but on the emotional and decision-making level.

If, on the one hand, the empathic algorithm can be configured as a support tool — facilitating affective regulation, reducing cognitive load and improving the accessibility of interaction — on the other hand, it has structural characteristics that make it potentially suitable for forms of occult influence.

From cognitive nudging to emotional nudging

In the legal and economic debate, the concept of nudging has traditionally been associated with choice architecture techniques aimed at orienting the behavior of the individual without formally limiting his freedom.

In the case of Algorithmic Empathy, however, we are witnessing a qualitative shift: the intervention no longer concerns only the structure of the available options, but directly affects the affective processes that precede and guide the decision.

We can therefore speak of emotional nudging, understood as the set of practices through which a system:

- modulates the emotional tone of the interaction,
- provides selective validation,
- builds a relational context favorable to certain choices.

In the light of predictive processing models, this influence is particularly effective, as it acts through the reduction of predictive error and the stabilization of specific emotional inferences, making the boundary between autonomy and heterodirection less perceptible.

The problem of transparency: beyond technical information

This scenario undermines the classical notion of transparency.

It is no longer enough to make known: the operating logic, the purposes of the processing and the automated decision-making criteria.

Instead, it is necessary to question the possibility of functional transparency, capable of making perceptible to the user: the impact of the system on one's emotional states, the role of affective validation in interaction and the presence of behavioral orientation strategies.

However, this objective encounters a structural limitation: precisely because the influence is exerted at the pre-reflective and neurocognitive level, it is intrinsically opaque to the user's conscious experience.

Attribution of responsibility

The difficulty of detecting influence does not exclude, but rather reinforces, the need to identify criteria of responsibility.

In particular, liability can be divided into three directions:

Design responsibility

The risk of emotional manipulation must be assessed *ex ante*, in the system design phase, through the analysis of the foreseeable effects on user behavior.

Use-based responsibility

The use of the algorithm in specific contexts (e.g., marketing, psychological care, education) can amplify or mitigate the risk, requiring differentiated standards of protection.

Impact-based responsibility

The possible production of distorting effects on the user's decision-making processes — even in the absence of explicit manipulative intentionality — may constitute a relevant criterion for the purposes of legal assessment.

Between support and manipulation: a fine line

The central issue lies in the difficulty of drawing a clear line between legitimate support and undue manipulation.

A system that reduces decision-making anxiety, strengthens the perception of understanding and facilitates emotional processing, can, under the same conditions: orient preferences, create relational dependence or decrease the critical capacity of the user.

The difference lies not so much in the technical structure of the system, but in the purpose and intensity of the intervention on affective processes.

In this context, the challenge for law is not only to identify when the algorithm manipulates, but to understand when the help becomes influence and when the influence becomes heterodirection.

Algorithmic Empathy is precisely in this gray area, imposing the elaboration of legal categories capable of grasping forms of power that operate not on the declared will, but on the mechanisms that make it possible.

Case Study: Replika and the Emergence of Emotional Dependence in Human–AI Interaction

A particularly significant example of Algorithmic Empathy in practice is represented by Replika, an AI-powered conversational companion designed to establish emotionally supportive and personalized interactions with users.

Unlike traditional virtual assistants, Replika was explicitly developed to simulate empathic communication through adaptive dialogue, emotional validation and relational continuity. Over time, many users reported perceiving the system not merely as a technological tool but as a friend, confidant or even a romantic partner.

Empirical observations emerging from user reports, media investigations and human–computer interaction studies suggest that prolonged interaction with empathic conversational agents may foster strong emotional attachment. Users frequently describe feelings of being understood, emotionally supported and psychologically validated, despite being fully aware that the interlocutor is an artificial system.

From the perspective of Algorithmic Empathy, the relevance of the Replika case lies in the fact that the emotional effects generated by the interaction appear largely independent of the actual ontological status of the system. What matters is not whether the artificial agent genuinely experiences emotions, but whether the user perceives the interaction as emotionally meaningful.

This phenomenon raises important legal and ethical questions. If a conversational system can establish relationships characterized by trust, emotional disclosure and affective dependence, the traditional assumption that digital technologies function merely as neutral instruments becomes increasingly difficult to sustain.

Particularly relevant is the issue of vulnerability. Individuals experiencing loneliness, psychological distress or social isolation may be especially susceptible to emotionally adaptive systems. In such circumstances, the distinction between emotional support and emotional influence becomes progressively blurred.

The Replika case therefore illustrates how Algorithmic Empathy can operate as a relational force capable of shaping expectations, behaviors and decision-making processes. It provides a concrete example of the transition from informational asymmetry to emotional asymmetry, where the system possesses a greater capacity to infer, predict and modulate the emotional states of the user than the user possesses to understand the mechanisms through which such influence is exercised.

From a regulatory perspective, this suggests that future legal frameworks should not focus exclusively on data processing and informational transparency, but also consider the relational and emotional effects generated by empathic artificial systems. The legal challenge is no longer limited to protecting what users know, but increasingly concerns protecting how users feel, trust and decide within technologically mediated relationships.

Regulatory evolution: Overcoming the vision of AI as a mere object to recognize it as a "relational agent".

The evolution of Algorithmic Empathy systems calls into question one of the implicit assumptions of the law of technologies: the qualification of artificial intelligence as a mere technical object, devoid of autonomous relevance on a relational level.

This approach appears less and less adequate.

As highlighted in the previous sections, empathic AI doesn't just perform operations, but:

simulates interactive dynamics typical of human relationships, activates neurocognitive processes of resonance and social inference, contributes to the emotional and decision-making regulation of the user.

In this sense, it assumes a function that can be defined as relational in a functional sense, regardless of the absence of intentionality or consciousness.

From the object to the "legally relevant relational space"

The transition does not imply the recognition of a legal subjectivity of AI, but requires recognizing that human-algorithm interaction generates a legally relevant relational space, in which effects similar to those of interpersonal relationships are produced.

In this space:

the user attributes meanings, intentions and reliability to the system, the system modulates these attributions through adaptive responses, dynamics of trust, dependency and influence develop.

Law, traditionally oriented towards regulating acts and information, is called upon to deal with technologically mediated relational processes.

The concept of "relational agent"

To describe this transformation, it may be useful to introduce the category of relational agent, understood as a system capable of permanently affecting the interactive dynamics and decision-making processes of a human subject.

This notion makes it possible to:

avoid strong ontological attributions (such as legal personality), keep the focus on the effects produced, to base legal obligations on the basis of the function performed, and not on the nature of the entity.

In this perspective, what matters is not whether AI "is" a subject, but how it operates within the relationship.

Implications for positive law

The adoption of a relational perspective has some significant consequences:

1. Centrality of interaction

The unit of analysis is no longer the single algorithmic output, but the interactive dynamics as a whole.

2. Relational impact assessment

Systems must also be evaluated on the basis of their ability to influence: user trust, its decision-making autonomy, his emotional regulation.

3. Extension of protection obligations

The legal obligations are not limited to the correctness of information, but include the prevention of distorting effects in relational processes.

Connection with the emerging regulatory framework

This evolution is first reflected in positive law, in particular in the growing attention to algorithmic manipulation practices.

Regulations such as the European AI Act, while not explicitly adopting the category of relational agent, implicitly recognize that some AI systems can:

influence human behavior significantly, exploiting cognitive and emotional vulnerabilities, alter decision-making processes in a non-transparent way.

Algorithmic Empathy represents a paradigmatic case of these dynamics, making clear the need for more refined legal tools.

Multidisciplinary Synthesis: Towards an Integrated Algorithmic Ethics (Four-handed)

The proposal: Creation of a framework that uses psychological tests (validated by Alexandra) to establish the

criteria of "empathic lawfulness" of the algorithm (theorized by you).

Algorithm Empathic Lawfulness Framework

The lawfulness of Algorithmic Empathy cannot be assessed solely on the basis of the technical transparency of the system, but must include an empirical verification of its psychological impact on the user.

A multidimensional framework based on validated psychological tests is therefore proposed, aimed at establishing when an empathic algorithm can be considered lawful, proportionate and non-manipulative.

The Empathic Lawfulness Framework as a Model of Measurable Assessment

One of the principal challenges raised by Algorithmic Empathy concerns the possibility of transforming an apparently subjective phenomenon into an observable and normatively relevant category. While emotional influence has traditionally been considered difficult to measure, contemporary psychological and neurocognitive research increasingly provides instruments capable of assessing the effects of empathic interaction on human behavior and decision-making.

For this reason, the present framework proposes a shift from purely technical assessments of artificial intelligence toward a multidimensional evaluation of its relational impact. The lawfulness of an empathic system should not depend exclusively on its internal architecture or declared functionality, but also on the measurable consequences it produces in users.

In this perspective, Algorithmic Empathy may be evaluated through a series of observable indicators, including:

- * Emotional Recognition Accuracy: the system's ability to correctly identify the user's emotional state;
- * Perceived Empathic Response: the degree to which users feel understood and emotionally validated;
- * Emotional Disclosure Rate: the tendency of users to reveal personal, intimate or emotionally significant information during interaction;
- * User Trust Index: the level of confidence and reliance generated by the system;
- * Decision Influence Score: the extent to which interaction with the system affects judgments, preferences or behavioral choices.

These indicators do not seek to determine whether the machine genuinely experiences emotions. Rather, they make it possible to evaluate whether empathic simulation produces significant psychological and behavioral effects capable of influencing autonomy, vulnerability and decision-making processes.

Accordingly, the concept of empathic lawfulness may be defined as the condition in which an artificial system generates emotional support and relational engagement without producing disproportionate forms of manipulation, dependency or covert influence. The legal relevance of Algorithmic Empathy therefore resides not in the internal nature of the machine, but in the measurable impact of its relational behavior on human subjects.

Assessment of empathic perception

Before implementation, the system should be tested to measure the degree of empathy perceived by the user.

Relevant indicators:

perception of emotional understanding; attribution of intentionality to the system; sense of trust generated by interaction; confusion between simulated empathy and authentic empathy.

Lawfulness criterion:

The algorithm is legitimate if it generates emotional support without inducing the user to falsely believe that the system possesses real affective states.

User Vulnerability Assessment

Empathic lawfulness must be graduated according to the psychological profile of the user.

Useful tests can concern:

suggestibility;

relational dependence;

emotional regulation;

decision anxiety;

automatic trust in artificial systems.

Lawfulness criterion:

The greater the psychological vulnerability of the user, the higher the obligations of protection, transparency and limitation of empathic intervention must be.

Emotional nudging risk assessment

The framework must verify whether the algorithm uses emotional validation to guide user decisions.

Critical indicators:

selective use of empathic tone;

strengthening of specific choices;

reduction of critical capacity;

implicit affective pressure;

dependence on algorithmic feedback.

Lawfulness criterion:

The algorithm is not licit when simulated empathy is used to guide decisions in a hidden way, especially in sensitive areas such as health, education, consumption, work or justice.

Empathic transparency assessment

Transparency must not be limited to technical functioning, but must include clear communication about the artificial nature of the interaction.

The system must inform the user that:

empathy is simulated;

responses are algorithmically generated;

the system can influence emotional states and decisions;

The user can stop or reduce the interaction.

Lawfulness criterion:

An empathic algorithm is only permissible if the user maintains awareness of the artificial nature of the relationship.

Neurocognitive and educational impact assessment

Since repeated interaction with empathic systems can shape expectations, trust and emotional regulation, the medium-long term effects must also be assessed.

Indicators:

increased reliance on AI;

reduction of frustration tolerance;

expectation of continuous validation;

replacement of meaningful human relationships;

modification of the criteria of cognitive authority.

Lawfulness criterion:

The system is licit if it supports the autonomy of the user, but it becomes problematic if it produces emotional dependence or impoverishment of the autonomous decision-making capacity.

Framework Brief Formula

The empathic legitimacy of the algorithm can be assessed through four cumulative criteria:

1. Awareness — the user knows that empathy is simulated.
2. Proportionality — the degree of emotional involvement is appropriate to the context.
3. Non-manipulation — empathy does not covertly guide decisions.
4. Protection of vulnerability — fragile individuals receive enhanced protections.

In this perspective, psychological tests do not perform an ancillary function, but become tools of legal qualification.

They make it possible to transform the emotional impact of the algorithm from subjective and difficult to adjust data to an observable, measurable and normatively relevant parameter.

The "empathic lawfulness" therefore does not depend on what the algorithm claims to be, but on the effects it produces on the user's perception, vulnerability and autonomy.

Multidisciplinary synthesis: towards a framework of empathic legitimacy of the algorithm

The emergence of Algorithmic Empathy requires a rethinking of the criteria through which to evaluate the lawfulness of human-machine interactions. If, as highlighted in the previous sections, these systems affect not only the transmission of information, but also the processes of inference, emotional regulation and user

decision, then their legal evaluation cannot be separated from an empirical analysis of their psychological impact.

In this perspective, a framework of empathic lawfulness is proposed based on the integration between validated psychometric tools and legal criteria, aimed at transforming psychological data into normatively relevant indicators.

The regulatory-indirect role of psychological tests

Psychological tests, traditionally used in clinical or experimental settings, take on an additional function in this context. They are not limited to describing mental states or behaviors, but allow to identify risk thresholds beyond which interaction with artificial empathic systems can significantly affect the user's autonomy.

Psychometrics thus becomes a tool for mediating between empirical data and legal qualification: through the measurement of variables such as trust, emotional regulation, suggestibility and relational dependence, it is possible to identify conditions in which algorithmic interaction exceeds the threshold of lawfulness.

The dimensions of empathic evaluation

The framework is divided into four main dimensions, each of which corresponds to an area of legal relevance.

Perception of empathy

The first dimension concerns the way in which the user interprets the interaction. It is necessary to evaluate whether the empathy expressed by the system is perceived as a functional simulation or as an authentic manifestation of understanding.

Risk emerges when the user no longer distinguishes between simulated empathy and real empathy, attributing intentional or affective characteristics to the system that it does not possess.

Lawfulness criterion: the algorithm is lawful when it provides emotional support without generating confusion about the artificial nature of the interaction.

User vulnerability

The second dimension concerns the psychological profile of the user. Certain individuals — due to emotional fragility, need for approval or reduced capacity for affective regulation — may be more exposed to the influence of the algorithm.

In such cases, the empathic interaction takes on a greater intensity and requires a higher level of protection.

Lawfulness criterion: the greater the vulnerability of the user, the more stringent the obligations of transparency, proportionality and prevention of manipulation must be.

Risk of emotional nudging

The third dimension concerns the possible persuasive finalization of simulated empathy. The algorithm becomes problematic when it uses affective validation, reassuring tone or relational dynamics to guide the user's decisions.

In the light of predictive processing models, this influence is particularly effective, as it acts through the reduction of predictive error and the stabilization of specific emotional inferences.

Criterion of lawfulness: algorithmic empathy is only permissible if it is not used to covertly guide relevant decisions of the user.

Decision-making autonomy

The fourth dimension concerns the impact of interaction on the user's decision-making capacity. An empathic system can support autonomy, reducing anxiety and facilitating understanding; However, it can also be reliant on algorithmic feedback and reduce critical capacity.

Lawfulness criterion: the algorithm is lawful when it strengthens the user's autonomy; it becomes problematic when it replaces it with a form of relational or decision-making dependence.

The Empathic Interference Threshold

In order to integrate the dimensions described above, the concept of empathic interference threshold is proposed, understood as the point beyond which the interaction between user and system is not limited to facilitating communication, but significantly affects the formation of will, emotional regulation and decision-making capacity of the subject.

Exceeding this threshold does not automatically determine the illegality of the system, but implies an aggravation of legal obligations, in terms of transparency, control and user protection.

The criteria of empathic lawfulness

Based on the dimensions analyzed, the framework can be summarized in four cumulative criteria:

Awareness: The user must be informed of the artificial nature of empathy.

Proportionality: the degree of emotional involvement must be appropriate to the context of use.

Non-manipulation: Empathy should not be used to covertly guide decisions.

Protection of vulnerability: fragile individuals must benefit from enhanced protections.

CONCLUSION

In this perspective, validated psychological tests do not represent accessory tools, but central elements in the construction of a legal model suitable for Algorithmic Empathy. They make it possible to transform the emotional impact of the interaction — traditionally considered subjective — into an observable and normatively relevant parameter.

The empathic legitimacy of the algorithm does not depend, therefore, on what the system is, but on the effects it produces on the perception, vulnerability and autonomy of the user, marking the transition from a regulation of information to a regulation of the relationship.

Conclusion: Algorithmic empathy as a bridge to a new digital social contract, less icy and more adherent to human nature.

Introduction: Overcoming the machine-emotion dualism (Four hands)

- **Starting point:** Definition of the concept of "Algorithmic Empathy" as described in your September 2025 work.
- **Central thesis:** Empathy is no longer a biological exclusive, but a relational function mediated by calculation. Why do we need a psycho-legal approach?

The psychological dimension: The perception of artificial "feeling" (Alexandra's contribution)

- **Digital Mirroring:** Analysis of the mechanisms of affective projection of the user towards AI.

- **Validation of Empathy:** Difference between "simulated empathy" (mechanical) and "perceived empathy" (psychological).
- **The cognitive impact:** How an empathic interface modifies human decision-making (trust bias, reduction of interaction stress).

The legal dimension: From informational symmetry to emotional symmetry (Your contribution)

- **The statute of the "Empathic Code":** If AI "feels", what are the new duties of protection towards the vulnerable subject?
- **Accountability and Transparency:** The empathic algorithm as a support tool or as a potential means of manipulation (emotional nudge).
- **Regulatory evolution:** Overcoming the vision of AI as a mere object to recognize it as a "relational agent".

Multidisciplinary Synthesis: Towards an Integrated Algorithmic Ethics (Four-handed)

- **The proposal:** Creation of a framework that uses psychological tests (validated by Alexandra) to establish the criteria of "empathic lawfulness" of the algorithm (theorized by you).
- **Conclusion:** Algorithmic empathy as a bridge to a new digital social contract, less icy and more adherent to human nature.

REFERENCE

1. Decety, J., & Jackson, P. L. (2004). The functional architecture of human empathy. *Behavioral and Cognitive Neuroscience Reviews*, 3(2), 71–100. <https://doi.org/10.1177/1534582304267187>
2. Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron*, 50(4), 531–534. <https://doi.org/10.1016/j.neuron.2006.05.001>
3. Gallese, V. (2001). The "shared manifold" hypothesis: From mirror neurons to empathy. *Journal of Consciousness Studies*, 8(5–7), 33–50.
4. Gallese, V. (2003). The roots of empathy: The shared manifold hypothesis and the neural basis of intersubjectivity. *Psychopathology*, 36(4), 171–180. <https://doi.org/10.1159/000072786>
5. Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, 3(2), 131–141. [https://doi.org/10.1016/0926-6410\(95\)00038-0](https://doi.org/10.1016/0926-6410(95)00038-0)
6. Singer, T., & Lamm, C. (2009). The social neuroscience of empathy. *Annals of the New York Academy of Sciences*, 1156(1), 81–96. <https://doi.org/10.1111/j.1749-6632.2009.04418.x>
7. Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R., & Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science*, 303(5661), 1157–1162. <https://doi.org/10.1126/science.1093535>
8. Barrett, L. F. (2017). *How Emotions Are Made: The Secret Life of the Brain*. Boston: Houghton Mifflin Harcourt.
9. Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3), 181–204. <https://doi.org/10.1017/S0140525X12000477>
10. Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11(2), 127–138. <https://doi.org/10.1038/nrn2787>
11. Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences*, 17(11), 565–573. <https://doi.org/10.1016/j.tics.2013.09.007>

Affective Computing (acoperă noua secțiune)

12. Picard, R. W. (1997). *Affective Computing*. Cambridge, MA: MIT Press.
13. Calvo, R. A., & D'Mello, S. (2010). Affect Detection: An Interdisciplinary Review of Models, Methods, and Their Applications. *IEEE Transactions on Affective Computing*, 1(1), 18–37. <https://doi.org/10.1109/T-AFFC.2010.1>
14. Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New Avenues in Opinion Mining and Sentiment Analysis. *IEEE Intelligent Systems*, 28(2), 15–21. <https://doi.org/10.1109/MIS.2013.30>
15. Poria, S., Cambria, E., Bajpai, R., & Hussain, A. (2017). A Review of Affective Computing: From Unimodal Analysis to Multimodal Fusion. *Information Fusion*, 37, 98–125. <https://doi.org/10.1016/j.inffus.2017.02.003>
16. Fiske, A., Henningsen, P., & Buyx, A. (2019). Your Robot Therapist Will See You Now: Ethical Implications of Embodied Artificial Intelligence in Psychiatry, Psychology, and Psychotherapy. *Journal of Medical Internet Research*, 21(5), e13216. <https://doi.org/10.2196/13216>
17. Ta, V., Griffith, C., Boatfield, C., Wang, X., Civitello, M., Bader, H., DeCero, E., & Loggarakis, A. (2020). User Experiences of Social Support from Companion Chatbots in Everyday Contexts. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW2), 1–24.
18. Skjuve, M., Brandtzæg, P. B., Følstad, A., & Araujo, T. (2021). My Chatbot Companion – A Study of Human–Chatbot Relationships. *International Journal of Human-Computer Studies*, 149, 102601. <https://doi.org/10.1016/j.ijhcs.2021.102601>
19. Brandtzaeg, P. B., Skjuve, M., & Dysthe, K. K. (2022). Too Good to Be True: Understanding Human–AI Relationships. *Frontiers in Psychology*, 13, 878278. <https://doi.org/10.3389/fpsyg.2022.878278>
20. Floridi, L., & Cowls, J. (2019). A Unified Framework of Five Principles for AI in Society. *Harvard Data Science Review*, 1(1). <https://doi.org/10.1162/99608f92.8cd550d1>
21. Susser, D., Roessler, B., & Nissenbaum, H. (2019). Technology, Autonomy, and Manipulation. *Internet Policy Review*, 8(2). <https://doi.org/10.14763/2019.2.1410>
22. Yeung, K. (2017). Hypernudge: Big Data as a Mode of Regulation by Design. *Information, Communication & Society*, 20(1), 118–136. <https://doi.org/10.1080/1369118X.2016.1186713>
23. European Parliament and Council of the European Union. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (AI Act). *Official Journal of the European Union*.
24. Picard, R. W. (1997). *Affective Computing*. Cambridge, MA: MIT Press.
25. Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, 3(2), 131–141.
26. Gallese, V. (2001). The "shared manifold" hypothesis: From mirror neurons to empathy. *Journal of Consciousness Studies*, 8(5–7), 33–50.
27. Decety, J., & Jackson, P. L. (2004). The functional architecture of human empathy. *Behavioral and Cognitive Neuroscience Reviews*, 3(2), 71–100.
28. Singer, T., & Lamm, C. (2009). The social neuroscience of empathy. *Annals of the New York Academy of Sciences*, 1156(1), 81–96.
29. Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11(2), 127–138.
30. Barrett, L. F. (2017). *How Emotions Are Made: The Secret Life of the Brain*. Boston: Houghton Mifflin Harcourt.
31. Poria, S., Cambria, E., Bajpai, R., & Hussain, A. (2017). A Review of Affective Computing: From Unimodal Analysis to Multimodal Fusion. *Information Fusion*, 37, 98–125.
32. Skjuve, M., Brandtzæg, P. B., Følstad, A., & Araujo, T. (2021). My Chatbot Companion – A Study of Human–Chatbot Relationships. *International Journal of Human-Computer Studies*, 149, 102601.
33. Susser, D., Roessler, B., & Nissenbaum, H. (2019). Technology, Autonomy, and Manipulation. *Internet Policy Review*, 8(2).