Health, Risk & Society

ISSN: 1369-8575 (Print) 1469-8331 (Online) Journal homepage: www.tandfonline.com/journals/chrs20

Algorithmic reconfiguration of mental health care: risk, trust and vulnerability in Italian professionals' accounts

Giulia Banfi & Noemi Crescentini

To cite this article: Giulia Banfi & Noemi Crescentini (2026) Algorithmic reconfiguration of mental health care: risk, trust and vulnerability in Italian professionals' accounts, Health, Risk & Society, 28:3-4, 153-171, DOI: 10.1080/13698575.2026.2663062

To link to this article: https://doi.org/10.1080/13698575.2026.2663062

© 2026 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

Published online: 20 Apr 2026.

Submit your article to this journal

Article views: 373

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=chrs20

Health, Risk & Society, 2026
Vol. 28, Nos. 3–4, 153–171, https://doi.org/10.1080/13698575.2026.2663062

Routledge
Taylor & Francis Group

Check for updates

RESEARCH ARTICLE

Algorithmic reconfiguration of mental health care: risk, trust and vulnerability in Italian professionals' accounts

Giulia Banfi[a] and Noemi Crescentini[b]

[a]*Department of Humanities, University of Ferrara, Italy;* [b]*Department of Social Sciences, University of Naples Federico II, Italy*

(*Received 21 July 2025; accepted 16 April 2026*)

Abstract

The growing integration of generative artificial intelligence (AI) into mental health care raises critical questions for risk studies about how trust, risk perception, and professional responsibility are reconfigured in algorithmically mediated therapeutic contexts. This study examines how Italian mental health professionals negotiate and interpret the introduction of AI into psychological practice, with particular attention to the social construction of risk and the conditions under which trust in algorithmic systems is extended or withheld. Data were collected in Italy between May and July 2025 through semi-structured interviews with 14 practicing psychologists, analysed using reflexive thematic analysis. Three interconnected dimensions emerged. First, professionals actively constructed risk perception through boundary work, distinguishing between acceptable instrumental automation and threatening encroachments on clinical judgement. Second, trust towards algorithmic systems and digital platforms was negotiated selectively and conditionally, shaped by algorithmic opacity and the reorganisation of therapeutic labour within platform economies. Third, vulnerable patients emerged as a site of amplified risk, where structural inequalities in the Italian mental health care system were compounded by unsupervised reliance on low-cost AI tools. These findings suggest that risk and trust in AI-mediated mental health care cannot be addressed through technical or regulatory frameworks alone, but require collective responses attentive to the relational, epistemic, and structural conditions under which care is practiced.

Keywords: Mental Health Care; artificial intelligence; professional risk perception; therapist-patient relationship; qualitative research

Introduction

In recent years, generative AI systems have rapidly moved from experimental tools to widely accessible infrastructures embedded in everyday practices, including domains

Corresponding author. Email: noemi.crescentini@unina.it

© 2026 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

characterised by high relational and psychological sensitivity such as mental health. While early public debates focused on isolated incidents,[1] such as problematic or harmful responses generated by conversational agents, these cases already pointed to broader dynamics. Such cases reveal how algorithmic infrastructures may reproduce and amplify vulnerabilities when relational dynamics shift towards automated systems, particularly in contexts where users are psychologically or socially exposed.

Large language models (LLMs) are increasingly mobilised as forms of relational support, particularly in situations of vulnerability. This growing integration raises critical questions concerning the conditions under which such systems are perceived as trustworthy, the types of risks they introduce, and the ways in which professional roles and therapeutic relationships are being reconfigured. Rather than understanding generative AI only in terms of technical performance or isolated failures, this study approaches these technologies as sociotechnical systems that mediate trust, redistribute expertise, and reshape how risk is perceived and managed in therapeutic contexts. It focuses on how mental health professionals interpret and negotiate the introduction of AI in relation to issues of uncertainty, responsibility, and relational care.

The emerging literature highlights deeply ambivalent dynamics: on the one hand, AI systems are associated with increased accessibility, personalisation, and continuous availability; on the other, they raise concerns regarding opacity, reliability, emotional dependency, and the erosion of professional boundaries (Giray, 2024; Haensch, 2025; Lawrence et al., 2024; Pandya et al., 2024). These tensions become particularly salient in mental health, where trust is not only a precondition for effective care, but also a fragile and continuously negotiated outcome of therapeutic relationships. In contexts mediated by opaque algorithmic systems (Burrell, 2016), trust must often be extended in the absence of full understanding, introducing new forms of epistemic uncertainty and asymmetries of knowledge between users, professionals, and technological infrastructures.

In this sense, risk cannot be understood merely as the possibility of technical failure or error, but rather as a relational and situated condition that emerges from the delegation of care, judgement, and interpretation to algorithmic systems. As interactions increasingly unfold through generative AI, both users and professionals are required to rely on systems whose internal logic remains largely inaccessible, thereby redefining the boundaries of responsibility and accountability within therapeutic relationships.

Generative AI systems are often perceived as less judgemental, more accessible, and continuously available compared to human professionals, contributing to experiences of relief, being heard, and perceived efficacy (Gerlich, 2024). At the same time, this perceived availability reflects a broader transformation in the temporal structure of care, as AI-mediated interactions tend to compress therapeutic timeframes and redefine what counts as sufficient or appropriate care (van Voorst, 2025).

However, this apparent ‘relational efficacy’ requires critical scrutiny.

The probabilistic and opaque forms of agency characteristic of these systems intertwine with human agency, reconfiguring not only the boundaries of the therapeutic relationship but also the roles, responsibilities, and forms of expertise involved in care. The introduction of AI contributes to ongoing transformations in professional competencies, knowledge hierarchies, and accountability structures (Moretti & Pronzato, 2024). These transformations call for closer examination of how trust is negotiated and potentially undermined in contexts where expertise is increasingly mediated by opaque

algorithmic systems, and where risk emerges as a relational rather than purely technical condition.

This study investigates how mental health professionals based in Italy involved in diverse areas of psychological practice interpret and negotiate the introduction of generative AI within therapeutic contexts, with particular attention to how trust, risk, and professional responsibility are reconfigured in AI-mediated interactions. Drawing on a qualitative research design based on semi-structured interviews with a purposive sample of practicing psychologists, the analysis focuses on the ways in which perceived risks associated with algorithmic tools are constructed, interpreted, and managed in everyday clinical practice-. Through the narratives of mental health professionals, it explores how expectations, resistance, and forms of selective appropriation emerge, alongside practices of negotiation and, at times, outright rejection. These dynamics contribute to redefining both the meaning of therapeutic work and the boundaries of professional expertise in increasingly algorithmically mediated environments. More specifically, the study addresses the following research question: *how do mental health professionals negotiate and interpret the introduction of artificial intelligence within the therapeutic relationship, and what conceptions of care, knowledge, and professional ethics emerge from these processes?*

The article is structured as follows. The next section develops the theoretical framework, focusing on the sociotechnical implications of generative AI in mental health care. This is followed by an analysis of the empirical findings, highlighting how professionals interpret and negotiate the role of AI in therapeutic practice. The final section discusses the results in light of the proposed framework, reflecting on the broader challenges posed by the integration of AI into mental health care.

Literature review

Understanding algorithmic risk in mental health AI

The integration of generative AI, as well as earlier forms of digital platforms, into mental health care brings into focus a tension at the core of therapeutic practice. While these systems promise to expand access, responsiveness, and continuity of support, they simultaneously reconfigure the relational conditions under which care is enacted. Chatbots and conversational agents have shown promising results in terms of accessibility and perceived empathy (Fitzpatrick et al., 2017; Hatch et al., 2025), yet their deployment raises critical questions that cannot be reduced to issues of technical performance alone.

Engaging with generative AI systems in therapeutic contexts presupposes an act of reliance under conditions of uncertainty. Turning to such systems for psychological support entails delegating processes of interpretation, judgement, and care to infrastructures whose internal operations remain largely inaccessible. In this sense, what is often framed as trust can be understood as a situated orientation towards systems that require action despite limited knowledge and asymmetrical visibility, with risk emerging from this very delegation.

This form of reliance resembles what has been described as trust in contexts of uncertainty, where actors place confidence in systems that cannot be fully inspected or verified (Luhmann, 1979), and where engagement with expert systems depends on routinised forms of trust that bracket uncertainty in order to enable action (Giddens, 1990). The interdependence between trust and risk becomes particularly

visible in moments of disruption, when unexpected or inappropriate outputs challenge the continuity of interaction. Under such conditions, the opacity of algorithmic systems (Burrell, 2016) prevents users from reconstructing the rationale behind a response, undermining their capacity to evaluate, contest, or reinterpret the interaction. Rather than simply diminishing trust, these breakdowns destabilise it, exposing the fragile and provisional nature of reliance on AI-mediated forms of care.

The systemic logics underpinning these technologies contribute to the reproduction of inequalities and exclusions. Training and data selection processes may reinforce the marginalisation of subjectivities that do not conform to the normative assumptions embedded in the models (Coghlan et al., 2023), particularly when datasets are incomplete or unrepresentative. The increasing reliance on automated systems in the management of psychological well-being therefore entails a redefinition of professional boundaries and responsibilities, with the risk of delegating aspects of care, diagnosis, and emotional support to opaque infrastructures.

While regulatory frameworks such as the AI Act (Regulation EU 2024/1689) and the European Health Data Space seek to address these challenges through principles of transparency, accountability, and risk management, they tend to conceptualise risk primarily in technical and legal terms. Such approaches risk overlooking the relational and situated dimensions through which trust is negotiated in practice. From a science and technology studies perspective, digital technologies are not neutral tools, but socio-technical arrangements that co-produce epistemic hierarchies, forms of governance, and professional boundaries (Jasanoff, 2004; Latour, 1992). So, the integration of generative AI into mental health care can be understood as a reconfiguration of the conditions under which trust is enacted and risk is experienced in therapeutic relationships.

Recent empirical studies show that AI models can produce a ‘simulation of empathy,’ often evaluated positively by both therapists and users, yet this linguistic performance risks conflating relational appearance with authentic experience (Hatch et al., 2025). What happens when the conversational agent fails to provide an adequate answer? What are the consequences at the limits of automated conversation?

These questions reveal a tension between the automation of therapeutic relationships and the humanisation of algorithms. While AI promises to expand access and standardise care, it simultaneously redefines the very meanings of empathy, trust, and professional accountability. This ambivalence also raises questions of responsibility and accountability, particularly in contexts where the effectiveness of AI-based tools remains uncertain and difficult to evaluate. In such cases, a responsibility gap may emerge, as potential harms cannot be easily attributed either to developers or to the systems themselves. The easy accessibility of generative AI, often available with minimal barriers and at no financial cost, has enabled its widespread use as a form of emotional support, offering constant availability in the absence of professional mediation. Moreover, the interactional style of these systems, typically accommodating and affirming, tends to align with users’ expectations rather than challenge them. While this may enhance perceived comfort and accessibility, it also risks reinforcing users’ perspectives, displacing the critical distance and interpretative tension that are central to therapeutic practice.- A tension thus emerges between the apparent democratisation of psychological support and the reconfiguration of the relational boundaries that traditionally structure therapeutic practice. The absence of temporal and professional constraints, combined with the always-available nature of algorithmic systems, can foster new forms of attachment and dependency that challenge

the normative conditions of care. It is through these sociotechnical arrangements that the dynamics of trust, risk, and professional practice are redefined.

Risk Perception as a Relational and Situated Construct

These dynamics call for a more precise conceptualisation of risk perception as a sociological – rather than merely technical – construct. Drawing on Brown and van Voorst (2024), risk in AI-mediated health care cannot be reduced to the probability of system failure or algorithmic error. Rather, it operates as a relational and organisational phenomenon, shaped by the ways in which professionals interpret uncertainty, assign responsibility, and negotiate the boundaries of legitimate delegation. In mental health care specifically, where therapeutic efficacy depends on the quality of relational interactions, risk perception is not a stable individual attribute but a situated and dynamic process, continuously produced and revised within professional practice. Professionals do not simply react to risks that AI systems generate; they actively construct risk through interpretive frameworks that reflect their epistemic commitments, professional identities, and understandings of what care requires. This active, constructive dimension of risk perception – as opposed to a passive or reactive one – is particularly salient in contexts where algorithmic systems introduce new forms of opacity and uncertainty that cannot be resolved through existing professional protocols.

Understanding these transformations requires moving beyond an analysis of individual systems towards a broader examination of the sociotechnical configurations through which AI and digital platforms are integrated into mental health care.

Sociotechnical implications of LLMs and digital platforms in mental health care

If the relationship between trust, risk, and opacity becomes visible at the level of interaction with AI systems, it is more fully structured within the broader sociotechnical arrangements through which these systems are deployed. In mental health care, AI is increasingly embedded within digital platforms that organise access to services, mediate interactions, and shape professional practices. Understanding its role therefore requires moving beyond individual systems towards an analysis of the infrastructures within which care is coordinated and delivered.

Digital platforms (Lupton, 2020; Van Dijck, 2021) have progressively reconfigured the organisation of mental health services by introducing new forms of coordination between patients and professionals (Moretti & Pronzato, 2024). Systems enabling remote interaction, therapist-patient matching, and the management of appointments and communication have become integral to contemporary care infrastructures. These platforms rely on algorithmic processes to align supply and demand, translating relational dimensions, such as proximity, availability, and specialisation, into calculable and operationalizable criteria. In doing so, they reshape how access to care is structured and how therapeutic relationships are initiated and maintained. Within these environments, the therapeutic relationship becomes mediated by infrastructures that organise access, interaction, and evaluation, introducing new forms of standardisation and coordination. In this context, trust extends beyond the interpersonal relationship between therapist and patient, becoming increasingly distributed across platforms, interfaces, and algorithmic procedures. Concurrently, the delegation of key processes, such as matching or preliminary

support, to algorithmic systems introduces new configurations of risk, tied not only to potential errors but also to the reorganisation of relational and professional boundaries.

These configurations of risk are deeply intertwined with the conditions under which trust is produced and sustained in algorithmically mediated care. It is important, however, to distinguish between two analytically distinct – though empirically interrelated – forms of trust at stake in these contexts. The first is interpersonal trust, which characterises the therapeutic relationship between patient and professional, and which depends on continuity, reciprocity, and the shared negotiation of meaning over time. The second is systemic trust, which Giddens (1990, p. 37) describes as confidence in abstract systems – expert knowledge, institutional procedures, technological infrastructures – that operate beyond the direct experience and verification of the actors who rely on them. In AI-mediated mental health care, both forms of trust are simultaneously at stake and subject to reconfiguration.

Trust, Confidence and Algorithmic Mediation

Platforms and algorithmic tools do not simply supplement interpersonal trust; they redistribute it, extending reliance onto systems whose internal logic remains opaque and whose outputs cannot always be evaluated through existing professional frameworks. This redistribution operates through a partial displacement of trust from interpersonal encounters to platform-mediated processes. Activities traditionally grounded in relational continuity, such as initial needs assessment, the matching between patient and professional, or the interpretation of symptoms, are increasingly structured by algorithmic systems and interface design. This shift also reconfigures the locus of trust. Trust is no longer situated exclusively within interpersonal relations but extends to the systems that mediate them. Access points (Giddens, 1990, p. 117) are no longer embodied only by professionals but increasingly by platforms themselves, which come to structure and mediate interaction. However, precisely because these systems are relatively new, opaque, and not fully intelligible or inspectable to users, they cannot be taken for granted as stable objects of confidence. Rather than providing a reliable background for interaction, they may instead emerge as objects of uncertainty and potential risk (Luhmann, 1979), requiring users and professionals to engage with their opacity and to negotiate their reliability in practice. Building on this distinction, trust in persons and confidence in systems should be understood as analytically distinct, while interrelated, levels. As Frederiksen (2016) argues, trust is not transferable across these levels: confidence in institutional arrangements does not automatically translate into trust in interpersonal encounters, nor vice versa. Rather, the perception of uncertainty becomes central in the classification of situations, shaping whether actors orient themselves towards trust or towards more reflexive and risk-based forms of engagement.

Understanding how these two forms of trust interact – and how their reconfiguration shapes professional practice and patient experience – is central to analysing the sociotechnical implications of AI integration in mental health care.

The integration of generative AI into these platforms further intensifies these dynamics. While such systems may support clinical activities or provide forms of emotional assistance, their statistical and probabilistic nature, combined with limited transparency, complicates processes of evaluation and accountability, particularly in contexts characterised by vulnerability and asymmetries of knowledge.

The automation of processes such as therapist-patient matching exemplifies how algorithmic mediation operates at the infrastructural level, facilitating the alignment between supply and demand based on proximity, specialisation, and availability. Predictive algorithms and recommendation systems are increasingly integrated into these platforms to enable more targeted pairings, potentially enhancing therapeutic adherence and the quality of the care relationship.

These developments are embedded within broader socio-political narratives that frame digital innovation as a solution to structural constraints in health care systems. The techno-optimistic vision promoted by neoliberal logics – is often centred on efficiency, personalisation, and continuous availability. However, the structural ambivalences of such technologies simultaneously generate new forms of empowerment and emerging modalities of control (Moretti & Pronzato, 2024).

It becomes essential to critically interrogate the narratives of technological inevitability, often fuelled by emergency rhetoric and resource scarcity, conditions that risk turning automation into a default rather than a deliberate choice (Brown & van Voorst, 2024; Pandya et al., 2024). Such narratives obscure the situated and contingent nature of technological integration.

Against this theoretical backdrop, the empirical analysis presented in this article examines three interconnected dimensions through which the integration of AI into mental health care is negotiated in practice. The first concerns how Italian psychologists construct and manage risk perception at the professional level – how they interpret the delegation of clinical tasks to algorithmic systems, define the boundaries of legitimate automation, and articulate the conditions under which AI can be integrated without compromising therapeutic accountability. The second dimension concerns the negotiation of trust towards algorithmic systems and digital platforms, examining how professionals extend, withhold, or qualify systemic trust in contexts marked by opacity and asymmetrical knowledge. The third dimension concerns the figure of the vulnerable patient as a site of amplified risk, exploring how structural inequalities in access to care – compounded by the availability of low-cost AI tools – reconfigure the conditions under which psychological support is sought, received, and evaluated. Together, these dimensions allow for an analysis of AI integration that moves beyond questions of technical performance, engaging instead with the relational, epistemic, and organisational transformations that algorithmic systems produce in mental health care.

Methodology

This study explores the current and potential uses of AI in psychological practice in Italy, focusing on the perceptions, experiences, and interpretive frameworks of mental health professionals. Its specific objective is to understand how AI-based tools are integrated – or perceived as integrable – into psychological practice, with particular attention to the conditions of acceptability, the epistemological implications, and the risk factors associated with their use.

A qualitative design based on semi-structured interviews was adopted (Bichi, 2005), a technique particularly suited to the investigation of situated meanings and the analysis of complex, evolving social phenomena. The sample was defined through a non-probabilistic, purposive strategy aimed at selecting individuals able to provide relevant information in relation to the research objectives, prioritising diversity of clinical context, geographical location, and professional experience (Corbetta, 2003). It included 14

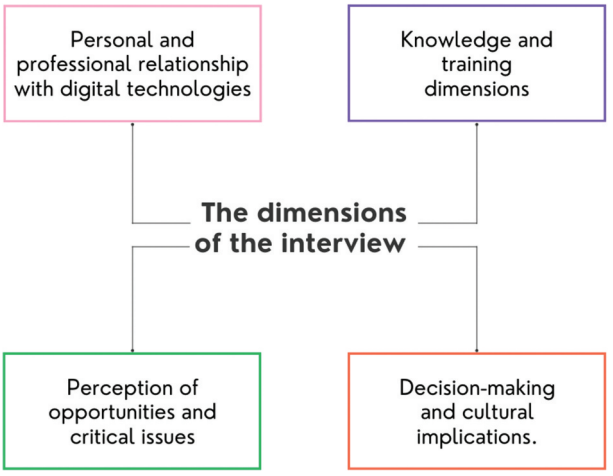

Figure 1. Visual model of interview dimensions.
Source: own elaboration.

psychologists (5 women and 9 men) aged between 30 and 50, actively engaged in psychological practice across different areas of Italy: seven from the south, three from the centre, and four from the north. Given the exploratory nature of the study and the small sample size, territorial differences were not analysed.

The interviews were conducted between May and July 2025 via Google Meet and were structured around 12 questions organised into four thematic areas (Figure 1): personal and professional relationship with digital technologies; knowledge and training; perception of opportunities and critical issues; and decision-making and cultural implications related to the adoption of technological tools. All interviews were recorded and transcribed in full.

Participants were guaranteed anonymity in accordance with the Personal Data Protection Regulation (Legislative Decree No. 196/2003) and were free to withdraw at any time. Ethical approval was obtained from the Institutional Ethics Committee of the University of Ferrara on 20 October 2025.

Analysis was conducted following the reflexive thematic analysis approach developed by Braun and Clarke (2019), which understands themes not as pre-existing patterns to be discovered in the data, but as actively constructed through the researcher's interpretive engagement with the material. This approach is particularly suited to theoretically informed qualitative research, as it allows for the integration of inductive coding – attentive to the participants' own categories and meanings – with a deductive theoretical reading oriented by the conceptual framework developed in the literature review. Concretely, the analytical process involved repeated close readings of the transcripts, the generation of initial codes, and their progressive organisation into broader thematic clusters through iterative discussion between the two researchers. NVivo 15 software was used to support systematic text management and the traceability of analytical decisions. The resulting themes were then interpreted in dialogue with the theoretical framework, focusing specifically on how professionals construct and negotiate risk perception and trust in relation to AI integration within therapeutic practice.

Findings

The reflexive thematic analysis of the interview material generated three analytical themes, which are presented below. These themes do not represent exhaustive categories but rather are theoretically oriented dimensions through which the interviewed professionals constructed and negotiated the integration of AI into their clinical practice.

Negotiating professional risk perception

The narratives of the interviewed psychologists revealed that risk, in relation to AI integration, was not experienced as an abstract or distant possibility, but as a situated and relational condition emerging within the everyday organisation of clinical work. Consistent with Brown and van Voorst (2024) conceptualisation of AI-related risk as an organisational and professional phenomenon rather than a purely technical one, the professionals interviewed did not simply react to risks generated by algorithmic systems – they actively constructed, interpreted, and managed risk through ongoing processes of boundary work (Gieryn, 1983), in which the legitimacy of algorithmic tools was continuously assessed against the relational and epistemic demands of therapeutic practice.

A first significant finding was that a broadly positive relationship with digital technology did not translate automatically into acceptance of AI as a clinical instrument. Across the sample, familiarity with digital tools – for scheduling, remote sessions, administrative management – coexisted with marked ambivalence towards algorithmic systems specifically. Professionals drew a clear boundary between technologies that extended their organisational capacity and those perceived as threatening to reconfigure clinical judgement itself. This distinction was not merely practical but deeply epistemological, reflecting a form of risk perception centred on the conditions of professional knowledge and accountability:

> GC: AI depends on how we want to use it and what we expect it to do. In psychology, I don't think it's a good thing, because the diagnostic moment is too important. That's when you label someone, you set the course for the therapeutic process, you give important guidance, and making a mistake at that point means a lot.

A particularly salient dimension of professional risk perception concerned what participants described, in various ways, as the risk of erosion of clinical agency – the gradual displacement of professional judgement by algorithmic outputs. This risk was perceived as especially acute in the diagnostic phase, where the attribution of meaning to a patient's distress was understood as an irreducibly interpretive and relational act, incompatible with the standardising logic of automated systems. As Timmermans and Berg (2003) have argued in relation to health care technologies more broadly, the negotiation of algorithmic tools within professional practice is always simultaneously a negotiation of professional identity and the legitimate conditions of expertise. This dynamic has been documented in other health care professions facing algorithmic integration: Lombi and Rossero (2024), in their qualitative study of radiologists, show how the introduction of AI reshapes professional autonomy precisely through the redefinition of boundaries between delegable and non-delegable tasks – a process strikingly similar to that described by the psychologists interviewed here. The interviewed psychologists articulated this tension explicitly:

> AM: When AI does everything for you, there is a risk of becoming irresponsible and paying less attention to fundamental elements, losing important nuances. Those with traditional training have a solid foundation and use technology as a support. But those who rely directly on AI risk skipping crucial steps, delegating too much to a tool that simplifies but does not replace expertise.

These narratives revealed a form of risk perception that was specifically epistemic in character: the concern was not only that AI might produce incorrect outputs, but that its routine integration might progressively reconfigure the practitioner's relationship to uncertainty, ambivalence, and interpretive responsibility. In this sense, risk was experienced not as an isolated event but as a structural condition tied to the reorganisation of professional practice – a finding that resonates with broader debates in the sociology of health care about the consequences of technological standardisation for clinical reasoning and relational care (Mol, 2008; Timmermans, 2020).

Crucially, the perception of professional risk was almost unanimously reframed by participants not in existential terms – as a threat to the survival of the profession – but in conditional ones, focused on the circumstances and governance of AI use. Practitioners expressed concern not about AI itself, but about the normalisation of uncritical delegation, a scenario in which algorithmic tools become the norm rather than deliberate choices:

> GC: My only hope is that, in some way, tools such as ChatGPT will be monitored. So, as far as our work is concerned, it would be useful to remember that if we can think, the tool is fine, but the tool must not be put before everything else, which is perhaps my greatest fear;
>
> MP: A smile at the right moment, perhaps a pat on the back in a certain context and a certain type of psychological support are worth much more than an immediate response, so I am confident that real human relationships will not be threatened.

This conditional reframing is theoretically significant. It reflects a mode of engagement with technological risk that Stilgoe and Guston (2016) associate with responsible innovation: an anticipatory and reflexive orientation in which professionals seek to govern the integration of AI by defining thresholds of legitimate delegation, maintaining accountability for clinical decisions, and preserving the epistemological foundations of therapeutic work. Rather than positioning themselves as passive recipients of technological change, the interviewed psychologists emerged as active risk managers, continuously negotiating the boundaries between what could be delegated to algorithms and what had to remain within the domain of human judgement and relational care.

Trust and distrust towards algorithmic systems and platforms

A second dimension that emerged from the interviews concerned the ways in which professionals negotiated trust – and distrust – towards algorithmic systems and digital platforms. Drawing on the analytical distinction introduced in the theoretical framework, the narratives revealed a complex and often asymmetrical relationship between interpersonal trust, which characterises the therapeutic relationship, and systemic trust, which extends towards platforms, interfaces, and algorithmic procedures that organise and mediate clinical work (Giddens, 1990).

A consistent finding across the sample was that professionals distinguished sharply between two regimes of automation. Instrumental automation – including standardised test scoring, clinical note summarisation, appointment management, and administrative tasks – was broadly accepted and, in many cases, valued as a means of reducing workload and preserving time for the therapeutic relationship. Decision-making automation, by contrast – encompassing diagnosis, case formulation, and the management of clinical uncertainty – was met with sustained hesitation and, frequently, outright resistance. This distinction mapped directly onto the trust-distrust dynamic: professionals extended systemic trust towards algorithmic tools insofar as those tools operated within the boundaries of administrative and organisational support, but withdrew it when automation encroached upon the interpretive and relational core of clinical work:

> EA: In my opinion, AI can be extremely helpful in diagnostics, but not in therapy, rehabilitation, or demand analysis;
>
> LC: We do not necessarily have to adapt to these tools, but we cannot ignore their presence: they must be integrated into clinical practice as much as possible.

EA's positive reference to diagnostics should be read carefully here: it concerns standardised assessment procedures – such as test administration and scoring – rather than the broader clinical act of diagnosis, which involves interpretive judgement, case formulation, and the negotiation of meaning with the patient. This nuance reinforces rather than contradicts the pattern described above.

This selective extension of trust reflected a broader professional logic of delegation (Parasuraman & Riley, 1997), in which the acceptability of algorithmic mediation was conditioned upon its capacity to extend – rather than replace – clinical attention and judgement. From a science and technology studies perspective, this dynamic can be understood as a form of co-production (Jasanoff, 2004): professionals were not simply adopting or rejecting pre-formed technologies but actively shaping the conditions under which algorithmic systems could legitimately participate in therapeutic work.

The evaluation of digital platforms – particularly telepsychology services – constituted a particularly revealing site for the negotiation of systemic trust. Platforms were simultaneously recognised as instruments of democratisation, expanding access to psychological support for economically fragile or geographically isolated users, and criticised as infrastructures that reconfigured therapeutic work according to the logic of the platform economy (Van Dijck, 2021).

Several interviewees described how algorithmic mediation at the platform level – automated matching, standardised communication, performance metrics – introduced new forms of organisational control that eroded professional autonomy and reconfigured the conditions of trust within the therapeutic relationship:

> CR: I strongly disagree with working through platforms. I consider them companies that have managed to carve out a niche market, attracting users who would otherwise not have access to therapy. However, I find the treatment of professionals degrading, including in financial terms. I understand that they offer a specific service for a certain type of user and professional, but I personally work differently. I would describe them as "the Ryanair of psychology.
>
> For this professional, platform-mediated work was experienced not simply as a practical inconvenience but as a structural reconfiguration of professional identity and worth — a process in which the logic of market efficiency progressively displaces the relational and ethical commitments that define therapeutic practice.

> NZ: I initially worked at Unobravo and I ran away from there. I really ran away, for me they are the evil of psychology. I believed so much in those projects initially, I really believed in them. I didn't do it for financial reasons, but because I had the idea that psychology would become accessible, instead it is becoming mainstream, terrible, and professionals are not valued.

These narratives pointed to a fundamental tension at the heart of platform-mediated care: while platforms extended access and reduced barriers to psychological support, they simultaneously introduced new configurations of risk tied to the reorganisation of relational and professional boundaries. The opacity of algorithmic matching procedures and the standardisation of interaction protocols meant that systemic trust could not be grounded in transparency or professional verification but had to be extended – or withheld – on the basis of partial and often insufficient information. This condition resonates with Burrell's (2016) analysis of algorithmic opacity as a structural feature of machine learning systems, which prevents users and professionals alike from reconstructing the rationale behind automated decisions.

The experience of platform-based work also revealed how professionals were required to navigate narratives of technological inevitability – the implicit assumption that digital platforms and algorithmic mediation represent an unavoidable feature of contemporary health care rather than a deliberate organisational choice. As Hoff (2023) has shown in his analysis of governmental sociotechnical imaginaries around AI and health care, these narratives function as political devices that pre-empt critical evaluation of automation by framing it as a structural necessity. Several interviewees resisted this framing, insisting on the conditional and governable nature of technological integration – consistent with Brown and van Voorst (2024) argument that risk work in AI-mediated health care requires the active interrogation of such inevitability narratives.

The experience of platform-based work also revealed what several interviewees described as a process of professionalisation under conditions of algorithmic control – a dynamic that Garofalo (2025) and Bonifacio (2024) have analysed as the platformisation of professional fields, in which therapeutic labour is progressively reorganised according to the metrics of efficiency, user retention, and workflow optimisation. For some interviewees, this process was experienced as a form of delegitimization, in which the relational and interpretive dimensions of psychological work were subordinated to the operational logic of digital platforms:

> EC: I have worked online, although not with the most well-known platforms. It certainly has its advantages, but also its limitations, like most things. It is good to be able to rely on these tools in today's hectic lifestyle, even if direct contact is sometimes lacking. When I am in person, I take care of every detail of the setting: the environment, the scent, the organisation of the space.

Taken together, these findings suggest that the integration of AI into mental health care involves a profound reconfiguration of the relational, epistemic, and organisational conditions under which care is practiced, producing new forms of risk and new demands on professional agency, institutional governance, and regulatory frameworks.

The vulnerable patient as a site of amplified risk

A third and particularly significant dimension that emerged from the interviews concerned the figure of the vulnerable patient as a site where the risks associated with AI integration were perceived as most acute and consequential. Professionals described a landscape in which structural inequalities in access to mental health care – compounded by the growing availability of low-cost or free AI tools – were reshaping the conditions under which psychological support was sought, received, and evaluated, often outside the boundaries of professional mediation.

A consistent finding was that patients' autonomous use of conversational AI systems – most frequently ChatGPT – was understood by professionals not primarily as a technological phenomenon, but as a response to systemic failures in the Italian mental health care system. Economic barriers, long waiting lists, fear of stigmatisation, and geographic disparities in service provision were identified as the structural conditions that drove vulnerable users towards AI tools as substitutes for professional support:

> AN: I believe that human beings adapt as best they can, especially in a country where the National Health Service is collapsing and psychotherapy is only guaranteed in the most serious cases. If a session costs at least €50 - €40 even on Unobravo - it is not affordable for those living on €1,000 a month or on welfare. In such cases, people adapt by looking for more accessible alternatives: for example, tools such as ChatGPT, which cost much less or are free;
>
> MP: When we think about these people who turn to ChatGPT for psychological advice, we have to blame the Italian health care system, which makes it impossible to access a psychologist. So, if someone is struggling to get psychological and psychiatric care, it's obvious that they'll do their best to get it at the lowest price or even for free.

These narratives pointed to a fundamental paradox at the heart of AI-mediated mental health support: the very populations most exposed to psychological risk – those facing economic precarity, social isolation, or institutional exclusion – were also those most likely to rely on algorithmic systems in the absence of professional alternatives. From this perspective, AI tools functioned as what Mol (2008) might describe as vernacular care technologies – pragmatic, low-threshold resources mobilised by social actors to address forms of suffering that formal services failed to reach. However, this apparent democratisation of access simultaneously introduced new and compounded forms of risk, concentrated precisely among those least equipped to evaluate or contest the outputs of algorithmic systems. This dynamic resonates with Pols's (2012) analysis of care at a distance, which demonstrates how remote and technologically mediated forms of care produce new relational configurations that are neither simply equivalent to nor fully replaceable by face-to-face interaction – a tension that becomes particularly acute when the mediating technology is an opaque algorithmic system rather than a monitored telecare device.

Professionals identified several specific risk configurations associated with patients' unsupervised use of AI. The first concerned the production of unvalidated self-diagnoses and the uncritical reinforcement of dysfunctional hypotheses. Without the interpretive mediation of a therapist, conversational AI systems were described as returning comforting or easily acceptable content that aligned with users' existing perspectives, rather than introducing the critical distance and productive tension that characterise therapeutic practice. This dynamic – which resonates with broader concerns about algorithmic

systems optimising for user satisfaction rather than therapeutic efficacy (Coghlan et al., 2023) – was understood as particularly dangerous for patients in psychologically fragile conditions:

> IG: I see a potential danger, especially when there is psychological vulnerability, when AI is used to compensate for problems, for example. That's where I see a potential danger. I also had a young patient who used a chat app that I wasn't even aware of, where you talk to someone who doesn't exist, so it's AI, and she did it because she felt lonely and it was a way to connect with someone who didn't exist.

A second and related risk configuration concerned the emergence of forms of emotional dependency on conversational AI systems, particularly among young or socially isolated patients. Several professionals reported cases in which patients had begun to use chatbots not only as sources of information but as substitutes for human relationships, simulating emotional bonds and filling relational voids through interaction with algorithmic systems. This dynamic introduced a particularly complex layer of risk: the compensatory relational function temporarily assumed by AI threatened to reinforce mechanisms of avoidance of real others and deepen patterns of social withdrawal, running counter to the relational and intersubjective foundations of therapeutic work.

A third risk configuration concerned what professionals described as the intrusion of AI-generated content into the therapeutic relationship itself. Several interviewees reported cases in which patients brought the outputs of their interactions with conversational systems into sessions, using them as a basis for self-assessment, comparison, or challenge. In such cases, AI assumed the role of a symbolic third party (Gieryn, 1983; Jutel, 2024) – a non-human entity whose outputs circulated within the therapeutic relationship, introducing new dynamics of authority, credibility, and interpretive competition:

> EC: If a patient uses artificial intelligence to explore what emerges during a session, it can be a useful support. In my cognitive-behavioural approach, I encourage personal work between sessions: ideally, therapy lasts all week. If AI helps to enhance this process, it is certainly welcome!

These contrasting evaluations highlighted the deeply ambivalent character of AI's role in relation to vulnerable patients. While some professionals recognised the potential of AI as an inter-session support tool within structured therapeutic frameworks, the majority expressed concern about its unsupervised use in contexts of vulnerability, where the absence of professional mediation transformed the apparent accessibility of AI into a source of amplified risk. This finding resonates with van Voorst's (2025) analysis of the temporal reconfiguration of care introduced by AI systems, in which the constant availability and immediate responsiveness of algorithmic tools redefine what counts as sufficient or appropriate support – potentially normalising forms of care that are structurally inadequate for the complexity of psychological distress.

Taken together, the three risk configurations identified in this dimension – unvalidated self-diagnosis, emotional dependency, and the intrusion of AI-generated content into therapeutic relationships – pointed to a broader dynamic in which the structural inequalities of the Italian mental health care system were not simply reproduced but actively amplified by the introduction of AI tools. In contexts where access to professional care was already severely constrained, the availability of low-cost algorithmic

alternatives did not compensate for systemic deficits but introduced new layers of risk, concentrated among those already most vulnerable.

Discussion and conclusions

In this study we have examined how a small, non-probability sample of Italian mental health professionals negotiated and interpreted the introduction of AI into therapeutic practice, analysing their opinions and the implications for the therapeutic relationship, professional standards, and quality of care. The research is grounded in a theoretical framework that considers the platformisation of society (Van Dijck, 2021) and the digitalisation of healthcare (Lupton, 2020) and focuses specifically on the ways in which risk perception and trust were constructed, managed, and reconfigured in algorithmically mediated clinical environments. The three analytical dimensions identified through reflexive thematic analysis offer a theoretically grounded account of these processes, contributing to ongoing debates in the sociology of health care about the social and relational implications of AI integration in mental health settings.

A first and overarching finding concerns the fundamentally relational and situated character of professional risk perception. Consistent with Brown and van Voorst (2024) conceptualisation of AI-related risk as an organisational and professional phenomenon, the interviewed psychologists did not experience risk as a technical abstraction but as a condition embedded in the everyday negotiation of clinical work. Risk was actively constructed through practices of boundary work (Gieryn, 1983), in which the legitimacy of algorithmic delegation was continuously assessed against the epistemic and relational demands of therapeutic practice. This framework highlights that professionals did not view AI as an existential threat, but rather as a tool whose value was negotiated through situated practices – practices that recognised AI's effectiveness in data management while reaffirming the inherently relational nature of the therapeutic process, and that positioned the psychologist as a critical mediator, capable of integrating digital tools without passively adhering to their internal logic.

This finding extends existing analyses of risk perception in health care professions by demonstrating that, in the specific context of mental health, risk is perceived not only in relation to potential clinical errors but in relation to the structural reconfiguration of professional knowledge, accountability, and interpretive responsibility. A central issue in this mediation was confronting algorithmic opacity. The often-indecipherable nature of AI systems obscured clinical decision-making processes, making it difficult for professionals to validate or challenge machine outputs. This transforms transparency into a prerequisite for professional autonomy (Moretti & Pronzato, 2024). The fear of clinical agency erosion – the gradual displacement of human judgement by algorithmic outputs – emerged as a distinctive form of professional risk perception, one that cannot be adequately captured by purely technical or regulatory frameworks.

A second significant contribution concerns the negotiation of trust in algorithmically mediated care. The analytical distinction between interpersonal trust and systemic trust (Giddens, 1990) proved particularly productive for interpreting the empirical material. Professionals extended systemic trust towards algorithmic tools selectively and conditionally, accepting automation in administrative and organisational domains while withdrawing it in the interpretive and relational core of clinical work. The evaluation of digital platforms revealed an additional layer of complexity. While platforms expanded

access to psychological support, they simultaneously introduced new configurations of risk tied to the opacity of algorithmic matching procedures, the standardisation of therapeutic interactions, and the reorganisation of professional labour according to the logic of the platform economy (Garofalo, 2025; Van Dijck, 2021). In this context, systemic trust could not be grounded in transparency or professional verification but had to be extended – or withheld – under conditions of irreducible uncertainty. This dynamic resonates with Luhmann's (1979) analysis of trust as a mechanism for reducing complexity in conditions where full knowledge is unavailable, and with Burrell's (2016) account of algorithmic opacity as a structural feature that fundamentally reconfigures the conditions of accountability in automated systems. The tension between efficiency and clinical autonomy existed within a broader context of structural weaknesses in the Italian healthcare system – ranging from underfunding to regional disparities – which exacerbated the risk of implicitly delegating care to low-cost technologies. Without an adequate ethical and technical framework, AI-based platforms could evolve from supplementary tools into substitute infrastructures, normalising an 'unethical temporality' (van Voorst, 2025) where superficial, immediate support obscures systemic neglect.

A third and particularly significant finding concerns the figure of the vulnerable patient as a site of amplified risk. The narratives of the interviewed professionals pointed to a fundamental paradox: the populations most exposed to psychological risk were also those most likely to rely on algorithmic systems in the absence of professional alternatives, driven by structural inequalities in the Italian mental health care system. The industrialisation of care, driven by market logic, has the potential to reconfigure therapeutic relationships by eroding their embodied and empathetic dimensions (Mol, 2008). It acts as a co-producer (Jasanoff, 2004) of new subjectivities, in which AI asserts itself as a symbolic third party – an algorithmic actor that reshapes expectations of listening, competence, and responsiveness, calling for careful examination of how relational algorithms reconfigure trust and power relations, particularly in the case of vulnerable users (Latour, 1992).

This finding has important implications for the sociology of health care inequality. The apparent democratisation of access to psychological support through low-cost AI tools did not compensate for systemic deficits but introduced new and compounded forms of risk, concentrated among those already most vulnerable. As van Voorst (2025) has argued in relation to the temporal reconfiguration of care introduced by AI, the constant availability and immediate responsiveness of algorithmic systems redefine what counts as sufficient or appropriate support – potentially normalising forms of care that are structurally inadequate for the complexity of psychological distress. The risk configurations identified in this dimension – unvalidated self-diagnosis, emotional dependency, and the intrusion of AI-generated content into therapeutic relationships – point to the need for a critical examination of how algorithmic systems interact with existing inequalities in access to mental health care, and of the governance frameworks required to protect vulnerable users from the unintended consequences of AI-mediated support.

Taken together, these findings suggest that the integration of AI into mental health care cannot be understood as a straightforward process of technological adoption or resistance. Rather, it involves a profound reconfiguration of the relational, epistemic, and organisational conditions under which care is practiced – a reconfiguration that produces new forms of risk and new demands on professional agency, institutional governance, and regulatory frameworks. The concept of responsible innovation (Stilgoe & Guston, 2016) captures something important about the orientation of the professionals

interviewed: their approach to AI was characterised by anticipation, reflexivity, and responsiveness – an implicit governance framework oriented towards preserving the conditions of legitimate and accountable care in an increasingly algorithmically mediated environment. However, individual professional reflexivity is insufficient as a response to systemic challenges. The findings point to the need for collective and institutional forms of governance that address the structural conditions – underfunding, inequality of access, regulatory gaps – that shape the context in which AI is integrated into mental health care in Italy and beyond.

This study has several limitations that should be acknowledged. The sample was small and purposive, and the exploratory nature of the research design limits the generalisability of the findings. The absence of patients' perspectives represents a further limitation, as the analysis relied exclusively on professionals' accounts and interpretations of patients' experiences. Future research should incorporate the perspectives of service users – particularly those from vulnerable and marginalised groups – to develop a more complete understanding of how AI integration reshapes the experience of care from multiple standpoints. It is also essential to clarify how clinical specialisation influences technological integration, and the role that universities should play in promoting critical thinking skills with regard to the opacity of digital systems – a competence that remains marginal in academic programmes today. Longitudinal and comparative designs would be valuable for tracking how professional practices and risk perceptions evolve as AI becomes more deeply embedded in clinical environments across different national health care systems.

Ultimately, integrating AI into mental health care involves what Adams (2022) describes as an ontological negotiation of the therapeutic encounter. For this reason, critical awareness and constant ethical reflection are essential in order to protect the intersubjective foundations of care from the neoliberal fusion of accessibility and automation. As these technologies advance, it becomes imperative to preserve the clinical value of intuition and human contact – what participants referred to as 'the smile at the right moment' - by resisting the pressures of perpetual availability and pure algorithmic efficiency. The challenge is not only to govern AI as a tool, but to ensure that the conditions under which care is organised, made accessible, and practiced remain oriented towards the relational and ethical demands of therapeutic work.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Author contributions

CRediT: **Giulia Banfi:** Conceptualization, Investigation, Writing – original draft, Writing – review & editing; **Noemi Crescentini:** Conceptualization, Data curation, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) reported there is no funding associated with the work featured in this article.

Note

1. In 2024, a widely circulated case drew public attention when a student from Michigan, while interacting with a generative language model developed by Google, received the response 'Human … please die' while seeking assistance for a university project. The company dismissed the episode as an instance of AI 'hallucination,' attributing it to still poorly understood internal processes within these systems (CBS News, 20 November 2024: https://www.cbsnews.com/news/google-ai-chatbot-threatening-message-human-please-die/).

References

Adams, J. (2022). A theory of infrastructural rhetoric. *Communication Design Quarterly Review*, *10*(3), 46–55. https://doi.org/10.1145/3507870.3507876

Bichi, R. (2005). *La conduzione delle interviste nella ricerca sociale*. Carocci.

Bonifacio, F. (2024). Professionisti di piattaforma. Un'analisi della piattaformizzazione dei campi professionali in ambito medico e psicologico. In I. Pai (Ed.), *Il welfare alla prova delle piattaforme. Lavoro e servizi di cura nella transizione digitale* (pp. 326–344). Fondazione Giangiacomo Feltrinelli.

Braun, V., & Clarke, V. (2019). Novel insights into patients' life-worlds: The value of qualitative research. *The Lancet Psychiatry*, *6*(9), 720–721. https://doi.org/10.1016/S2215-0366(19)30296-2

Brown, P., & van Voorst, R. (2024). The influence of artificial intelligence within health-related risk work: A critical framework and lines of empirical inquiry. *Health, Risk & Society*, *26*(7–8), 301–316. https://doi.org/10.1080/13698575.2024.2412374

Burrell, J. (2016). How the machine "thinks": Understanding opacity in machine learning algorithms. *Big Data & Society*, *3*(1). https://doi.org/10.1177/2053951715622512

Coghlan, S., Leins, K., Sheldrick, S., Cheong, M., Gooding, P., & D'Alfonso, S. (2023). To chat or bot to chat: Ethical issues with using chatbots in mental health. *Digital Health*, *9*, 1–11. https://doi.org/10.1177/20552076231183542

Corbetta, P. (2003). *Social research: Theory, methods and techniques***.**. Sage Publications.

Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial. *JMIR Mental Health*, *4*(2), e19. https://doi.org/10.2196/mental.7785

Frederiksen, M. (2016). Divided uncertainty: A phenomenology of trust, risk and confidence. In S. Jagd & L. Fuglsang (Eds.), *Trust, organizations and social interaction* (pp. 43–67). Edward Elgar Publishing.

Garofalo, L. (2025). (Tele) therapist, platform worker, data manager: Therapeutic labor and the new therapeutic exchange. *Economic Anthropology*, *12*(2), e70005. https://doi.org/10.1002/sea2.70005

Gerlich, M. (2024). Societal perceptions and acceptance of virtual humans: Trust and ethics across different contexts. *Social Sciences*, *13*(10), 516. https://doi.org/10.3390/socsci13100516

Giddens, A. (1990). *The consequences of modernity*. Polity Press, Cambridge.

Gieryn, T. F. (1983). Boundary-work and the demarcation of science from non-science: Strains and interests in professional ideologies of scientists. *American Sociological Review*, *48*(6), 781–795. https://doi.org/10.2307/2095325

Giray, L. (2024). Cases of using ChatGPT as a mental health and psychological support tool. *Journal of Consumer Health on the Internet*, *28*(4), 367–377.

Haensch, A. C. (2025). "*It listens better than my therapist": Exploring social media discourse on LLMs as mental health tool* [Preprint]. LMU Munich; University of Maryland, College Park.

Hatch, S. G., Goodman, Z. T., Vowels, L., Hatch, H. D., Brown, A. L., Guttman, S., Le, Y., Bailey, B., Esplin, C. R., Harris, S. M., Holt, D. P., Jr., O'Connell, M., McLaughlin, P., Ritchie, K., Rothman, L., Top, N., Jr., Braithwaite, S. R., & Braithwaite, S. R. (2025). When Eliza meets therapists: A Turing test for the heart and mind. *PLOS Mental Health*, *2*(2), e0000145. https://doi.org/10.1371/journal.pmen.0000145

Hoff, J. L. (2023). Unavoidable futures? How governments articulate sociotechnical imaginaries of AI and healthcare services. *Futures*, *148*, 103131. https://doi.org/10.1016/j.futures.2023.103131

Jasanoff, S. (2004). The idiom of co-production. In S. Jasanoff (Ed.), *States of knowledge* (pp. 1–12). Routledge.
Jutel, A. (2024). *Putting a name to it: Diagnosis in contemporary society*. Johns Hopkins University Press.
Latour, B. (1992). Where are the missing masses? The sociology of a few mundane artifacts. In W. Bijker & J. Law (Eds.), *Shaping technology/building society: Studies in sociotechnical change* (pp. 225–258). MIT Press.
Lawrence, H. R., Schneider, R. A., Rubin, S. B., Matarić, M. J., McDuff, D. J., & Jones Bell, M. (2024). The opportunities and risks of large language models in mental health. *JMIR Mental Health*, *11*, e59479. https://doi.org/10.2196/59479
Lombi, L., & Rossero, E. (2024). How artificial intelligence is reshaping the autonomy and boundary work of radiologists: A qualitative study. *Sociology of Health & Illness*, *46*(2), 200–218. https://doi.org/10.1111/1467-9566.13702
Luhmann, N. (1979). *Trust and power*. Wiley.
Lupton, D. (2020). Special section on "sociology and the coronavirus (COVID-19) pandemic". *Health Sociology Review*, *29*(2), 111–112. https://doi.org/10.1080/14461242.2020.1790919
Mol, A. (2008). *The logic of care: Health and the problem of patient choice*. Routledge.
Moretti, V., & Pronzato, R. (2024). The emotional ambiguities of healthcare professionals' platform experiences. *Social Science & Medicine*, *357*, 117185. https://doi.org/10.1016/j.socscimed.2024.117185
Pandya, A., Lodha, P., & Ganatra, A. (2024). Is ChatGPT ready to change mental healthcare? Challenges and considerations: A reality-check. *Frontiers in Human Dynamics*, *5*, 1289255. https://doi.org/10.3389/fhumd.2023.1289255
Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, *39*(2), 230–253. https://doi.org/10.1518/001872097778543886
Pols, J. (2012). *Care at a distance: On the closeness of technology*. Amsterdam University Press.
Stilgoe, J., & Guston, D. (2016). *Responsible research and innovation*. MIT Press.
Timmermans, S. (2020). The engaged patient: The relevance of patient-physician communication for twenty-first-century health. *Journal of Health and Social Behavior*, *61*(3), 259–273. https://doi.org/10.1177/0022146520943514
Timmermans, S., & Berg, M. (2003). The practice of medical technology. *Sociology of Health & Illness*, *25*(3), 97–114. https://doi.org/10.1111/1467-9566.00342
Van Dijck, J. (2021). Seeing the forest for the trees: Visualizing platformization and its governance. *New Media & Society*, *23*(9), 2801–2819. https://doi.org/10.1177/1461444820940293
van Voorst, R. (2025). "No longer morally justifiable" temporal dynamics of care, or how AI made waiting unethical. *Health, Risk & Society*, *27*(3–4), 130–144. https://doi.org/10.1080/13698575.2025.2495331