Cordula Kropp* and Tobias Renner

# Automation, Co-Agency, and Distributed Responsibility: Caring for Hybrid Therapeutic Networks

https://doi.org/10.1515/auk-2026-3005

**Abstract:** Drawing on insights from sociology and science and technology studies, we develop an account of agency as an emergent and relational property constituted within hybrid configurations of humans and machines. We speak of *co-agency* to capture the ways in which responsible action is jointly produced within human-machine entanglements, with both parties demonstrating purposively intended as well as determined forms of action. By emphasizing the interdependence of sociotechnical systems, therapeutic practices, and institutional automation infrastructures, we argue that responsibility cannot be attributed to discrete actors – whether human or digital – but must instead be understood as generated within the relational processes through which co-agency emerges. This reconceptualization shifts attention from locating isolated bearers of responsibility to examining the socio-technical arrangements that structure action, thereby suggesting recommendations for ethical guidelines aimed at supporting responsible automation in healthcare.

**Keywords:** responsibility; co-agency; healthcare; autonomous machine agency; sociotechnical networks; ethical guidelines for care in hybrid networks

***Corresponding author: Cordula Kropp**, Institute for Social Sciences, University of Stuttgart, 70174 Stuttgart, Germany, E-mail: cordula.kropp@sowi.uni-stuttgart.de
**Tobias Renner**, Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, University Hospital of Psychiatry and Psychotherapy, Osianderstr. 14-16, 72076 Tübingen, Germany; and Medical Director, Department of Pediatric Psychiatry and Psychotherapy, University Hospital Tuebingen, Tübingen, Germany

# 1 Introduction: Structural Tensions Between Automation and Responsibility

Smart devices, AI (Artificial Intelligence) assistants, and robots are increasingly assuming tasks, organizational functions, and decision-making responsibilities across a wide range of domains – including healthcare. Machine learning techniques are intended to enhance flexibility and precision by moving beyond rules-based systems toward stochastic pattern recognition and the identification of relationships between individual characteristics and contextual factors. These digital agents promise relief from the burden of repetitive tasks and an improved overview and convey an appearance of attentiveness to individual needs and desires, while presenting themselves as tirelessly available to meet them. Despite well-known oligopolistic structures and the biases embedded in their design and operation,[1] healthcare professionals and institutions are inclined to rely on these 'intelligent' systems, often entrusting them with sensitive and consequential tasks. The systems can be used to diagnose diseases, predict their risk, develop personalized therapies and remotely monitor patients, all of which is linked to "a new form of agency capable of coupling, re-coupling, and de-coupling different parts of the healthcare system [...] in ways that may fundamentally transform the meaning of health, the nature of healthcare, and the practice of medicine" (Morley and Floridi 2025, 30).

As Suchman (2007) observed, these techniques pose a classic management dilemma: "The more empowered these agents are, and the more capable of pursuing their own self-interests rather than ours, the less reliable they are" (Suchman 2007, 219). From a robotics perspective, the reliability of these agents is inseparable from their degree of automation, yet comes with a potentially problematic lack of transparency, which raises fundamental questions about the significance of so-called autonomous machine agency in healthcare and about the capacities of designers, users, and institutions to take responsibility for the outcomes of automated decision-making. We deliberately take on the concepts of 'autonomous systems' and 'autonomous control' from computer science and engineering – despite the irritation they may provoke from a philosophical perspective. In these contexts, *automation* refers to the predefined execution of tasks, whereas

**1** This refers to the oligopolies of data collection and processing, platform economies, and global technology corporations, whose ability to capitalize on vast amounts of data (extraction) allows them to control the development of markets, digital infrastructures, and modes of organization (Rahman and Thelen 2019) as well as to set research agendas.

*autonomy* denotes a system's ability to achieve certain goals independently under conditions of uncertainty and without human intervention (Antsaklis 2020, 16). Every autonomous system is considered to be a control system (ibid). Etymologically, autonomy derives from the Greek *autos* (self) and *nomos* (law), meaning "one who gives oneself his/her own law" (17). For humans, it describes the capacity of self-governance, that is the ability to act according to one's own reasons and/or values rather than external coercion.

However, debate persists over whether autonomy requires rational self-governance and how this applies to individuals or communities constrained by power relations or other forms of restraint. Within a health technology assessment framework oriented towards trustworthy implementation, a central question emerges: given the relational nature of co-agency among medical staff, patients, digital instruments and machine learning systems, as well as the distributed and opaque character of automated action, what principles can guide responsibility in the development and use of AI-supported, automated healthcare systems?

We explore this question by focusing on the interplay of automation, agency, and responsibility within sociotechnical networks composed of people, organizations, digital technologies, and other artifacts. Our approach is relational: agency is understood as emerging from the dynamic interactions within these interconnected worlds, shaped by mutual dependencies. Just as a robot cannot act autonomously, but only in conjunction with human deployment and maintenance, such deployment and maintenance cannot occur independently of the robot, nor of the infrastructures, tools and organizational knowledge that support it. This perspective draws on actor-network theory (ANT; Latour 2005), feminist technoscience (Haraway 1985), and practice-oriented understandings of the entanglement between technology, social contexts, agency, and responsibility (Suchman 2007; Puig de la Bellacasa 2017). Since our inquiry focuses on therapeutic settings, ethical questions of care and responsibility play a central role in the analysis. Before introducing these settings, we first outline a common understanding of automation, agency and responsibility, which is further elaborated in dialogue with a case analysis in Section 2 and with existing literature in Section 3.

*Automation* is commonly understood as both a technical and social process through which human labor or decision-making is replaced, supplemented, or governed by machines, software, or other technological systems. Technically, this means tasks previously performed by humans are increasingly carried out by mechanical, electrical, or digital technologies. Socially, it entails delegating human decisions to technical agents. These technologies promise to lower costs, increase productivity, and reduce temporal, energetic, and cognitive effort, while enabling enhanced monitoring and higher precision. A defining characteristic of contemporary automation, in what Brynjolffson and McAfee (2014) termed the 'Second

Machine Age', is that algorithms, machines, and robots – integrated with vast arrays of inexpensive sensors and large-scale databases – do not operate in isolation. Instead, they communicate, coordinate, and continuously learn, allowing them to perform complex tasks and processes. Whereas earlier forms of automation primarily replaced manual labor by substituting individual workers, the automation of knowledge work through intelligent networks extends interconnected digitalization across nearly every domain, from medical diagnostics to autonomous driving and scientific text generation, signalling a move beyond humans working alongside machines toward scenarios in which the "full replacement" of human labor becomes conceivable (Brynjolffson and McAfee 2014).

From a sociological perspective, automation requires that the accelerated and technically standardized execution of tasks, aligned with particular models, knowledge and interests, be embedded within organizational settings, triggering shifts in sociotechnical systems and social relations (Howcroft and Taylor 2022; Leonardi 2011). In this process, definitions of situations and organized practices increasingly orient toward technically mediated, hybrid forms of agency, reconfiguring work environments, organizational rules, and scripted practices considered appropriate in context (Rammert 2016; Suchman 2007; van den Berg 2008). Consequently, automation and AI reshape not only work but also preferences, cultural practices, and entire markets, raising ethical concerns (Milano et al. 2020; Stinson 2021).

Although a vast body of literature explores the impacts of AI-enabled automation on work, organizations, social structures, and standards, debates often run the risk of oversimplifying the socio-economic shaping of sociotechnical change and the fluid, relational boundaries between human and technical agency (Howcroft and Taylor 2022; Suchman 2023). Individuals, while rarely rendered fully dependent on technical devices, may also act automatically, implement automated decisions, or consciously choose to deviate from them, exercising judgement within complex sociotechnical networks.

The core feature of the sociological understanding of *agency* lies precisely in the "capability of the individual to 'make a difference' to a pre-existing state of affairs or course of events" (Giddens 1984, 14). It refers to the fundamental capability of knowledgeable agents to act independently and to deliberately alter processes. Although society and technology are mutually constitutive (rather than separate spheres), influencing one another in multiple ways and each evolving in relation to the other, the distinction between human and technical agency does not simply dissolve (Sayes 2014). Rather, the issue at stake is precisely how human and technical agency are entangled, a question we explore in more detail in Section 3.

Since automation and machine learning operate fundamentally according to if-then rules and are therefore more purposive than most human behavior, one cannot simply distinguish human from technical agency by attributing *purposive* *action* to humans alone. Instead, the concept of co-agency is required to adequately capture their interplay. When vast amounts of information and stochastic processes inform task execution, prioritizing insights that humans typically rely on only implicitly – if at all – these forms of displacement of human decision-making, intuition, and rule- or knowledge-based, context-sensitive action raise critical concerns regarding accountability and responsibility.

To whom can errors be attributed? Who is liable? And to whom can one appeal in special cases or request exceptions? In fact, *accountability and responsibility* cannot simply be assigned; they are evaluative and attributional concepts circulating within networks of responsible parties (Lenk 2017). *Responsibility* is generally understood as an obligation to ensure, within a given framework, that the necessary and appropriate actions are taken to prevent harm to oneself, others, or protected goods; at the same time, making judgments about whether a person is morally responsible for their behavior – and holding them accountable for actions and their consequences – is a fundamental and familiar part of moral practices and interpersonal relationships. In our context, this means that responsibility is about using digital technologies in therapeutical and ethically appropriate ways, while accountability is about justifying these uses and being answerable for their outcomes. This presumes the ability to assess and evaluate the consequences of one's actions and to control uses accordingly. Philosophically, however, this capacity is neither given nor universal; it is constructed, relative and distributed, grounded in reasons for action. What counts as necessary and appropriate depends on interpretation, attribution and contextually valid justifications. Every assumption of responsibility requires distinguishing responsible from irresponsible action, intended effects from unintended side-effects or the contingent factors of an overly complex environment. Furthermore, responsible action depends on the freedom and resources available to act. We therefore regard responsibility as an ethical standard of justification rather than an absolute quality.

*Responsibility* is thus a relational construct of attribution and evaluation (merely in the sense of a productive fiction), enacted in real, often sociotechnical contexts (Lenk and Maring 2001, 259–60): Someone (a person or organization) is held responsible for something (actions or consequences) with respect to an object of responsibility (protected goods such as health or the environment), based on a judgement (prescriptive, normative or ethical criteria). While accountability attaches to the agent answerable for the intended or unintended consequences of actions and omissions (backward-looking), responsibility attaches to the agent ethically expected to act or refrain from acting with respect to desired outcomes (forward-looking). Both concepts presuppose a causality, in which controllable chains of effects are assumed. Advocates of automation tend to follow linear cause-and-effect reasoning, whereas critics highlight uncertainties, ambivalence and the huge complexity of sociotechnical systems.

Against this background, the network of subjects and objects, humans and technologies with determined or autonomous capacities for action, must be examined closely to assess the relationship between automation, agency and responsibility. The central point of this text is to consider automation on both sides of the human-machine interaction, enmeshed in hybrid networks. To do so, we first introduce the entanglements of human and technical agency and responsibility through a case study (Section 2). In Section 3, we conceptually classify these blurred situations and trace the shifts in agency caused by automated decision-making systems. In conclusion, we address what ethical guidelines for care can be derived from these considerations that make the use of such automated technologies responsible.

## 2 Automation as a Relational Process – An Example from Healthcare

In sociology, research questions are more often examined empirically than purely conceptually, by studying social interactions within the context of given structures. In this article, the relationship between automation and responsibility is explored in interdisciplinary cooperation with clinical science, using a discussion example drawn from child psychiatry and psychotherapy. The case condenses current observations while avoiding data protection concerns. It is based on two research projects investigating the use and effectiveness of sensor technology in psychotherapy for children and adolescents. The studies explore whether video-based tele psychotherapy can be enhanced through smart sensors that inform patients and therapists about stress-related states indicative of the need to adapt the therapeutic process (Klein et al. 2024; Alt et al. 2025).

With the following report, we aim to place readers in a position to understand and engage with the intertwined themes of automation, responsibility and co-agency. The case study is intended to demonstrate, in particular, the extent to which rational, goal-oriented action on the one hand and determined, automated behavior on the other cannot be attributed to humans or technologies as such but must instead be reassessed situationally. We return to the discussion example later as a backdrop for further considerations.

When the COVID-19 pandemic reached Germany in early 2020, Noah had been attending kindergarten for a year. The family was worried about the risks posed by the pandemic, especially as Noah's mother was expecting her second child, and they immediately introduced hygiene measures. Noah learned to wash his hands regularly – after being outside, after visits, and before meals – and did so conscientiously, so that everyone could rely on him. Soon, he internalized and automated the behavior, to the point that it became second nature. However, as further measures such as social distancing, bans on visits, and daycare lockdowns were implemented, he grew increasingly uneasy in social interactions, and his parents' worries added to the strain.

Two years later, what had started as a reasonable precaution had developed into a compulsion. The fear of germs and infection accompanied Noah constantly. He felt the need to wash his hands frequently and intensively to relieve stress and regain a sense of control. He began insisting that the sink be perfectly clean, that the soap be used only by him, that baby items for his sister not be left nearby. Later, he even demanded that his father use a different sink. His compulsive thoughts and actions, that had started as measures to avert harm, turned into an urge to perform formalized behaviors that increasingly affected the family. The household eventually adapted to his strict rules to avoid escalating stress, and the pressure intensified, especially when guests were visiting.

Three more years later, it was clear that Noah's hygiene rituals provided short-term relief but reinforced his suffering and entrenched the symptoms of his obsessive-compulsive disorder. By imposing automated-like processes of compulsions and ritualized handwashing, the disorder significantly reduced his freedom of action and overall quality of life. At this point, his family finally secured therapy for Noah. His persistent efforts to regain control through handwashing and distancing are now addressed through therapeutic reflection and behavioral therapy. In a recent session, the family was offered the opportunity to participate in a study testing a "smart wristband." Using integrated data on skin conductance and heart rate variability measured via sensors, this sensor-based device alerts wearers with a gentle impulse when their stress level increases. If variability is high and regular, the wristband signals that the wearer should engage cognitive strategies to counteract compulsive urges; if variability is consistently low, it indicates chronic tension. The wristband can also store data and associate these with specific time points, although comparative data for children are still limited.

The researchers hope to use the case in an exploratory study with a clinical focus to gain data-based insights into the connections between stress levels, compulsive behaviors, heart rate variability, and environmental factors, and to explore whether technological support can enhance therapy. The therapists expect the wristband to provide more continuous care beyond the weekly 45-minute face-to-face sessions, enabling the child's compulsions to be addressed before they escalate. Noah himself is eager to try the device, to allow his inflamed, overwashed hands to heal and hopefully put an end to the highly uncomfortable situations he finds himself in when washing his hands at school. His mother supports participation in the study, in the hope of bringing greater calm to the household, but his father fears new complications and dependence on a device whose mode of operation is not yet fully validated.

If the study reveals positive therapeutic effects, the use of technical media to communicate and stabilize therapeutic goals beyond the clinical encounter will likely spread rapidly, given the scarcity of therapy places. As such technologies proliferate, therapeutic preferences and routines, the availability of medical technology, billing regulations, and the expectations of patients and relatives will all adapt accordingly.

This vignette highlights several questions that arise at the intersection of automation and responsibility. Understanding responsibility as an evaluative and attributional concept in overcomplex environments suggests that young Noah operates within a productive fiction of responsibility. Though lacking the means and knowledge to protect his family during the pandemic, he was led to believe that the responsibility was partially his and did his best. Now, though he wishes to overcome his compulsions, he cannot fully control his behavior in stressful situations and may at times even choose to maintain the disorder temporarily as a way of relieving stress. Interestingly, his therapists also face limits in their means and knowledge. They rely on available information, professional guidelines, subjective experience, and ethical standards, all within a network of agents and events that remains only partially controllable. Inevitably, they depend on collaboration with individuals from other contexts, such as parents and technicians, as well as on scientific and technical processes like the accurate digital processing of sensor data, of which they have limited understanding and can exert little control. Moreover, they cannot know whether, once such technologies enter commercial healthcare markets, the strict data-protection standards of medical research will be upheld or compromised by competing economic interests.

Responsibility must therefore be understood as inherently distributed, a state of what is described in ethics as joint “co-responsibility” (Lenk 2017, 220). Sociologists tend to emphasize the substantively, socially, and temporally fragmented nature of “joint” or “distributed” responsibility. Responsibility is seen as unfolding in networks of human and non-human agents, where agency becomes a matter of configuration and coordination (Kropp and Wortmeier 2021, 100; Rammert 2016, 175; Suchman 2007). The less such networks are coordinated, the more fragmented the possibilities for responsibility and accountability become, as participants know too little about one another. Given their differing strategies and constraints, a networked approach would be required (Kropp and Wortmeier 2021; Saurwein 2019).

Within this complex interplay, inter- and transactive technologies, such as sensor technology, robotics, and AI, which transcend the passive nature of traditional tools and introduce new levels of machine autonomy (Rammert and Schulz-Schaeffer 2002, 49; Winfield et al. 2019; Grote et al. 2024), pose particular challenges. While a blood pressure monitor is a passive device that can simply be switched off, a pacemaker intervenes autonomously in cases of malfunction. A wristband that automatically optimizes its warning functions based on detected alterations (e.g. in skin resistance) already qualifies as a transactive technology whose control requires networked expertise and demonstrates the opacity of automated machine agency. In such complex cyber-physical systems, it often remains unclear which tasks are taken over by technical agents, and which decision-making models are involved. As a result of such opacity, the use of the smart wristband generates short-term risks of misdiagnosis due to inappropriate calibration or flawed data evidence, as well as long-term risks of reshaping what behaviors are considered compulsive and what therapeutic approaches are deemed appropriate (Morley and Floridi 2025).

The attribution of agency and responsibility is therefore not straightforward but highly dependent on functions, competences, and domain-specific ethics, which makes it a matter of perspective and sociotechnical design. For instance, how do engineers in medical contexts responsibly balance the principles of safety, accuracy, and cost efficiency when selecting sensor systems? From an ethical standpoint, prevention-orientated design that accounts for possible side effects is recommended (Lenk 2017). Yet in the healthcare sector, economic considerations increasingly play a decisive role. Responsibility remains an evaluative attribution, albeit a productive and indispensable one, because without assuming and attributing responsibility, whether individual or collective, neither trust nor the identification of culpable irresponsibility would be possible. Without social attributions of responsibility, members of society would face boundless contingency and a lack of accountability.

## 3 Concepts of Distributed Agency, Responsibility and Control

In this section, we deepen the discussion of automation and agency and, in particular, explain why their mutually constitutive interdependence calls for professional-ethical guidelines for an ethics of care. Such guidelines are needed in order to compensate for the limitations of distributed responsibility, which can no longer be adequately assumed by individuals or managed sequentially.

The example discussed here highlights the ambiguity of automation: it is not a unidirectional process by which human action or decision-making is simply replaced, supplemented, or governed by machines, software, or other technological systems. Rather, automation simultaneously assumes decision-making and process control while demanding individual attention. For instance, Noah is now automatically made aware of his rising stress level and draws conclusions for action. While his compulsive behaviors follow relatively automated patterns, his response to the wristband's impulse requires considerable cognitive effort, modulated by emotional processes. Similarly, in organizational contexts, automation does not merely relieve workers from repetitive routine tasks; it can also generate (techno)stress as operators remain responsible for monitoring the accelerated execution of automated tasks. This effect is described as 'ironies of automation' (Bainbridge 1983).

Moreover, automated behavior can occur even in the absence of any technical device when input information leads to an output without consideration of individual circumstances – for example, when therapy choices are mechanically derived from disease classifications. These considerations underscore the importance of distinguishing between determined and autonomous automation. Traditionally, technical automation has been seen as determined "*self-activity*", while human autonomy has been understood as autonomous "self-governance" or self-regulation (cf. Brödner 2019, 89). Today, however, we must recognize that both human and technical agents possess varying degrees of autonomy while remaining interdependent. The more tightly the networks between them are interwoven, the more blurred the dualistic distinction between autonomy and determination – between self-governance and self-activity – becomes. Therapeutic devices powered by AI, such as the wearable discussed here, can strongly influence how patients behave while remaining opaque in their operation. Yet this entanglement is neither entirely new nor wholly unintended.

Luhmann (1997, 524) described technology as a "functioning simplification" that channels human action through technically induced tight coupling. The use of technology enforces intended outcomes, reduces disruptions, saves costs and helps build consensus, because acting differently seems irrational. Scholarship in Science and Technology Studies (STS) similarly emphasizes how the design and implementation of technologies embed the interests, values, and assumptions of those who develop and deploy them, stabilizing processes and scripts of use (van den Berg 2008). Increasing reliance on technical operations has made many social practices contingent on functioning technologies, with alternative modes of operation becoming obsolete or unavailable. As Luhmann (1997, 532) noted, this dependence means that a breakdown in technical systems entails a corresponding collapse in the social practices that depend on them. From an ANT perspective, individual entities lose their capacity to act when a previously stabilized network of actors, technologies and artifacts disintegrates. Thus, agency is not an inherent property of isolated entities but is distributed across heterogeneous associations constituting a network (Callon 1986; Latour 2005). The identities and capacities of all actors – whether human or non-human – are constituted through their interdependent relations with one another.

What, then, is the role of AI in such networks? It is often perceived as more powerful than the individual actors who enrol and mobilize it, while the couplings it generates are even more opaque (Burrell 2016). This dynamic can lead users to feel both empowered by AI and diminished in their autonomy, and helpless in the event of malfunction or failure. Automation is therefore not merely a technical procedure but a sociotechnical process, replacing or supplementing individual action and decisions with shared standards, apparatuses or systems. Paradoxically, the more automation substitutes human judgment in routine cases, the more it relies on human responsibility in exceptional cases: the wristband itself cannot detect technical malfunctions. So far, it has been considered the responsibility of operators to intervene in the event of unexpected deviations or irregularities. They must suspend or override automation when necessary, exercising reflective judgments while remaining embedded within the broader network. Without such meaningful assumption of responsibility, ethical considerations risk being displaced by technical feasibility or purely algorithmic calculations.

Sociologically, a conceptual distinction is made between human action, understood as subjectively meaningful, goal-oriented and context-bound, and human behavior, which is seen as an almost instinctive reaction or an observable manifestation of action. While individual actions are largely shaped and enabled by social practices and structures that render them expectable and intelligible, behavior tends to occur quasi-automatically, without deliberative reflection. *Agency*, however, refers to the fundamental capability of actors to act both within and independently of structures, to make conscious decisions that "'make a difference' […] that is, to exercise some sort of power" (Giddens 1984, 14). In social arenas, this kind of agency manifests either to sustain existing structures as useful rules and resources for living together, or to challenge and transform them through altered practices, circumvention, or resistance. For our discussion, this means that agency, while constituted within a network, emerges in the capability to recognize when automation should be resisted and alternative solutions found.

Technoscientific advances increasingly transfer human intentions not only to social actors and organizations but also to technical programmes and AI – just as Noah's smart wristband is intended to influence his behavior. When patients' or therapists' decision-making depends heavily on AI models in healthcare technologies, they may perceive their own agency as diminished. Concepts of agency have accordingly been systematically extended to include non-human entities, from sensors to entire sociotechnical networks. ANT, STS, and later new materialist and posthumanist theories highlight that human and non-human agency emerges within networks of interacting and mutually defining relations. For example, Kropp et al. (2026) show how socio-digital co-design practices shape building projects and reorganize coordination – always depending on socio-technical interactions. Nevertheless, technical or material agency raises questions of intentionality, freedom of action, and responsibility.

With the aim of clarifying responsibility within sociotechnical networks dealing with AI, Johnson and Verdicchio (2019) propose distinguishing causal, intentional and triadic agency. Their argument is that while users often depend on interaction with technological artifacts to carry out intentional actions, accountability arises only within the triad of users, system designers and computational artifacts. In this triad, they conclude, the crucial distinction is that "AI is computational whereas intentions are not" (Johnson and Verdicchio 2019, 645) – a claim that neuroscientists would not necessarily endorse. Notably, their analysis largely overlooks the fact that humans frequently behave automatically, at times akin to computational devices.

Noah's compulsive washing illustrates the difficulty of drawing a line between free will and determined behavior. Similarly, the algorithmic control of his wristband may be entirely determined, or it may adjust thresholds 'unsupervised' through machine learning as it continues to be used, acquiring a degree of autonomous decision-making. The conceptual starting point becomes even more complex when we consider that Noah's hand washing was initially subjectively meaningful but later became compulsive. In fact, the wearable is intended to help him regain the capability to make conscious decisions. Here, a mutually constitutive *co-agency* emerges between Noah, the wristband and his therapists. While the wearable initially responds to programmed threshold values with actuation, algorithmic processing of user data becomes increasingly opaque, and machine autonomy grows as computational learning expands across users. Responsibility shifts gradually: the wristband is expected to adapt autonomously and support Noah's development to the point where he no longer needs it. Otherwise, if the wristband assumes too much control, therapeutic tools risks reinforcing disorder-related dependency. From an external perspective, this represents a decentring of the human subject within technoscientific therapy, which can still be the free will of the patients. The long-term course of mutual influence therefore becomes crucial.

Studies in the history and sociology of science and technology further illuminate distributed agency by accounting for interactive trajectories. Pickering (1993) provided a constructive clarification by describing human and material agency as "reciprocally engaged by means of a dialectic of resistance and accommodation" (Pickering 1993, 559), enacted in an iterative dialectic of tentative enabling versus technical or material non-feasibility. Pickering considers human agency as emerging temporally in practice, i.e., in a socially embedded, mostly pre-reflective but skilful enactment of actions that can be enabled or restricted by technical devices and other material agents. Although humans differ from non-humans precisely in that their actions have intentions, both human and technical agency are temporally emergent in response to technical "affordances" and cultural "constraints" as Leonardi (2011) later specified. For example, the researchers might seek to "accommodate" high performance of the intelligent wristband by manipulating the granularity of measurements through material and sensor quality, while at the same time metadata and underlying programming, potentially drawn from software libraries with specific but un-known restrictions, might resist some of the desired adaptations.

Conversely, Noah and his family learn from the timing and frequency of the countermeasures triggered by the wristband which situations are particularly stressful for him, interacting with the technology in ways that shape emergent therapeutic goals. However, if they were to start avoiding potentially stressful situations, the therapists would regard this outcome as undesirable, as it would actually reinforce Noah's anxious feelings and restrict his freedom. Since the "trajectories of emergence of human and material agency are constitutively enmeshed in practice by means of a dialectic of resistance and accommodation" (Pickering 1993, 567), the goals of researchers, therapists and patients partly emerge only in interaction with the technology. For these reasons, we regard networked social arrangements as sociotechnical systems or as socio-materiality (Leonardi 2012) in order to highlight the organized interweaving of cultural-institutional and technical-material conditions in their development.

The wristband under discussion exists as one element within a tightly knit network of, among other things, behavioral automatisms considered dysfunctional, selected therapeutic approaches, family expectations and tolerances, digitally proposed solutions, sensors and sensor data, cost pressures and data protection rules in therapeutic settings. Social, legal, cultural, economic, algorithmic and material factors shape the network. Each component could have been defined differently, and the emerging network is the relational result of ongoing coordination and adaptation. It must be actively and continuously maintained, particularly in response to potential 'dissidents' whose interests or needs are not adequately met within the network (Callon 1986). These may include human factors, such as dissatisfied patients; technical issues, such as material problems with the wristband itself; or economic factors, such as increasing reliance on proprietary AI-devices and data-analysis techniques provided by large IT companies.

The mutually constitutive, networking character of the interactions makes it ineffective to either assign responsibility to individual components or determine an *a-priori* distribution of responsibility between human, organizational and technical factors, especially when taking into account what is all too easily referred to as 'artificial intelligence'. In the case study, AI seems to emerge from the algorithmic processing of data on skin resistance and heart rate variability, impulse strengths and a child's reaction patterns. However, material properties, metadata, data protection regulation and psychological knowledge simultaneously influence the algorithmic learning. Algorithms, data and AI bring about two different effects within the therapeutic network by acting as both intermediaries and mediators, to use a distinction made in ANT. As intermediaries, they transport "meaning or force without transformation: defining its input is enough to define its output" (Latour 2005, 39), providing a connection and enabling transaction. As mediators, by contrast, they "transform, translate, distort, and modify the meaning or the elements they are supposed to carry. [...]. No matter how apparently simple a mediator may look, it may become complex; it may lead in multiple directions."

Algorithms and (transactive) AI almost systematically act as mediators that not only process and pass on information, but also select, weigh and transform it when creating a connection. Occasionally, their operations result in bias or discrimination, possibly towards overweight or elderly patients who do not meet the usual standards, but always the operations mediate social processes and create a gap between what is consciously designed and what the social, ethical and legal implications are (Mittelstadt et al. 2016; Baumgartner et al. 2023). While activated with therapeutic intent, the wristband has the capacity to act without direct therapeutic intervention or control and AI may operate independently of reflective intention, influencing outcomes and shifting responsibility within the network (Leonardi 2012, 36).

Over time, social and technical agencies co-evolve, forming an "integrated organizational structure" (Leonardi 2012, 37). In digital therapy, sensors, algorithms, and organizational practices interlock, producing both new technologies and new organizational forms by being "recursively implicated" (Leonardi 2011, 150). The wristband's design is shaped by existing therapeutic priorities and prior technological solutions which constrain and enable future developments. Social and technical agencies are constitutively entangled, extending even to biological functions. This includes the emergence of digitally mediated therapeutic routines, in which biological, technical and organizational elements are jointly patterned rather than separately developed. Neither side is starting from scratch or limited to a particular status quo, but the successful entanglement of the two agencies creates infrastructures that people and organizations will draw on when making future decisions (Leonardi 2011). In this sense, organizational arrangements and technical configurations become mutually cemented: therapeutic quality criteria, existing control codes taken from previous applications and data documentation formats gradually materialize in the artefact, while the artefact itself prestructures subsequent therapeutic action. The initially chosen design of the wristband and the data processing will afford certain uses and actions and constrain others, so that further developments of digital therapy will be influenced by variants originally specified.

Almost all social phenomena are also technical (and biological), and no technical phenomenon exists that is not also social (cf. Orlikowski 2007, 1437). Yet, while both domains are intertwined, humans can deliberately change technology, whereas technology cannot change itself. For this reason, responsibility cannot readily be delegated to AI components; rather, explicit forms of technology assessment and governance are required. When it comes to the emerging capacities and constraints, in particular of taking responsibility, the use of AI, often portrayed as superior and independent of social arbitrariness, shifts accountability toward supposed 'intelligent' technological components, while social factors, such as the lack of therapy places, are obscured despite being constitutive.

Ultimately, certain actors authorize the introduction of AI – and none of those actors is AI itself. For this reason, concepts of human "head status" and technological "complement status" (Taylor et al. 2001, cited in Leonardi 2011) highlight the role of authorizing agencies in reconfiguring responsibility.

> It is not that Agent 1 (some human) had purposes and Agent 2 (tool) does not (they both incorporate purposes, and both have a history that is grounded in previous mediations), but we generally attribute head status to the human agent and complement status to the tool. In this way, the human subject's agency is given recognition even as the subject's agency (the capability of acting) is constituted objectively, as an actor in mediated communication with others. (Taylor et al. 2001, 71, cited from Leonardi 2011, 150)

In this sense, the decisive moment is the organizational decision to integrate AI into pre-existing responsibility structures, not any intrinsic capability of the system itself.[2] Such reconfigurations are typically enacted collectively rather than individually. Often, existing alliances derive their mandate from integrated expertise or the constitutive authorization of a joint research project.

We have argued that automation is not a technical replacement for human action but an element of a mutually responsive transformation of social and technical agencies within sociotechnical systems. Not every trajectory of interlinking is conceivable, however. Automation and organizations evolve in constitutive interaction, gradually blurring boundaries of agency, responsibility, and control. Our analysis therefore draws attention to the patterned, yet open-ended, stabilization of heterogeneous networks and to the emergent, hybrid forms of *co-agentic* capacities for action and governance that result. Control and accountability are distributed and embedded within the network, which makes ethical reflection not an external add-on but a practical engagement with this indivisible yet organized co-agency.

## 4 Conclusion: Who Cares? Guidelines for Distributed Responsibility in Hybrid Networks

Empirical studies on sensor-assisted therapeutic support are promising. Childhood obsessive-compulsive disorders (OCD) result in significant impairments in the daily lives of affected children and their families and, without adequate, guidelines-based

2 NB: The shift from more social to more technical responsibility cannot be assessed as inherently positive or negative. There are cases where a therapist may become too personally involved with a patient, which could hinder appropriate advice if it risks causing distress. An algorithm, by contrast (at least for now), is free from such social modulation.

treatment, carry a high risk of chronicity. The first-line treatment of choice is cognitive behavioural therapy (CBT), through which patients acquire effective strategies for managing compulsive symptoms. Successful treatment further depends on the implementation of exposure exercises in the child's natural environment. The use of a technology-based approach to enhance treatment in the children's everyday lives requires, on the one hand, data regarding the level of psychological tension, which can be collected using wearable sensor systems and leveraged for individualized interventions. On the other hand, it is essential that young patients and their families can trust that these data are handled responsibly, that their age-appropriate accuracy is ensured, and that they are applied in a therapeutically sound manner. There is widespread concern about sensitive personal data appearing on the internet, incorrect conclusions being drawn from flawed sensor measurements, and technology-based therapeutic approaches displacing conventional methods merely on the basis of cost-effectiveness, potentially carrying the risk of unintended consequences.

In Section 3, we clarified that responsibility can only emerge within the network of heterogeneous components, making accountability a distributed evaluative attribution that can neither be fully assumed by a single entity nor transparently delegated. Rather, it arises as a distributed capacity to meaningfully control intended and unintended effects, and is dependent on careful design of the sociotechnical entanglements. This networked configuration generates the characteristic risk that individual actors – therapists, technicians, families – do not feel ethically responsible either, assuming that others, or institutions, will take responsibility. In order to ensure accountability and responsible, yet distributed control, ethical guidelines for *care* are required for these emergent networks. Such guidelines need to extend beyond individual and even interpersonal concerns and foreground the complex interrelations between responsibility, co-agentive entanglement, and mutual dependency within sociotechnical networks. Moreover, the considerations above highlight the importance of enabling independent decision making, even though technologies impose certain limits on human autonomy. The interactive evolution over time obscures how each component could have been defined otherwise, but once established, shapes the trajectory of the entire sociotechnical constellation.

Following Puig de la Bellacasa (2017), we argue that hybrid networks imply an 'ethico-political commitment' to the entanglements we co-produce, as an essential part of scientific knowledge-making (Puig de la Bellacasa 2017, 40). Care is *not* to be understood as a discrete, human-centred individual practice, but rather as an affective, ethical, and practical mode of engaging with the ongoing remaking of the world, inherently situated, relational, and hybrid. Traditional understandings of the ethical are displaced, as care emphasizes material interdependence and involvement, reflecting a decentred conception of human agency enacted through entangled co-agencies and sociotechnical practices. The necessary guidelines of care also encompass how sociotechnical stories of autonomous machine agency are told, who and what is recognized, marginalized, or neglected (cf. Suchman 2007; Haraway 1985). Accordingly, the ethical stakes concern not only actions within the network but also the narratives through which the resulting agency is rendered intelligible.

With this in mind, we derive guidelines for trustworthy technology development from our analysis of conditions for responsible co-agency in complex, extended networks involving inter- and transactive technologies and multiple interests (cf. Rammert 2016). The guidelines are meant to empower individuals to engage in these networks with confidence, and should support alliances, that is the collectives shaping these sociotechnical networks, in critically identifying gaps and deficiencies in accountability, with the aim of mitigating harm and manging risk in the sense of health technology assessment.

The guidelines address seven topics, as follows:

1. *Revisability*: Actions, motivations, and their justification evolve not only over time but also through their entanglement with technical affordances and constraints. Sociotechnical networks, along with new configurations of practice, therefore, give rise to adapted justifications. The more therapeutic formats are technicised, the harder it will become to take responsibility for minority concerns or extraordinary needs. Habermas (1971) analysed the interplay between knowledge and interest, warning against technocratic tendencies in which what is deemed desirable becomes dependent on what is technically feasible, thereby marginalizing values such as solidarity and their democratic definition. Co-responsibility in sociotechnical systems requires keeping decision-making processes open to evolving dynamics of value formation. Design must remain explainable and revisable so that future therapeutic responses can be tailored to individual and societal concerns rather than predetermined by inherited technical specifications.
2. *Social responsiveness.* Standardization and automation cannot be confined to isolated human-technology interactions; their logic extends to adjacent practices and social settings. Not only do Noah's constraints automatically find reflection in the family and lead to responsive adaptations, but the automation of one action also makes automation in another area more likely. The step from an intelligent wristband to an intelligent collar, to be applied when someone deviates from the norm, may seem extreme but actually becomes smaller with each additional application. Responsibility therefore requires attention to potential extensions and generalizations of local solutions, including their political economy. This implies caring about what technoscientific promises are

made and how they reflect particular interests. Current trajectories of automation policy – whether within the Trump administration or the Chinese State Council – can serve as reminders that sociotechnical interventions are never normatively neutral.

3. *Internal and external coherence.* The circumstances under which an application is used vary not only within the system itself, whether Noah is sitting in school or playing sports, but also in its external context. For example, in research settings, data processing might be confined to protected servers of the participating institutions, but, once commercialized, can migrate to external cloud infrastructures, a transition that often remains invisible. Yet, such seemingly minor changes can have a significant impact on fundamental protected goods, such as participation in social life or the right to privacy. Accountability therefore requires systematic attention to how infrastructural changes affect normative commitments and protected interests. The full range of system-determining options must be explored as comprehensively as possible and remain flexible to allow coherent functioning, transparent adaptation and variability.
4. *Multicausality*: Influencing factors and constellations of conditions are not monocausal. For instance, skin conductance and heart rate variability may change in discordant ways, sending contradictory signals to the sensors. Only control rules that are sufficiently saturated with experience can then determine which changes should trigger an impulse. Developing such rules requires sustained empirical work on relevant correlations (the frequently overlooked co-agency of control algorithms) and on the epistemic limits of simulation techniques and data sets, which often lack the necessary granularity. If systems are fixed prematurely, based on falsely assumed causalities, this can result in severe harm (cf. Kropp and Wortmeier 2021). Therefore, data processing, underlying models, and calibrations must be continuously validated.
5. *Unintended consequences*: The promised liberation from repetitive or unpleasant actions, such as compulsive hand washing, can turn into a new form of constraint through technological mediation, for example, via the constant monitoring enabled by the wristband. If the consequences of such co-agency are not responsibly assessed and incorporated into development, individual self-efficacy may be weakened. Technical dependency can evolve into a form of heteronomy, whereby socio-technical systems introduce new and potentially inappropriate coercions. For instance, the wristband may prompt stress-relieving cognitive exercises even when Noah is participating in a sporting activity or experiencing an early moment of physical intimacy. A negative consequence could be that Noah develops little individual awareness due to the system's rigidity and reduced opportunities for emotional learning without technical

support. Just as pilots practice landing an aircraft without autopilot in simulators, users of complex technical systems and AI must remain capable of making and implementing responsible decisions independently. Likewise, therapists must be able to assess their patients' progress both with and without documented sensor data and remain accountable to their patients' well-being. There must be the ability to integrate, interpret or disregard data, depending on situational demands: otherwise, automation threatens to erode both co-agency and human agency.

6. *The precautionary principle for complex entanglements*: Recognizing the mutual constitutive entanglements of diverse elements in ongoing sequential trajectories heightens awareness of the heterogeneity of the networked components, their precedents and side effects, and the risk of error propagation. Automated systems, particularly those incorporating unsupervised machine learning, can be assumed to exhibit complex systemic interactions that easily exceed the bounds of linear reasoning and disciplinary knowledge (Lenk 2017). Responsible control over intended procedures therefore requires integrative, multi-disciplinary deliberation. Similarly, the responsible mitigation of AI-related risks requires alignment between control tasks and accountability structures among those involved in system development and implementation (Grote et al. 2024).

   In this context, control refers to the capacity to influence a situation such that it unfolds or remains in accordance with the intentions of those held as (co-) responsible agents (Flemisch et al. 2016, 73). Admittedly, control relationships in dynamic networks tend to be overly complex, involving a multiplicity of agents with significantly varying degrees of agency. To navigate these networks, detailed mapping of interactions and potential side effects, as pursued, for example in responsibility-related decision ladders (Townsend et al. 2025) is indispensable for precautionary reasoning. To enhance the prospects for responsible and democratic technological development, these interrelationships and their potential direct and indirect consequences must be considered comprehensively within the broader public.

7. *Human-centred design*: Kranzberg's (1986, 547) first law reminds us: "Technology is neither good nor bad; nor is it neutral." For four decades, this law has served in technology assessment as a reminder not to rush into viewing technologies as either improvements or problems, and certainly not as disinterested entities without significant societal consequences. Technologies embody values, reflect social priorities and stabilize assertive interests (Howcroft and Taylor 2022). Given the distributed nature of agency, human-centred design provides a critical normative orientation: it demands that hybrid networks be examined for the kinds of human and non-human

agencies they enable, constrain or systematically privilege. Human-centred thus becomes a mode of caring for the humanness of the hybrid networks we co-create, requiring case-by-case contextualization.

There is no doubt that digital health technology can help to make life easier for many patients and, as in the example discussed, offer the opportunity for unrestricted participation in social life. The question, therefore, is not so much whether these digital tools should be used or not, but rather how they can be rendered trustworthy. Continued adaptability, social responsiveness, internal and external coherence, awareness of multicausality and necessary degrees of independence, together with the precautionary principle and human-centred design define, as stated above, what we believe to be the seven guidelines for assuming shared co-responsibility in hybrid networks with various forms of automation.

For these reasons, the paper does not attempt to provide a definitive answer on how to organize responsibility for automation and a responsible use of AI. Instead, it seeks to systematize the ways in which the capacity to take responsibility is constrained and redefined within networked constellations of co-agency. By articulating these guidelines, we aim to elucidate the structural conditions under which an ethico-political commitment (Puig de la Bellacasa 2017) can inform the responsible design of sociotechnical networks, attentive to their potential implications and normative ambivalences.

**Research funding:** Part of this work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy – EXC 2120/1 – 390831618.

# References

Alt, Annika, Carolin Klein, Anja Pascher, Annette Conzelmann, Franziska Kosel, Jan Kühnhausen, et al. 2025. “Acceptance of a Sensor-Based Online Psychotherapy for Adolescents with Obsessive-Compulsive Disorder (SSTeP-KiZ).” *Digital Health* 11. https://doi.org/10.1177/20552076251327046.

Antsaklis, Panos. 2020. “Autonomy and Metrics of Autonomy.” *Annual Reviews in Control* 49: 15—26.

Bainbridge, Lisanne. 1983. “Ironies of Automation.” *Automation* 19 (6): 775—9.

Baumgartner, Renate, Payal Arora, Corinna Bath, D. Burljaev, K. Ciereszko, B. Custers, et al. 2023. “Fair and Equitable AI in Biomedical Research and Healthcare: Social Science Perspectives.” *Artificial Intelligence in Medicine* 144. https://doi.org/10.1016/j.artmed.2023.102658.

Brödner, Peter. 2019. “Grenzen und Widersprüche der Entwicklung und Anwendung ‘Autonomer Systeme’.” In *Autonome Systeme und Arbeit: Perspektiven, Herausforderungen und Grenzen der Künstlichen Intelligenz in der Arbeitswelt*, edited by Hartmut Hirsch-Kreinsen, and Anemari Karacic. Transcript Verlag.

Brynjolffson, Erik, and Andrew McAfee. 2014. *The Second Machine Age: Work, Progress and Prosperity in a Time of Brilliant Technologies*. W.W. Norton & Co.
Burrell, Jenna. 2016. "How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms." *Big Data & Society* 3 (1): 1–16.
Callon, Michael. 1986. "Some Elements of a Sociology of Translation: Domestication of the Scallops and the Fisherman of St. Brieuc Bay." In *Power, Action and Belief: A New Sociology of Knowledge?* edited by John Law. Routledge.
Flemisch, Frank, David Abbink, Makoto Itoh, Marie Pierre Pacaux-Lemoine, and Gina Weßel. 2016. "Shared Control is the Sharp End of Cooperation: Towards a Common Framework of Joint Action, Shared Control and Human Machine Cooperation." *IFAC-PapersOnLine* 49 (19): 72–7.
Giddens, Anthony. 1984. *The Constitution of Society: Outline of the Theory of Structuration*. Polity Press.
Grote, Gudela, Sharon K. Parker, and Kevin Crowston. 2024. "Taming Artificial Intelligence: A Theory of Control-Accountability Alignment Among AI Developers and Users." *Academy of Management Review*, https://doi.org/10.5465/amr.2023.0117.
Habermas, Jürgen. 1971. *Knowledge and Human Interests*. Translated by Jeremy J. Shapiro. Beacon Press.
Haraway, Donna. 1985. "Manifesto for Cyborgs: Science, Technology, and Socialist Feminism in the 1980's." *Socialist Review* 80: 65–108.
Howcroft, Debra, and Phil Taylor. 2022. "Automation and the Future of Work: A Social Shaping of Technology Approach." *New Technology, Work and Employment* 38 (2): 351–70.
Johnson, Deborah G., and Mario Verdicchio. 2019. "AI, Agency and Responsibility: The VW Fraud Case and Beyond." *AI & Society* 34: 639–47.
Klein, Carolin, Annika Alt, Anja Pascher, Jan Kühnhausen, Lennart Seizer, Winfried Ilg, et al. 2024. "Cognitive Behavioral Therapy for Pediatric Obsessive-Compulsive Disorder Delivered via Internet Videoconferencing: A Manualized Sensor-Assisted Feasibility Approach." *Child and Adolescent Psychiatry and Mental Health* 18: 154.
Kranzberg, Melvin. 1986. "Technology and History: Kranzberg's Laws." *Technology and Culture* 27 (3): 544–60.
Kropp, Cordula, Yana Boeva, and Kathrin Braun. 2026. "Socio-Digital Co-Design Practices: A Case Study on Human-Computer Entanglements in Architecture." *Science & Technology Studies* 39 (1): 21–40.
Kropp, Cordula, and Ann-Kathrin Wortmeier. 2021. "Intelligente Systeme für das Bauwesen: überschätzt oder unterschätzt?" In *Digitalisierung souverän gestalten*. *Innovative Impulse im Maschinenbau*, edited by Ernst A. Hartmann. Springer Vieweg.
Latour, Bruno. 2005. *Reassembling the Social: An Introduction to Actor-Network Theory*. Oxford University Press.
Lenk, Hans. 2017. "Ethics of Responsibilities Distributions in a Technological Culture." *AI & Society* 32: 219–31.
Lenk, Hans, and Matthias Maring. 2001. "Responsibility and Technology." In *Responsibility – the Many Faces of a Social Phenomenon*, edited by Ann-Elisabeth Auhagen, and Hans-Werner Bierhoff. Routledge.
Leonardi, Paul M. 2011. "When Flexible Routines Meet Flexible Technologies: Affordance, Constraint, and the Imbrication of Human and Material Agencies." *MIS Quarterly* 35 (1): 147–67.
Leonardi, Paul M. 2012. "Materiality, Sociomateriality, and Socio-Technical Systems: What Do These Terms Mean? How Are They Different? Do We Need Them?" In *Materialiy and Organizing. Social Interaction in a Technological World*, edited by Paul M. Leonardi, Bonnie A. Nardi, and Jannis Kallinikos. Oxford University Press.
Luhmann, Niklas. 1997. *Die Gesellschaft der Gesellschaft*, Vol. 1. Suhrkamp.

Milano, Silvia, Mariarosario Taddeo, and Luciano Floridi. 2020. "Recommender Systems and their Ethical Challenges." *AI & Society* 35: 957–67.

Mittelstadt, Brent D., Patrick Allo, Mariarosaria Taddeo, Sandra Wachter, and Luciano Floridi. 2016. "The Ethics of Algorithms: Mapping the Debate." *Big Data & Society*: 1–21. https://doi.org/10.1177/2053951716679679.

Morley, Jessica, and Luciano Floridi. 2025. "The Ethics of AI in Healthcare: An Updated Mapping Review." In *Ethics and Medical Technology*, Vol. 113, edited by Matthew C. Altman, and David Schwan. The International Library of Bioetics.

Orlikowski, Wanda J. 2007. "Sociomaterial Practices: Exploring Technology at Work." *Organization Studies* 28 (9): 1435–48.

Pickering, Andrew. 1993. "The Mangle of Practice: Agency and Emergence in the Sociology of Science." *American Journal of Sociology* 99 (3): 559–89.

Puig de la Bellacasa, María. 2017. *Matters of Care: Speculative Ethics in More than Human Worlds*. University of Minnesota Press.

Rahman, K. Sabeel, and Kathleen Thelen. 2019. "The Rise of the Platform Business Model and the Transformation of Twenty-First-Century Capitalism." *Politics & Society* 47 (2): 177–204.

Rammert, Werner. 2016. *Technik — Handeln — Wissen. Zu einer pragmatistischen Technik- und Sozialtheorie*. Springer Fachmedien.

Rammert, Werner, and Ingo Schulz-Schaeffer. 2002. "Technik und Handeln: wenn soziales Handeln sich auf menschliches Verhalten und technische Artefakte verteilt." In *Können Maschinen handeln? Soziologische Beiträge zum Verhältnis von Mensch und Technik*, edited by Werner Rammert, and Ingo Schulz-Schaeffer. Campus.

Saurwein, Florian. 2019. "Automatisierung, Algorithmen, Accountability. Eine Governance Perspektive." In *Maschinenethik. Normative Grenzen autonomer Systeme*, edited by Matthias Rath, Friedrich Krotz, and Matthias Karmasin. Springer VS.

Sayes, Edwin. 2014. "Actor-Network Theory and Methodology: Just What Does it Mean to Say that Nonhumans Have Agency?" *Social Studies of Science* 44 (1): 134–49.

Stinson, Catherine. 2021. "Algorithms are Not Neutral: Bias in Collaborative Filtering." *AI and Ethics* 2 (4): 763–70.

Suchman, Lucy A. 2007. *Human-Machine Reconfigurations: Plans and Situated Actions*, 2nd ed. Cambridge University Press.

Suchman, Lucy A. 2023. "The Uncontroversial 'Thingness' of AI." *Big Data & Society* 10 (2). https://doi.org/10.1177/20539517231206794.

Taylor, James R., Carol Groleau, Lorna Heaton, and Elisabeth Van Every. 2001. *Computerization at Work: A Communication Perspective*. Thousand Oaks, CA: Sage Publications.

Townsend, Beverley, Katie J. Parnell, Sinem Getir Yaman, Gabriel Nemirovsky, and Radu Calinescu. 2025. "Normative Conflict Resolution Through Human-Autonomous Agent Interaction." *Journal of Responsible Technology* 21: 100–14.

van den Berg, Bibi. 2008. "Self, Script, and Situation: Identity in a World of ICTs." In *The Future of Identity in the Information Society. Privacy and Identity. IFIP — The International Federation for Information Processing*, Vol. 262, edited by Simone Fischer-Hübner, Penny Duquenoy, Albin Zuccato, and Leonardo Martucci. Springer.

Winfield, Alan F., Katina Michael, Jeremy Pitt, and Vanessa Evers. 2019. "Machine Ethics: The Design and Governance of Ethical AI and Autonomous Systems." *Proceedings of the IEEE* 107 (3): 509–17.