AI and Ethics
https://doi.org/10.1007/s43681-022-00184-2

REVIEW

Responsibility in Hybrid Societies: concepts and terms

Stefanie Meyer[1] · **Sarah Mandl**[2] · **Dagmar Gesmann-Nuissl**[1] · **Anja Strobel**[2]

Received: 22 March 2022 / Accepted: 30 May 2022

Abstract
With increased digitalization and new technologies, societies are expected to no longer only include human actors, but artificial actors as well. Such a future of societies raises new questions concerning the coexistence, tasks and responsibilities of different actors. Manifold disciplines are involved in the creation of these future societies. This requires a common understanding of responsibility, and of definitions of actors in Hybrid Societies. This review aims at clarifying aforementioned terms from a legal and psychological perspective. Building from this common ground, we identified seven capacities in total which need to be met by actors in societies to be considered fully responsible, in both a legal and moral sense. From a legal perspective, actors need to be autonomous, have capacity to act, legal capacity, and the ability to be held liable. From a psychological perspective, actors need to possess moral agency and can be trusted. Both disciplines agree that explainability is a pivotal capacity to be considered fully responsible. As of now, human beings are the only actors who can, with regard to these capacities, be considered morally and legally responsible. It is unclear whether and to which extent artificial entities will have these capacities, and subsequently, whether they can be responsible in the same sense as human beings are. However, on the basis of the conceptual clarification, further steps can now be taken to develop a concept of responsibility in Hybrid Societies.

Keywords Hybrid Societies · Responsibility · Moral agency · Legal agency · Capacities of responsibility · Artificial agents

1 Introduction: responsibility in Hybrid Societies

With increased digitalization and new technologies, societies are expected to no longer only include human actors, but artificial actors as well—actors of different kind will live together in so called Hybrid Societies. Such a future of societies raises new questions concerning the coexistence, tasks and responsibilities of different actors. Human–machine hybrids, autonomous robot systems and artificial intelligence (AI) will find their way into numerous areas of life, taking tasks that were previously reserved for humans. With the rapid development in the field, artificial actors will act and decide with a greater level of autonomy than before, will sometimes perhaps make decisions that a human being would not have made or make independent decisions that are detached from the original program of the developer [1]. Such decisions often will not be neutral with respect to purpose and societal impact and will even imply to make choices of ethical or moral relevance, respectively [2]. Considering this, responsibility as an overarching construct gains importance. The often-quoted phrase 'With great power comes great responsibility' undoubtedly applies to autonomously acting artificial agents as well. But what does responsibility, or responsible action, mean in this context? Confusions already arise with terms that were tailored to the purely human side but are now transferred to the technical level, as for example 'person', 'legal subject', 'fundamental rights', or 'guilt' [3]. Human–machine interaction can only be guaranteed in a trustworthy manner if there are reliable

Stefanie Meyer and Sarah Mandl have contributed equally to this work.

✉ Stefanie Meyer
stefanie.meyer@wiwi.tu-chemnitz.de

✉ Sarah Mandl
sarah.mandl@psychologie.tu-chemnitz.de

[1] Professorship of Private Law and Intellectual Property Rights, Faculty of Economics and Business Administration, Chemnitz University of Technology, Thüringer Weg 7, 09126 Chemnitz, Germany

[2] Professorship of Personality Psychology and Assessment, Faculty of Psychology, Chemnitz University of Technology, Wilhelm-Raabe-Str. 43, 09120 Chemnitz, Germany

Published online: 27 June 2022

rules for the responsibility of the respective individuals. In this context, one cannot only talk about possible errors, but the considerations have to go much further: Can existing approaches and models from different disciplines adequately capture the decisions and actions of these hybrid systems, or do we need to develop more comprehensive and shared concepts of responsibility specifically for the interaction of humans and machines in Hybrid Societies that go beyond existing ones? Do specific types of hybrid systems also require different concepts of responsibility, or can overarching and generalizable concepts be developed and applied? Such questions can only be answered by approaching them from different perspectives taking account insights from different disciplines. With this review, we aim at integrating the psychological and the legal perspective as Radbruch [4] described the inseparable connection between moral concepts and jurisprudence as early as 1932.

> 'Only morality is capable of establishing the binding nature of law. [...] There can be no question of legal norms, legal requirements, legal validity, legal obligations, as long as the requirement of law to the individual conscience is not endowed with moral binding force.'

By drawing from legal and psychological expertise, terms are discussed and defined from these perspectives and a coherent taxonomy and a working model for responsibility in Hybrid Societies are derived.

1.1 Necessity of a literature review

There are two ways to determine the meaning of a word in a generally valid way: by mere language use or by creating a definition [5]. Ultimately, however, language use is by far the more important, since it takes into account not only the intention of the user, but above all the understanding of those who are potential addressees of a term [6]. However, it may also happen that certain terms are understood differently, possibly even contrary, even within a linguistic community. Various scientific disciplines have created their own definitions for terms that are unclear and difficult to delimit. With respect to 'Hybrid Societies' it is noticeable that in scientific literature in different disciplines even terms like 'robot', 'machine', 'artificial intelligence', 'machine intelligence', or 'computer', are not always used in the same way [7–10]. Therefore, the aim of this review was to examine the basic concepts of agents, the entities themselves and the system of accountability, and to search the existing literature which terminologies are used from a psychological and legal point of view. As a first step, we identified and further specified which entities are involved in Hybrid Societies. We then identified the necessary preconditions for responsibility in the legal and psychological sense by reviewing the current literature. Finally, we identified the necessary preconditions for responsibility and applied them to the different actors in Hybrid Societies. In the process of combining findings from different fields, it became clear that a large number of terms are used, some of which are very closely related and some of which have completely different meanings. In addition, terms describing the same phenomenon may differ from each other making interdisciplinary work on Hybrid Societies more difficult than it needs to be. Although a common vocabulary may be difficult to achieve, a common understanding should be the goal for scientists dealing with questions in this field.

2 Methods

Literature research was carried out from June 2020 until November 2021. We included literature from 1797 (Kant) to 2022 (Mandl et al.), with the majority of papers stemming from 1999 to 2021. Apart from classical literature such as Kant (1797) and Hegel (1805–07), we included key publications from the last decades until now. To adequately capture the sheer abundance of disciplines involved in the description of Hybrid Societies, we used the snowball method. We started with recent publications involving the keywords 'artificial agent', 'ethical agent', 'moral agency', 'moral responsibility', 'machine ethics', 'artificial intelligence', 'robot law', 'robot rights', 'legal entity', 'legal personhood', 'responsibility for human–machine-interaction', 'liability', 'capacity to act', or 'legal capacity' and proceeded to other relevant titles from there. To counteract the disadvantage of only searching retrospectively, we used search for the terms 'moral agency', 'machine ethics', 'artificial morality', 'artificial moral agent', 'ethical agent', 'moral responsibility', 'robot law', 'robot rights', 'legal entity', 'legal personhood', 'liability', 'capacity to act', or 'legal capacity'. We restricted articles included in terms of language (English or German) (Fig. 1).

This left us with a total of $N = 163$ articles, books, commentaries, case law reviews, and book chapters from $N = 12$ disciplines to be included in this review. While mainly focusing on literature from law and psychology and related disciplines, we did not restrict research to these disciplines, but included a variety of them (Table 1).

3 Results

3.1 Actors of a Hybrid Society

To approach the issue of responsibility, in a first step, the actors of Hybrid Societies have to be identified and defined in more detail. By evaluating the actors of Hybrid Societies

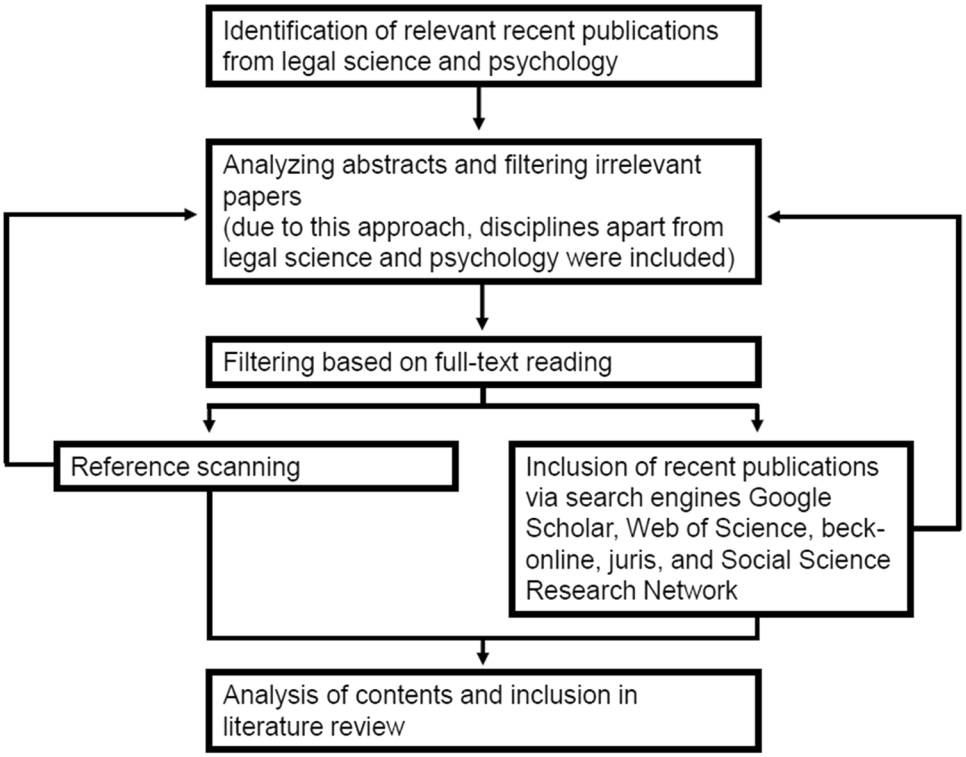

Fig. 1 Flowchart depicting the process of literature research

Table 1 Disciplines included in literature research

Discipline	Number of articles
Law	$N=37$
Psychology	$N=20$
Social Science	$N=3$
Philosophy	$N=45$
Robotics	$N=3$
Science-/Engineer Ethics	$N=1$
Engineering	$N=2$
Theology	$N=1$
IT/Computer Science	$N=14$
Political/Economic Studies	$N=8$
Physics	$N=3$
Communication Science	$N=2$
Others	$N=27$
	Total Number of articles: $N=162$

from a legal and psychological perspective we contrast these perspectives. For an overview see Fig. 2.

3.1.1 Human being

It may seem strange to define what a human being is. However, a closer look reveals that each discipline makes different typical connections to the entity of a human being; in some cases, different consequences are associated with the classification as a human being.

3.1.1.1 Legal science According to Kant [11], a person or a human being is the subject whom actions can be attributed to. The concept of person or legal subject can only mean a human being who is capable of responsibility, which in Kant's words means that the ability to act according to the categorical imperative must be given. In the following years since Kant's Metaphysics of Morals, a reorientation of the concept of person has taken place: the legal subject human being is no longer the bearer of actions just because they possess qualities such as rationality or physical mobility—they are the originator of his actions and their effects (and this qualifies him as a human being); the actions are attributed to them [12]. In the legal system, the human being also represents the central element of allocation, which is reflected in the fact that they are entitled to comprehensive fundamental and human rights. Of particular interest for the applicability of legal regulations is the attribution of typical human characteristics. Matthias considers a total of five criteria to be decisive: (1) intentionality, (2) receptivity to reasons, (3) second-order desires, (4) rationality, and (5) intended and foreseeable forms of action [13]. John attributes only three intrinsic qualities to human beings: (1) self-knowledge, (2) social competence, and (3) legal expediency [14]. The last point in particular is of immense importance from a legal perspective: judging whether one's actions are within the

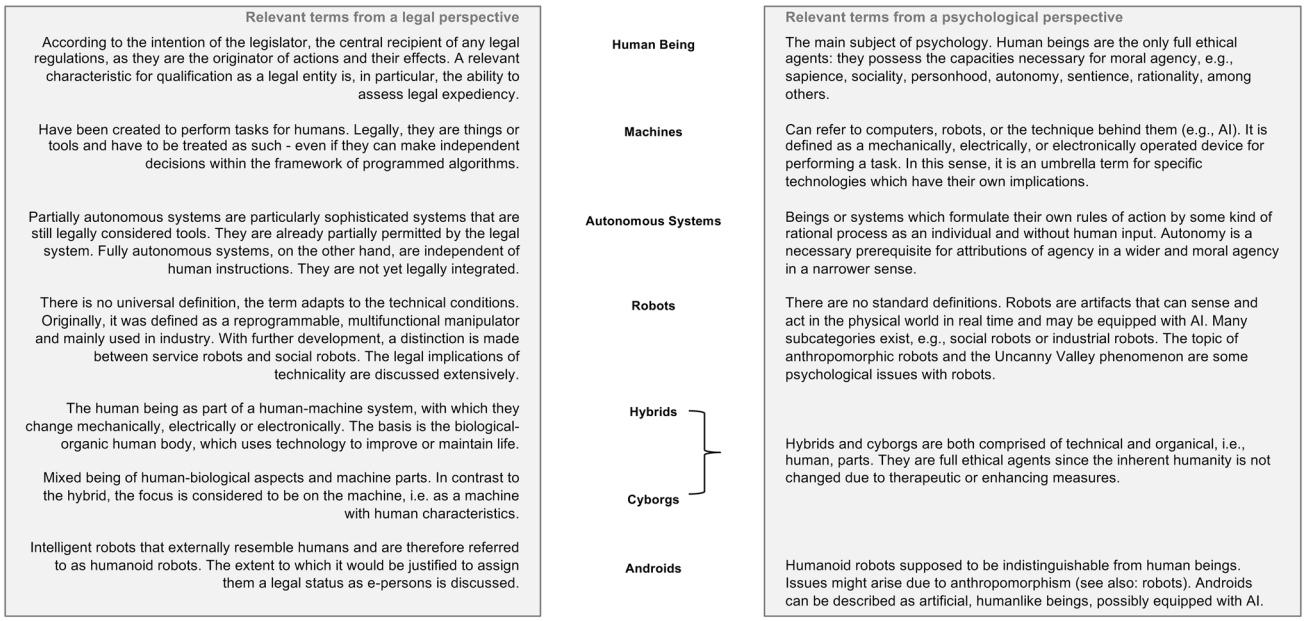

Fig. 2 Actors of a Hybrid Society

bounds of what is legally permissible requires a moral-legal ability to assess one's responsibility to third parties.

Consequently, these typically human characteristics, regardless of how many of them are adopted, are necessary for any entity operating in Hybrid Societies to ensure responsible coexistence. The prerequisite for legal responsibility is that humans are capable of bearing or assuming rights and duties [15].

3.1.1.2 Psychology Human beings are the main subject of psychology. They are the only full ethical agents; they make ethical decisions and justify them [16] and possess the capacities necessary for moral agency such as sociality and personhood, normative understanding, autonomy, sentience, rationality and action, and intentionality [17]. Even though humans might not be the gold standard for moral reasoning [18], they are the only beings currently existing which have sapience—and therefore moral agency—and can subsequently be held morally responsible [19].

3.1.2 Machines

3.1.2.1 Legal science Machines were originally designed to perform tasks that humans cannot perform or can only perform with great effort; or to make certain tasks easier or faster. Legally speaking, machines are therefore nothing other than a thing (Section 90 of the German Civil Code) or a tool (Art. 36 JICOSH; US OSH Act), which they were also designed to be [20]. Due to their degree of mechanization, they cannot execute programs other than those developed by humans [21], even if these programs are configured to reassign certain tasks. Machines do not have an inner world, rather they require code: even machines equipped with artificial intelligence can perceive all impressions only as data [21]. Therefore, machines or computer systems cannot be agents or legally capable of acting, no matter how automated, independent, or intelligent they may be [22].

3.1.2.2 Psychology Machines are tools which are self-sufficient, self-reliant, or independent [23] (as cited in Ref. [24]). The Merriam-Webster Dictionary defines a machine as a mechanically, electrically, or electronically operated device for performing a task [25]. In a wider sense, today, machine can refer to computers, the embodied artificial agents, that is, robots, or to the technique behind it, that is, AI, among others. Therefore, the term machine ethics refers to a machine that follows 'an ideal ethical principle or set of principles, that is to say, it is guided by this principle or these principles in decisions it makes about possible courses of action it could take' [26].

3.1.3 Autonomous system

3.1.3.1 Legal science The term autonomous system encompasses both partially autonomous and fully autonomous systems. The term is also highly indeterminate; Decker even describes these systems as being rich in unspoken presuppositions [27].

Partially autonomous systems are already permitted by the legal system (e.g., in the area of autonomous driving) [28], provided that their use is within the scope of the permitted risk and that the system is used appropriately.

Conceptually, these partially autonomous systems are rather tools, albeit remarkably sophisticated tools, used by humans [29]. The latter still has responsibility and control of the process and can take over the tasks of the system [29]. With regard to the classification as a tool or product, there is no consensus [30–32].

Fully autonomous systems, on the other hand, are not human tools. According to Vladeck, these should be machines that are used along with humans and can act autonomously, i.e., independent of direct human instruction [29]. This system obtains the intuition to act on the basis of its own analyses, on the basis of which it can make momentous decisions that were not always foreseeable for the human who programmed it [29]. What the establishment of these fully autonomous systems means is currently unclear and widely discussed in the literature [33]. While the term was initially used in a narrow context (e.g., in the context of autopilot), its use is now expanding to other systems (e.g., autonomous driving). This expansion of the use of the term is associated with the discussion of the autonomy of the systems in the legal sense (i.e., its status as a legal subject).

3.1.3.2 Psychology As autonomous systems are defined by their ability to be autonomous, from a psychological viewpoint, the definition of autonomy takes precedence. Autonomy refers to a being or system which formulates its own rules of action by some kind of rational process [34], the capacity to act as an individual [35], and without human input [36]. Autonomy furthermore encompasses the ability to critically reflect values, exert self-control and to set goals [17]. Autonomy is inextricably combined with the term of 'agency'. Agency includes different capacities such as thought, communication, planning, recognition, emotion, memory, morality, and self-control. These capacities define to which extend a character is capable of, for example, morality. Therefore, moral responsibility is tied to attributions of agency as well [37].

3.1.4 Robot

3.1.4.1 Common ground The term most commonly used for EDTs, but often in very different ways, is that of robot. This term is frequently applied to entities that are specifically discussed in this review, so there is overlap in language use that refers to this term. The term was first used by Czech author Karel Čapek in 1920 in his play R.U.R. Rossumovi Univerzální Robotí (2004) [38]. By today's standards, Čapek's robots would have been called androids, artificial humans, since they were constructed in a biochemical way and were indistinguishable from real humans. In general, the definition process started with a purely technical approach and was expanded step by step to include new components.

3.1.4.2 Legal science According to the VDI (Verein Deutscher Ingenieure) guideline of 1990 [39], the robot was defined as follows: 'A robot is a free and reprogrammable, multifunctional manipulator with at least three independent axes to move materials, parts, tools or special devices on programmed, variable paths to perform a wide variety of tasks.' This definition was essentially adopted for ISO (Internationale Organisation für Normung) Standard 8373 [40] to describe the industrial robot (see below). The Robot Institute of America (RIA) [41] also defines robots as programmable multipurpose handling devices for moving materials, workpieces, tools, or special equipment. The Japanese Robot Association [3] defines robots as handling devices that do not have a program but are guided directly by the operator; however, the term also includes 'intelligent robots', as devices that have various sensors and are thus able to adapt the program sequence automatically to changes in the environment.

Since robots are no longer designed intended to work merely as manipulators, but rather as intelligent machines that extend the human ability to move [42], Christaller has expanded the definition of a robot as follows: 'Robots are sensorimotor machines that extend the human ability to act. They consist of mechatronic components, sensors and computer-based control functions. The complexity of a robot differs significantly from other machines due to the greater number of degrees of freedom and the variety and scope of its behaviors' [43]. Bekey went a step further by saying that it should be a machine that perceives, thinks, and acts accordingly [44]. The entity 'robot' can be distinguished into two characteristics: (1) the physical characteristics. It needs sensors to perceive the environment, processors to perform certain cognitive functions and actuators to act in its environment [45]. In addition, it needs (2) software bots that determine the robot's behaviour through a computer code that defines the machine's scope for decision-making in the particular situation [32]. Due to the latter characteristic, the robot is often understood to be congruently with an autonomous system and is used interchangeably [45]; whereby with the robot the physical component has to be compellingly present. Balkin also does not make a distinction between the robot on the one hand and the artificial intelligence on the other (cf. above) [46]. As a result of the growing use of robots, a distinction has to be made between industrial robots and service robots. Industrial robots correspond to the basic definition of the VDI guideline and the *ISO standard* mentioned at the beginning. They are considered to have certain capabilities, such as high speed and precision, high force and quasi-unlimited repeatability of movements [27]. Service robots, on the other hand, perform useful tasks for humans, society or institutions, with the exception of tasks in automation technology [40]. Their focus is less on the autonomous execution of tasks than on cooperation and interaction with

humans [27]. According to Calo, these robots also have a social meaning for humans, although they do not react to robots as if they were humans, but rather as they do to animals [47]. There is a further development within the entity of the robot: depending on the area of application, they are becoming increasingly humanoid. This can be summarized under the concept of the 'android', which will be discussed below.

3.1.4.3 Psychology As of now, no standard definition of 'robot' exists, even though different organizations and authors have working definitions. Robots come in many forms and contexts, so for clarification, it is indispensable to define a subcategory when speaking of robots. As it is out of scope of this review, the authors exclude military-issue robots a.k.a. military drones. The Robotics Industries Association (RIA) provides working definitions for 'collaborative robots': robots designed for direct interaction with a human within a defined shared workspace, and 'industrial robots': an automatically controlled, reprogrammable multipurpose manipulator programmable in three or more axes which may be either fixed in place or mobile for use in industrial automation applications [48]. Bryson and Winfield proposed that robots are artifacts that sense and act in the physical world in real time which not only includes industrial or social robots, but also smartphones as (domestic) robots [35]. From a psychological perspective, two kinds of robots require stronger focus: social and industrial robots, for different reasons. Social robots are physically embodied artificial agents that have features which enable humans to perceive an agent as a social entity, for example eyes and other facial features, and are capable of interacting with humans via a social interface. Additionally, Bartneck and Forlizzi proposed that social robots need to follow behavioural norms expected by people with whom the robot interacts. Social robots are able to communicate verbal and/or non-verbal information to humans and can be designed in different ways such as abstract, humanoid, anthropomorphic, or non-humanoid [49–51]. They interact with humans in different settings. So, users of these social robots need to be certain that they are interacting with an artificial, not with a human being at all times.

3.1.5 Hybrid

3.1.5.1 Legal science Zimmerli used the synonym of the 'centaur' to describe what is explained here under the term 'hybrid': According to this, we are beings who live in symbiotic connection with the technologies that surround us. Accordingly, humans are part of a human–machine system in that they change mechanically, electrically or electronically [51]. If this description is taken literally, the definition he creates already includes the fact that we use technical devices (such as a cell phone, computer, etc.); a fusion of man and technology is therefore not necessary.

Nevertheless, it is also possible to speak of a hybrid when life-organizing or life-sustaining systems are integrated into the body of a human being. Faßler describes the hybrid as follows: '[It] is conceived as an additional spatio-temporal continuity, a kind of autonomous intermediate reality, with which technologically, physically, mathematically, in terms of the history of coding and culturally and programmatically, an attempt is made to generate an independent material reality of cultural origin.' [52] Furthermore, he names two approaches from which the hybrid can be approached. In this basic biological definition, it is non-producible organic products consisting of two agents that are clearly distinguishable. Alternatively, it is to be understood as technology that is biologized and thereby fused into new viable systems—though they are not organismically intertwined with the human being with whom they are fused [52]. From these assessments, it remains apparent that, generally speaking, the hybrid remains a biological system (i.e., a human) that uses technology to enhance or sustain life. Thus, the human being remains in essence the agent and the responsible person in the sense of the law. This approach is supported by Beck, who describes the following series: human clones (in the sense of reproduction of one and the same body)—chimeras (in the sense of fusion of different biological cells)—and finally the hybrids (in the sense of fusion of a basic biological structure with technology) [3].

3.1.5.2 Psychology The term 'hybrid', in the context used in this review, encompasses beings which are comprised of technical and organic parts, like cyborgs. It is not commonly used in psychological literature, since hybrids are still inherently human, hence definitions of human beings apply to them.

3.1.6 Cyborg

3.1.6.1 Common ground In fact, it was rather the science fiction literature and film scene that coined the term cyborg; consider, for example, Arnold Schwarzenegger's incarnation of the Terminator in the film of the same name. In literature, therefore, cyborgs are described as bred and chemically transformed artificial humans [53]. This term has been adopted by so-called cyborg activists such as Neil Harbisson or Moon Ribas, both of whom have had a technical body part implemented (an antenna on the head and a sensor in the foot, respectively, which measure vibrations) to perform artistic performances by means of the vibrations. In reality, they are as unspectacular as a pacemaker is, although no one would think of calling a person a cyborg for that reason [54].

3.1.6.2 Legal science The term cyborg is closely related to the concept of hybrid and is not always strictly demarcated in the literature and is sometimes used synonymously. Cyborgs, after all, are also mixed creatures of human and machine, of which there is no precise definition—according to this, all persons with artificial hip joints or pacemakers would be cyborgs [3, 45]. However, the same could be said for the classification as hybrid, which is why a clarification is necessary here.

Cyborgization in medicine is used to describe the fact that humans use technology in all areas of their lives, whereby here the term is restricted to technology that is located under the skin [54]. Spreen describes this more precisely by saying that by means of cyborgization the digestive system is abolished [55]. The cyborg, still, has biological elements such as a skeleton, muscles, skin and a brain, which, however, consciously controls the previously involuntary functions of the body (as if through an implanted neurochip), because osmotic pumps are located at the crucial points of the organism, which, depending on the need, supply it with nutrients, activating substances or, conversely, with substances that lower the basal metabolic rate [55]. From a normative-legal point of view, it does not matter whether the brain waves are influenced from the outside by electrodes or from the inside by a neurochip—only from an ethical point of view it is relevant whether the appropriate information and risk assessment has taken place [54].[1]

In fact, a variety of definitions include people with technology that is external to the organism. The long-term aim of mechanization may also be the creation of indestructible body parts that allow adaptation to any environment (deep-sea, other planets, etc.) [56]. According to Faßler, these indestructible body parts can be: myoelectronic arms, synthetic bones, artificial hearts, breast and penis prostheses or artificial hips [57]. It is important in this definition that the machine parts function without consciousness, i.e., they must not disturb the vegetative nervous system [57]. In contrast to hybrids as a human being with machine components, a cyborg is therefore rather to be understood as a machine with human characteristics [57]. This leads to a considerable difference from a legal point of view: While an action of a hybrid, despite the fusion of human and technology, will ultimately be attributable to the human being (see below), this is not necessarily the case of humanized technology. Since these machine parts function independently without connection to the nervous system of the human being, progress or errors, which arise due to a machine action, are to be evaluated possibly differently than those of the hybrid.

[1] Just look at *Elon Musk's* project: a chip in the head that communicates with the smartphone and can bridge nerve damage.

3.1.6.3 Psychology Cyborgs are comprised of technical and organic parts (unlike robots or android, which do not possess any organic parts), like hybrids. But opposed to hybrid, cyborg is a commonly used term especially in medical- and neurotechnical debates [58]. Heilinger and Müller propose that cyborgs need to have a 'substantial' amount of organic human parts. They also pose the question of what 'substantial' means, without defining this further but pointing out that current definitions do not specify on this issue. By the most conservative definitions, even something as small and commonplace as glasses or an artificial hip joint makes humans, per definition, cyborgs. Heilinger and Müller consider this as self-cyborgization ('Selbst-Cyborgisierung'). They suggest a scale ('scala cyborgensis') [58], on which a continuum between a natural human ('homme naturel') to the textbook cyborg ('Bilderbuch-Cyborg') exists on which small aspects of cyborgization, for example glasses, prostheses, memory chips, define where certain individuals stand. For the sake of further discussions, from a psychological viewpoint, a distinction between therapeutic and enhancing features is reasonable [58]. Therapeutic measures would include implants, prostheses, and objects to restore originally existing properties, whereas enhancing measures would aim towards the redesign and improvement of human beings. By this, potential pitfalls of less desirable social attributions to cyborgs being applied to humans could be avoided. Where human users of bionic prostheses are perceived as competent and warm, cyborgs face a rather detrimental judgment. They are perceived as threatening, that is, cold but competent [59]. As for ethical considerations of hybrids and cyborgs, since the inherent humanity is not changed due to therapeutic or enhancing measures, the status as a full ethical agent still stands.

3.1.7 Android

3.1.7.1 Legal science The android as a collective term is given its own definition: Androids are humanoid robots that are almost indistinguishable from humans and in some cases even superior to them; they are therefore imagined to be personalizable and therefore able to satisfy human needs of all kinds [60]. Humanoid robots are those that are generally modelled on the human physique, i.e., they have two arms, two legs, a torso, a head and joints [27]. Anthropomorphic robots go even further, since they are literally of human shape—from a purely legal point of view, the exact difference to humanoid robots is not always clear [27]. Androids and gynoid robots are to be defined in the literal sense as male- or female-like robots [27]. It is therefore an 'artificial system defined with the aim of being indistinguishable from humans in its external appearance and behavior' [61].

What this classification actually means from a legal perspective is still unclear. This applies in particular since the

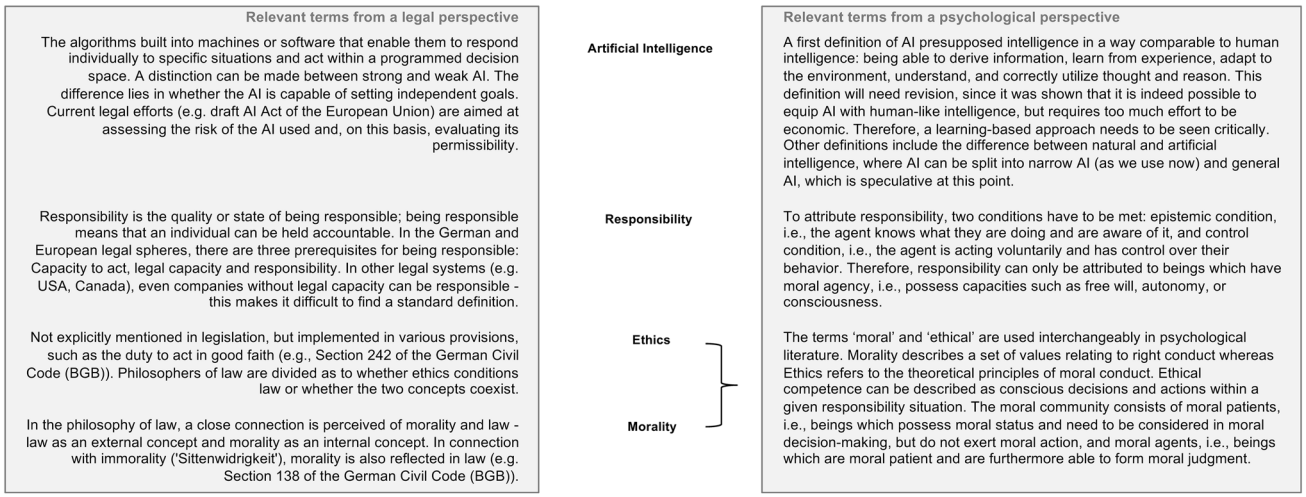

Fig. 3 Responsibility and the basis of rules

beings are attributed a consciousness and they are called intelligent, which is why the actions and decisions are autonomous [60]. With reference to the European Commission's 2017 Communication on 'Building a European Data Economy' [31], Wagner describes that robots or androids should be given a special legal status, which would mean granting them recognition as electronic persons [32]. This would make them liable for damages caused by autonomous behavior. However, the questions about an independent classification as a legal subject do not only arise when robots cause damage themselves, but also how to classify it when damage is committed to robots. Especially in the case of anthropomorphic robots, for example, the question arises as to whether mistreatment is relevant under criminal law, cf. Darling [61].

3.1.7.2 Psychology Where robots are supposed to be visually distinguishable from human beings, despite having superficial human-like features in some cases, androids aim at being as human-like as possible in terms of appearance. This extreme similarity could also have adverse effects on how the android robot is perceived by users: it could evoke fear, uneasiness, or even disgust. This effect is known as the Uncanny Valley effect [62]. Even though the existence of this effect is still under debate [61–66] to integrate robots into our everyday society, they may have to be virtually indistinguishable from human beings at least in some areas of application. From a psychological viewpoint, for androids as artificial beings, similar moral questions in terms of moral decision making and transparency, among others, apply.[2]

3.2 Meaning of AI

3.2.1 Artificial intelligence

3.2.1.1 Common ground Artificial intelligence is a term every scientist and researcher associates something with; at first glance there is broad consensus on the definition of the term, and yet differences can be found in wording that sometimes make a big practical difference (Fig. 3). AI as a term was first used in the proposal for the Dartmouth Summer Research Project on Artificial Intelligence [67]. The authors proposed that all aspects of learning and intelligence can be simulated by a machine. Seven aspects were mentioned, them being automatic computers, usage of language by a computer, neuron nets, theory of the size of a calculation, self-improvement, abstractions, and randomness and creativity. These aspects have endured decades and are still in the center of the development of AI. In 2007, McCarthy answered the question what AI is as follows: '(…) the science and engineering of making intelligent machines, especially intelligent computer programs.' [10]. From a computer science perspective, there are several waves of AI that include adaptive systems of varying degrees of sophistication [68].

[2] The controversial topic of sex robots replicating human beings should be excluded from this article, since the issues surrounding and influencing it would go beyond the scope of this review.

3.2.1.2 Legal science Artificial intelligence, generally speaking, is automated thinking, which, due to its automatism, should better be called machine intelligence [21]. Hildebrandt, with reference to Plessner [21, 69], clarifies that machine intelligence differs from human intelligence in that human intelligence is artificial; machine intelligence is only automatized. Artificial intelligence is ultimately exhibited by those machines that are capable of automated reasoning and, accordingly, have a specific type of agency that can be defined as data-driven or code-driven [21]. Humans, animals and machines may also have different degrees of intelligence, that is, the ability to think, which need to be considered in a definition. Russel and Norvig described that there are a total of eight conceivable definitions of AI, which can be divided into four categories: (1) thinking humanly, (2) acting humanly, (3) thinking rationally, and (4) acting rationally [70]. However, according to the general tenor, two groups of AI, weak AI and strong AI, are sufficient. In the Standardization Roadmap of the German Institute for Standardization (Deutsches Institut für Normung e.V.), a definition is described as follows: strong AI is a general intelligence that can set goals for itself; weak AI, on the other hand, is an AI system that has been developed for a specific purpose [71]. In short, artificial intelligence refers to machines that are capable of performing tasks that (when performed by a human) require a certain level of intelligence [9].

This can include both software and hardware components (i.e., a 'robot' or a computer program) [9]. In part, it is still emphasized that a sharp distinction should be made between the entity 'robot' and Artificial Intelligence, since robots or other interactive entities do not need to be constructed in a particular way [47]. Balkin, on the other hand, does not ascribe such importance to the distinction between robots as entities and artificial intelligence as 'algorithms of action', since this probably will not matter much in terms of how they are viewed by humans or the effects they bring, due to ongoing innovation in legal assessment [46]. Robots may cause people to see them as alive because they move; AI systems may be seen as alive because they talk [46].

Either way, to conclude in Scherer's words [9], the concept of AI should be regularly reviewed and adjusted as needed to reflect changes in the industry. The European Commission has recognized this need and on April 21, 2021, published its highly anticipated draft regulation to regulate the use of artificial intelligence ('AI') [72]. The draft regulation establishes harmonized rules for the development, marketing and use of AI systems in the European Union. It is an important step in the comprehensive European AI strategy. The draft takes a risk-based approach, dividing AI applications into four categories based on their potential risk: 'unacceptable risk,' 'high risk,' 'low risk,' and 'minimal risk.' The draft focuses on comprehensive regulation of those AI systems that pose a high risk according to this approach.[3]

3.2.1.3 Psychology It is noticeable that McCarthy presupposes the term of 'intelligence' in such a way that it is defined in a purely humane way, as psychologists might define it: 'the ability to derive information, learn from experience, adapt to the environment, understand, and correctly utilize thought and reason.' [10, 74]. A dual approach was proposed by Hoffmann and Hahn: (1) an algorithm has to satisfy at least the first two out of three conditions: (a) autonomous and complex decision-making with respect to some abstract goal, (b) the ability to learn and improve, (c) a cognitive representation of the self in an outside world. Furthermore, the authors propose that AI is '(…) any data analysis technology that enables decisions and reasoning about these decisions in a complex environment similar or superior to what humans can achieve.' [75]. But our current expectations of future use of AI may need to be revised, since researcher have shown that it is indeed possible to equip AI with human-like intelligence, for example, the ability to attribute false beliefs to others (Theory of Mind, ToM) [76], but it takes a lot of effort [77]. Both authors envisioned deep learning to be comparable to human learning mechanisms. In the case of Rabinowitz et al., the Sally-Anne-like test was passed by trained AI, but it required 32 million samples to perform at a level similar to that of a six-month-old infant. Therefore, Cuzzolin et al. cast doubt over the possibility of a pure learning-based approach for computational ToM. Ryan defined AI as being differentiated from natural intelligence and split between narrow AI, such as we are using now, and general AI (AGI), which, in his opinion, is still speculative [78].

3.3 Responsibility and basis of rules

After identifying and defining the actors involved in Hybrid Societies, terms associated with responsibility require closer attention. Therefore, the following section will again approach these terms, namely responsibility, morality, and ethics, from a legal and psychological perspective. For an overview see Fig. 2.

3.3.1 Responsibility

Responsibility can be distinguished in different types: the moral and the legal responsibility. Therefore, the following descriptions follow a different path.

[3] see for an application example of gaming platforms [73].

3.3.1.1 Legal science The concept of responsibility is one of the central legal issues when approaching Hybrid Societies. Hilgendorf explicated the concept of responsibility as follows: 'The responsible person X is responsible for the event Z according to rule Y' [6]. The rule Y is here that norm which the law or morality prescribes. It is important to notice that the legal subject can, but does not have to be a living organism. Therefore, a human being, as well as a company, can be a legal subject. In legal terms, the most important form of responsibility is that of responsibility under civil law—but there is also responsibility under criminal law or police law [6].

However, there are differences between these forms with regard to the persons who can be responsible. In German civil law, as in many other legal systems, not only people but also companies, for example, can be held responsible if contractual warranty claims are made. In criminal law, on the other hand, companies cannot be held criminally liable, as they lack the corresponding capacity to act. In many other countries,[4] though, this is possible—the requirements for criminal liability are thus not regulated in a consistent manner in all countries [79]. This fact complicates the task of finding a consistent definition of responsibility that can be adapted in all legal systems.

According to the basic German interpretation—and that of the European countries—three aspects are necessary for the attribution of responsibility: Capacity to act, legal capacity and capacity to take responsibility. This is the reason why concepts such as ethics and morality, but also agency and patience, have to be explained under the heading of 'basics' in this review. Only if a person is capable of being a bearer of rights and duties, they can manifest their will to the outside in such a way that they can cause legal effects. And only then can a person be legally responsible. Gruber describes the ability to be responsible as the ability to act according to the laws of the categorical imperative [12]—which brings us back to Kant [11]. Therefore, it is also important to know what is meant by the different entities that are listed—it is crucial to be able to classify their legal capabilities to sharpen the concept of responsibility. The following definition of 'responsibility' can be summarized in the legal sense: Responsibility is the quality or state of being responsible; being responsible means that one can be held accountable [80, 81].

As far as Hybrid Societies are concerned, it is the case that rarely an individual will be solely responsible for a matter. Thus, the achievement of successes and their imputability can often only result from the interaction of many participants and the accumulation of the various amounts (this is also accompanied by a problem of provability) [82]. If machines have their own scope for decision-making, this is of great importance with regard to the imputability of responsibility in the sense of criminal law, but also in the sense of civil law. Therefore, Asaro formulates the need for theories in which responsibility and agency are designed and aligned in such a way that large organizations of humans and machines produce desirable outcomes but can also be held accountable [83].

3.3.1.2 Psychology Moral decision making comes with distinct implications such as the question of *who is responsible* [84–88]*?* which have legal and moral effects. Moral responsibility can, from a psychological viewpoint, only be attributed to beings which have moral agency, i.e., possess different capacities such as autonomy, free will, or consciousness, among others [17, 37, 84, 87]. These attributions exclude by their nature machines and other inanimate objects. The case of AI is an ambiguous one: it encompasses questions of responsible development, responsible use, and responsibility in the narrow sense of *being* responsible for e.g., failure [85]. One is responsible only if they know what they're doing and actually doing it (or have done it). To attribute responsibility, two conditions have to be met: epistemic condition (i.e., the agent knows what they are doing and are aware of it) und control condition (i.e., the agent is acting voluntarily and has control over their behaviour) [89]. Since AI cannot fulfil these conditions as of now, it cannot be a responsible, neither an irresponsible moral agent, but rather an a-responsible agent [85]. Since moral decisions are and will continuously be made by AI, a gap in responsibility opens up and needs to be considered [85, 86, 88]. To bridge this gap, different options are available. For one, humans could be able to override AI if a decision could bring harm or seems faulty [85]. For several reasons, one of them being made vulnerable to human arbitrariness, this is not always possible and poses its own risks [90], even though if applicable, it certainly makes for a good option. Limiting automation is another option, such as opening only distinct streets for autonomous vehicles or limiting moral decision making to advice to the user—the second variant is a human-in-the-loop scenario where the final decision will be made by a human being, aided and informed by AI, but ultimately taking responsibility [85, 86]. Because of the sheer amount of people involved in the implementation and use of AI, and the amount of singular parts making up the final product, the dilemma of 'many things and many hands' needs to be taken into account when questioning responsibility and subsequently liability [84, 85].

[4] First and foremost the USA, Canada, New Zealand, England, Israel; Australia only until 1995.

3.3.2 Ethics and morality

3.3.2.1 Legal science *Ethics*: The concept of ethics is rarely mentioned in the pure application of law and in the study of law. Merely in the assessment of acting in good faith, the concept of ethics is found insofar as Sec. 242 of the German Civil Code (BGB)[5] in German civil law represents a fundamental and, above all, generally valid legal and social ethical principle [91]. A corresponding principle is also applied in Anglo-American common law countries, such as in the Uniform Commercial Code,[6] which exerts a uniform legal influence in the USA [92]. This application in legal practice illustrates where the principle of ethics actually assumes its greatest role: in philosophy of law. Hildebrandt has examined in detail the core statements of two philosophers of law [21]. Plessner and Radbruch, which are immensely relevant again, especially in the age of mechanical intelligence [4, 69]. As far as there is no written law and the coexistence of humans is determined by natural law, law is reduced to ethics [21]. This statement coincides with Radbruch's assertions, so that without legal certainty (i.e., through the certainty guaranteed by the legal system, i.e., written law) all action would be determined by the ethical inclinations of human beings or by those of the authorities [69]. It becomes clear that the respective contents of the legal norms depend on peoples' ethical conceptions, whereby only a minimum of ethical conceptions receives a legally binding written form. In contrast, for Kelsen ethics is the term for science, which is directed towards the knowledge and description of moral norms and social norms; jurisprudence and ethics coexist as norm sciences [93, 94]. The concept of medical ethics refers to human autonomy. According to Beauchamps and Childress [95], who follow Kant and Mill in their explanation, it is understood as the principle according to which everyone has to recognize that others may form their own opinion and act on the basis of personal values and ideas [54]. Ethical ideas often go beyond what is required by law [54].

In summary, the term ethics is still intensively discussed today, especially in relation to technologized entities. At all levels, the concept of ethics—referring to the Lexicon of Philosophy [96]—is defined as the branch of philosophy that deals with the prerequisites and evaluation of human action [71]. This can be seen for example in the Standardization Roadmap AI, published in 2020 by the *German Institute for Standardization* with the participation of leading figures from business, politics, science and civil society, the concept of ethics. Henceforth, the term is enshrined in law in a way that previously existed only in the field of medical law. With its enshrinement in the Standardization Roadmap AI, it is now incorporated into all application errors and acquires a legal effect.

Morality: The concept of morality is rather formative for jurisprudence from the perspective of legal philosophy than for the actual application of law. In the first decades of the twentieth century, several legal philosophers devoted themselves to the relationship between law and morality: In addition to Radbruch and Plessner, these include Kant, Mill, and Kelsen—with very different views. For Kelsen, the concept of morality is a conditional, positive type of norm that arises through habit and conscious laws. Due to this classification as a separate type of norm, he denies any connection between the two types of norms, so that laws with immoral content or laws that have emerged immorally nevertheless claim validity [93]. Radbruch understands law and morality as two cultural concepts that differ in that law is external, whereas morality is internal. Because of these characteristics, he assumes an inseparable connection between law and morality, so that any law that contradicts morality cannot have any validity [94]. Conceptually, these two views can be divided into the Connection Thesis and the Separation Thesis [94].

As with the concept of ethics, the concept of morality is also applied in the interpretation of certain circumstances: in German civil law, for example, in Section 138 of the German Civil Code (BGB)—immorality.[7] A contract that is contrary to morality is unlawful. The commentary literature specifies that 'good morals' does not mean morality in the sense of ethics, but rather moral views recognized in the community [91]. Once again, then, it becomes clear that morality is something that is rooted in the nature of human beings, for—according to Neuhäuser [97]—human beings possess moral qualities: Perception, language, higher-level intentionality, normative competence, and moral judgment; therefore, they can take moral standpoints. The human species is thus also capable of moral judgment [97], i.e., it has the ability to distinguish morally relevant from morally irrelevant points of view in a concrete situation and can therefore judge which specific concerns are to be weighed against each other.

What has been said in the last two sections, the classification of ethics and morality from the perspective of legal philosophy, probably leads to the general conclusion that ethics as a concept tends to address the prerequisites of the social thinking of human beings, while the concept of morality rather describes the ability, e.g., to assume responsibility. Both characteristics, and here is the fundamental problem of

[5] Comparable concepts can also be found in Art. 1337 of the Italian Codice Civile, in Art. 1134 para. 3 of the French Code Civil or in Art. 1258 of the Spanish Código Civil.

[6] 'Every contract or duty within the Uniform Commercial Code imposes an obligation of good faith in its performance and enforcement.' pre UCC Sec. 1–304.

[7] Whereby the German legislator does not use the German term 'immoral', but that of 'Sittenwidrigkeit'.

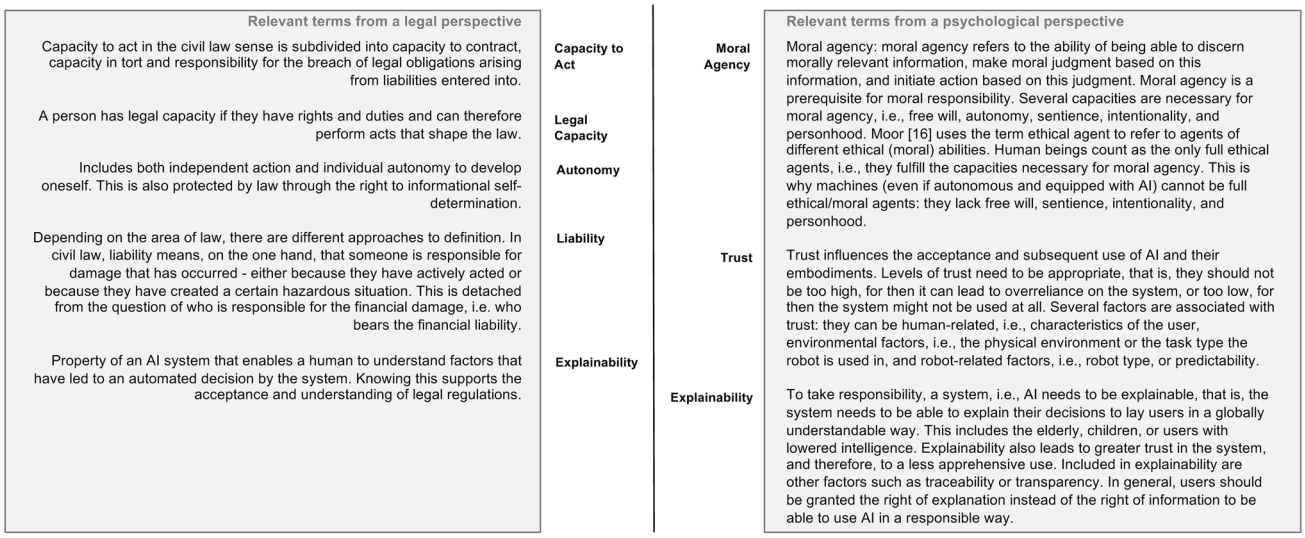

Fig. 4 Capacities necessary for responsibility

the current state of science, would have to be possessed by those entities with which humans act in Hybrid Societies.

3.3.2.2 Psychology Since the terms 'moral' and 'ethical' are regularly used interchangeably in psychological literature, this will be done in this part of the text as well. Morality, 'a system of beliefs or set of values relating to right conduct, against which behavior is judged to be acceptable or unacceptable' [98], and Ethics, the theoretical principles of moral conduct [99], have been subject to philosophical debate since antiquity. Where philosophy is concerned with ethical theories, moral psychology focuses on human thought and behaviour in ethical contexts, that is, how humans follow ethical theories [100]. Ethical theories which could apply to robots and pros and cons thereof will be revisited later.

To derive the relevant moral or ethical competencies, it is useful to consider well-established models of moral development and behaviour: one of the most prominent models is based on Kohlberg's classical hierarchical model of six moral stages [101], they proposed a sequential four-component model consisting of: (1) moral sensitivity to recognize an existing moral problem, (2) moral judgment as reasoning about morally correct actions, (3) moral motivation to establish moral intent, and (4) to engage in and persevere with moral actions [102, 103]. Hannah et al. extended this model by describing processes and associated capacities that are necessary within these stages [104]. Moral sensitivity and judgment correspond to so-called moral cognition processes, that is, the awareness and the processing of moral issues. Moral conation processes, on the other hand, refer to moral motivation and behavior. There are individual capacities that influence these processes: Moral maturation capacities enable elaborated storage, retrieval, processing, and integration of moral information and help to develop more complex cognitions or models with regard to the logic of morality [104]. Moral conation capacities, in turn, underlie enacting morally motivated action. They enable a person to feel responsible and to be motivated to take moral action [104]. Regarding the moral maturation capacities, Hannah et al. suggest three constructs to be critical in driving moral cognition processes, namely moral complexity, meta-cognitive ability, and moral identity, while moral ownership, moral efficacy, and moral courage are relevant moral conation capacities [104, 105]. Regarding moral action of humans, there are only recently considerations about what constitutes ethical competence [106, 107]. Ethical competence can be defined as 'conscious decisions and actions within a given responsibility situation. It implies to feel obliged to one's own moral principles and to act responsibly taking into account legal standards as well as economical, ecological, and social consequences. Ethical competence requires normative knowledge and the willingness to defend derived behavioural options against occurring resistance' [106, 107]. The moral community includes two entities: moral patients, that is, beings which possess some moral status and need to be considered in moral decision-making since they are capable of suffering moral harms and experience moral benefits, but they themselves do not exert moral action, and moral agents, that is, beings which are moral patients but furthermore are able to form moral judgment [37, 75, 87, 108, 109].

Table 2 Five major capacities necessary for legal responsibility

Capacity to act	Capacity to contract Capacity in tort Responsibility for the breach of legal obligations arising from liabilities entered into	Abbott [7] Calo [47] Christaller [43] EU Commission [31] Hilgendorf [6] Kirn and Müller-Hengstenberg [110] Klement [111]
Legal capacity	Legal obligations and rights	Asaro [83] Beck [60, 82] Darling [112] Gaede [113] Gellers [114] Gruber [12] Hildebrandt [21] Maia Alexandre [115] Matthias [13] Schirmer [116]
Autonomy		Reed [117] Sheriff [118]
Liability	Liability for an act Financial liability	Balkin [46] Chinen [81] Haagen [119] Wagner [32]
Explainability		Pasquale [120] Beck [28]

4 Capacities necessary for responsibility

Over the course of literature research, a handful of capacities came up more often than others. We see them as pivotal for responsibility but are at the same time aware that this list is not exhaustive. In the next section, we will elaborate on capacities necessary for legal and moral responsibility in Hybrid Societies. For an overview see Fig. 4.

4.1 Legal science

As mentioned above, there are some key characteristics that an actor has to possess to be held responsible in a Hybrid Society (Table 2). In addition to the capacity to act and the legal capacity, these include, in particular, autonomy and explainability. This is completed by the capacity for liability. In the following, these concepts are further divided and related to existing literature that has dealt with capacities in depth.

4.1.1 Capacity to act

If the legal side of agency is considered, this presupposes that legal capacity must also exist on the entity's side. This is also followed by Chopra and White, according to whom an agent must be a person, who in turn can be (1) an individual, (2) an organization, (3) a government, political subdivision, or otherwise an institution or entity created by the government; but (4) also any other entity that has the capacity to enter into rights and obligations [121]. In the legal context, the concept of capacity to act describes the ability to manifest one's will externally, in other words, to produce legal effects through one's own action [91]. It must be strictly distinguished from legal capacity (i.e. the ability to be the bearer of rights and obligations—a legal subject [122]). Capacity to act in the civil law sense is subdivided into capacity to contract, capacity in tort and responsibility for the breach of legal obligations arising from liabilities entered into [122]. In law, everyone is considered to have the capacity to contract unless exceptions are made. The capacity to contract is largely age-dependent, in Germany, for example, under Section 104 of the German Civil Code, anyone who has not yet reached the age of seven is legally incompetent. In addition, pathological disorders also have an influence on a person's capacity to contract. Persons who are legally capable but incapable of acting are represented by authorized representatives (parents or organs). Capacity in tort refers to a person's ability to be responsible in tort for harm they cause to another and to be liable to pay damages.[8] A restriction of the capacity to commit a tort is also made in particular in the case of minors, in this case it is a matter of the required sanity. Apart from a few exceptional cases (when equity so requires), persons who lack the capacity to commit a tort are not responsible for the damage they cause. The third category of capacity—liability for breach

[8] See, for example, Sect. 823 of the German Civil Code (BGB).

of legal obligations—concerns fault. In principle, liability is imposed only for one's own fault; however, the legal system recognizes exceptions, such as strict liability in road traffic or product liability.

4.1.2 Legal capacity

A person has legal capacity if they have rights and duties and can therefore perform acts that shape the law [15]. By its very nature, legal capacity is attributed only to human beings, since they are the sender and addressee of the commandments of the legal order. According to the concept of law, only they are capable of understanding the meaning of the commandments and acting in accordance with them [123]. In addition, associations of persons or legal entities may also be affected by legal obligations and rights. Who has legal capacity and is thus a legal subject is determined by the respective legal system.

4.1.3 Autonomy

To classify whether an entity can make decisions independently, it is necessary to know first what is meant by autonomy. Kant describes, on the one hand, autonomous action and, on the other, the individual autonomy of human beings [54]. The former means action according to self-chosen purposes; he defines the individual autonomy of the individual as personal self-determination. This idea was echoed by Frankfurt and Dworkin, who also focused on the individual's self-determination, describing autonomy as implying that the individual grasps motivations with which they can identify and which significantly guide self-determination [124]. However, autonomy also has limits: protection from self-harm, preservation of identity, authenticity, and the essence of being human, and interference with other common goods such as equality of opportunity [54].

In jurisprudence, the autonomy of the human being is decisively reflected in the entire legal system. On the constitutional level as a whole, the focus is always on the human being and his or her fully autonomous actions. In addition to the human rights standardized in the European Charter of Fundamental Rights (see in particular Art. 6 CFR), this is also reflected at the national level. In the German Basic Law (GG), this is reflected in the general right of personality of Art. 2 (1) in conjunction with Art. 1 (1) GG, which protects the free development of the personality. However, this can also only be guaranteed to the extent that the general public is not disadvantaged [125]. Protection against self-harm is also enshrined in law: in Sect. 228 of the German Criminal Code (StGB), bodily harm cannot be justified by consent if it is contrary to public morals.[9]

Legal considerations reach their limits when it is noticed that the autonomy of machines and subsequent entities is increasing immensely: the scope for decision-making granted to them is increasingly broad and, accordingly, unintended actions can be attributed to them [3]. The autonomy granted to machines, though, is difficult to compare with the autonomy of humans, which has just been described. In this context, autonomy should rather be understood as independence, i.e., the ability of a machine etc. to perform tasks on the basis of its internal state and environment without the influence of humans [71].

4.1.4 Liability

Liability is a concept that has been primarily coined by legal science. However, it can be described in different ways in relation to the various areas of law. In civil law, for example, a distinction can be made between liability for the act or damage and financial liability, i.e. compensation, which usually takes the form of money. Due to the immense scope and the specifics of the respective practical applications in which liability is relevant, here the term shall be limited to the liability of human beings or EDTs in circumstances that occur in connection with autonomous beings.

Even in this context, the concept of liability is subject to immense change, considering the increasing technical, but also moral-ethical changes and advances; after all, the welfare of human beings is given a higher priority [7]. Thus, according to Balkin, strict liability could be a suitable approach in its tradition, but it stifles innovation in its bud for fear of financial consequences and can never be a suitable solution—for example in the context of criminal law [46]. It is therefore necessary to apply the different concepts of liability appropriately with regard to the respective damage situation.

Liability means, on the one hand, that someone is responsible for damage that has occurred—either because he has acted actively or because he has created a certain hazardous situation. These two aspects are divided into the concept of fault liability and strict liability [122]. Most injuries caused by people under fault liability are assessed according to an intentional or negligent standard [7], which means that liability is established if the conduct is unreasonable. This—according to the current state of the law—cannot also be applied to any form of computer, since a more stringent standard of liability applies there if they cause the same injuries [7]. Ultimately, a stricter liability is inherent in machines

[9] At this point, again, a circle can be drawn with regard to moral and ethical action.

Table 3 Three major capacities necessary for moral responsibility

Moral Agency	Allen et al. [127] Bandura [138] Bigman and Gray [36] Cave et al. [18] Floridi and Sanders [87] Gray et al. [36] Hakli and Mäkelä [17] Johansson [109] Mabaso [128] Misselhorn [129] Tigard [130]
Trust	Hancock et al. [131] Ivanov et al. [132] Naneva et al., [51] de Visser et al. [133]
Explainability	Bostrom and Yudkowsky [134] Cave et al. [18] Coeckelbergh [85] Floridi et al. [135] Miller [136] Vanderelst and Willems [137]

than in human beings. This is detached from the question of who is responsible for the financial damage, i.e., who bears the financial liability. This legal system attempts, as far as possible, to link the actions of autonomous machines and the consequences to humans [81], whether by means of individual liability based on use, product liability, agency, aiding and abetting aspects or command responsibility. Product liability refers to the responsibility for the commercial distribution of a product that causes harm because it is defective or its characteristics have been misrepresented [7]. Products law in the U.S. combines tort law, contract law, and commercial and statutory law. In this sense, the law is consistent with prevailing views on moral responsibility [81].[10]

As the interaction between humans and machines in public spaces increases, this principle may shift to a strict liability approach [32]. This corresponds to strict liability, as is already the case, for example, in road traffic law (Section 7 of the German Road Traffic Act (StVG)). This liability rule would only require proof of the damage, the harmful functioning of a robot and a causal connection [32]. As an alternative to this strict liability with regard to robots, the European Parliament [126] proposes the so-called risk management approach, which does not focus on the entity that acted, but on the entity that was in a position to avert the risk, but did not act. This approach is—according to *Wagner*—rather unsophisticated [32].

[10] Here again, the close link between morality and law can be seen.

4.1.5 Explainability

Regarding the topic of explicability, there is rather less to say in terms of purely legal usage, because it has hardly gained any influence in the language of the legal profession. Nevertheless, this term is an immensely important one when we think about the implementation of AI and robotic entities in our society [75]. The term can be defined as follows: The property of an AI system that led to an automated decision of the system can be understood by a human [71]. This understanding is what also promotes the recognition of legal concepts. By understanding why a system acts the way it does, it is easier to understand why legal conditions are created in a certain way, so that this promotes acceptance and trust in these systems.

4.2 Psychology

Several preconditions have to be met by agents of any kind to be considered responsible. Even though more factors can surely be found, we limit the necessary preconditions to those regularly referred to by researchers and which we found non-negotiable: moral agency, trust, and explainability (Table 3). In the following, we will evaluate them in more detail.

4.2.1 Moral agency

The broad term 'agency' includes different capacities such as thought, communication, planning, recognition, emotion, memory, morality, and self-control. These capacities define to which extend a character is capable of, for example, morality. Therefore, moral responsibility is tied to attributions of agency [37]. The moral community consists of two entities which need consideration: moral agents and moral patients (or receivers). Moral patients are agents whose well-being is morally relevant but who cannot be held morally responsible, that is, animals or babies—or, under certain circumstances, robots [37, 75, 87, 108, 109]. Moral agents are able to discern morally relevant information, make moral judgments based on this information, and initiate action based on the judgment—their actions can be morally right or wrong [37, 87, 109]. Subsequently, they have rights and responsibilities in a moral community [129]. Therefore, moral agency and moral responsibility are mutually dependent [84, 88].

As for capacities necessary for moral agency, consensus has been found for some, where others are more critically evaluated. Hakli and Mäkelä proposed six classes of capacities which are necessary for moral agency: sociality and personhood, normative understanding, autonomy, sentience, rationality and action, and intentionality [17]. There is a broad consensus on these capacities: sociality

and personhood includes e.g., the ability to communicate, to form a judgment, and to process information as well as having memory, social commitment, memory, morality, and thought among others [17, 18, 21, 36, 37, 87, 109, 124, 129, 130]. Normative understanding, that is, the awareness of responsibility or moral reasoning, was also included by several researchers [17, 37, 122, 129, 138]. Autonomy includes e.g., the abilities to critically reflect values, setting goals, and exerting self-control [17, 36, 37, 87, 127, 129, 138]. Sentience is tied to having consciousness, emotions, empathy, and self-awareness [17, 37, 129, 138]. Rationality and action subsumes reasoning, action and omission, and decision-making and planning [17, 18, 36, 37, 109, 122, 129]. Intentionality, i.e., the ability of believing, desiring, having higher-order internal states, has been proposed as a precondition for agency by several researchers [17, 18, 129, 138]. By looking at these classes of capacities, it becomes clear why the notion of ascribing moral agency to robots seems to be premature at least [16, 20, 25, 33, 85, 88, 109].

Moor's taxonomy [16] is often employed when it comes to defining what qualifies as which agent [18, 139–141]. The taxonomy includes four stages of ethical agents: ethical impact agents which only have ethical impact via the task they are programmed to do but not by acting itself. Implicit ethical agents, in contrast to explicit ethical agents, do not have any ethics explicitly added in their programming. Full ethical agents are able to make and explain explicit judgments since they have consciousness, free will, and intentionality. This definition makes in clear why machines, even if we describe them as autonomous, cannot be full ethical agents as of now [20].

Despite robots probably never being full ethical agents, they—or the AI behind them—might be considered as explicit ethical agents. That is to say, they will make decisions that follow a certain set of moral rules. First apprehensive steps are taken with the introduction of Autonomous Vehicles or with AI being used in employment to process resumes [142]. But when it comes to the question which set of moral rules should be used, opinions and ideas differ greatly. The worldwide Moral Machine Experiment [143], which included 40 million decisions made by millions of people from 233 countries and territories, has identified three major clusters. These clusters were able to agree on a basic set of ethical principles in answer to one very specific moral dilemma used for autonomous vehicles. Despite this experiment being fascinating for the sheer number of decisions, it only provides insight into one single moral choice. This shows that the agreement on which ethical principles to follow is not trivial. Tolmeijer et al. identified seven ethical theories of importance in terms of the taxonomy of moral machines: consequentialism (act and rule utilitarianism), deontological ethics (agent and patient-centered), virtue ethics, particularist view, hybrid theories (hierarchically specific or nonspecific), configurable ethics, and ambiguous theories [139]. Artificial moral agent (AMA) is a term which frequently is used in the discourse of robot ethics. Even though there is critique to robots being seen as AMAs, such as the lack of an organic brain which is crucial for experiencing emotions and empathy, a lack of syntactic and semantic understanding, and the absence of mental states and subsequent intention to act [109], research on AMAs has flourished. AMAs can be modelled on various ethical theories, which poses the problem of deciding which moral principles should be implemented [127]. Bauer proposed AMAs which follow two level utilitarianism, i.e., they follow those rules that would tend to maximize good. Because of the stringent following of rules, this would make it ethically better than typical human behaviours [19]. On a first level, the AMA would instinctively follow a set of rules that has been shown through experience to result in the most good. On a second level, if these rules conflict or no rule exists, the agent is directed to calculate utility as act utilitarianism requires. Utilitarian choices seem preferable on a metalevel, since when questioned, users would not use agents equipped with utilitarian ethical theories [144]. This might be that, given the moral dilemma situation, an AMA equipped with utilitarian theories might sacrifice the user to save bystanders and users would prefer a self-protective model for themselves [145]. Furthermore, utilitarian choices were, while expected from and permissible for robots, considered morally as wrong, even though less so for robots than for humans [146]. Recently, the implementation of virtue ethics in AMAs has been considered as an alternative. It has been shown to be a promising moral theory for understanding and interpreting the development and behavior of AMAs [147].

Instead of focusing on one singular theory, a pluralistic approach based on several key concepts, which would avoid a theoretical bias, might be a practicable way, even more so if the paramount aspect is to avoid the immoral [139]. The importance of considering different ethical frameworks has to be broadened, as already suggested by Awad et al. [143], by taking cultural differences of the users into account: following moral advice from a robot is strongly influenced by cultural orientation, e.g., that individualistic cultural background interferes with making honest choices, especially if the moral framework was drawn from virtue ethics [148]. A more recent approach proposed a hybrid relational-normative model of moral cognition: by implementing a role-oriented moral core as a system of norms, a cornerstone is laid and expanded by moral cognition and moral decision-making. The robot can therefore evaluate different actions for their own decision-making [149].

4.2.2 Trust

A factor which influences the acceptance and subsequent use of AI and their embodiment, that is, robots, greatly, is trust in robots. Inappropriate levels of trust can culminate in two extremes: if very high trust is placed in the system, this can lead to an overreliance on and misuse of the system. If very low trust is placed in the system, it might be disused entirely [131]. Three different factors of trust development were identified in HRI: human-related, robot-related, and environmental factors [131]. Human-related factors encompass characteristics of the user such as demographics, personality traits, attitude towards and comfort with robots, self-confidence, and propensity to trust [51, 131, 132]. Furthermore, attentional capacity, expertise, and competency among others need to be taken into account. Environmental factors such as in-group membership, communication, task type and complexity, and physical environment play an important role in the trust development when humans are working or sharing a space with robots. As for robot-related factors, these can be either performance-based or attribute-based, where attribute-based includes robot personality and type, level of anthropomorphism, proximity or co-location, and adaptability. In terms of performance-based factors, dependability, reliability, predictability, and transparency play a pivotal role. Failure rates and false alarms, as well as the level of automation and behaviour of the robot are additional performance-based factors [131]. De Visser et al. identified three stages of trust when it comes to humans sharing workspheres with robots [133]: trust formation, trust violation, and trust repair. In the first stage, trust formation, people react rather deferent to automation in such a way that they exhibit a positive bias and place greater trust in the robot. In the second stage, trust violation, De Visser et al. summarize that those effects associated with trust violation from robots and from fellow humans may be differ qualitatively, since human beings are seen as fallible, where artificial agents are perceived as perfect [133]. Trust repair was shown to be associated with anthropomorphism. Adding human features in terms of appearance or social ability increased trust resilience, that is, a higher resistance to breakdowns in trust. Furthermore, more machine-like agents were found to be perceived as initially more trustworthy, but were not able to regain trust as easily as more anthropomorphic agents did [133].

4.2.3 Explainability

As mentioned earlier, Floridi et al. proposed seven ethical factors for responsible AI [135]: (1) falsifiability and incremental deployment, (2) safeguards against the manipulation of predictors, (3) receiver-contextualized intervention, (4) receiver-contextualized explanation and transparent purpose, (5) privacy protection and data subject consent, (6) situational fairness, and (7) human-friendly semanticization. Factor (4), receiver-contextualized explanation and transparent purpose, has gained special attention [18, 134, 136, 137]. Explainable Artificial Intelligence, i.e., AI which is able to explain their decisions to users in an understandable way [147], is one of the promising fields. Currently, the lack of trust in AI can lead to an apprehensive use. By increasing transparency, increased trust is being achieved. Decisions have to be explicitly explained and understood by all users—including lay people [136]. Traceability was suggested as a way to operationalize responsibility and explainability [85]: users usually are aware of the intended consequences (i.e., the goal) but not necessarily of the non-intended consequences and moral significance thereof. So, users should be granted the right of explanation instead of the right of information. This should be based on specifically asking affected users what and how much explanation is needed and making sure that the explanation is understandable for all affected parties, i.e., older people, people with disabilities, children.

5 Discussion

This review aimed at finding a common basis between two disciplines that are of immense relevance in shaping future Hybrid Societies: legal science and psychology. First, we identified and defined actors in Hybrid Societies. Second, we discussed constructs of and related to AI. Third, we identified the capacities necessary for legal and moral responsibility as core competencies for actors in societies. Our next step is to integrate these findings into a preliminary model of responsibility in Hybrid Societies. The concepts discussed here provide the basis for being able to sufficiently sharpen the topic of responsibility in Hybrid Societies from a legal and psychological perspective. The spectrum of necessary terms can be continuously expanded to develop a model of responsibility that can be applied to all areas of the technologized society. For example, concepts such as digitalization, obligations, foreseeability or error, as well as fairness, transparency and causality could be discussed here—but this would unduly exhaust the scope of this review.

5.1 Actors

By contrasting the different entities involved in Hybrid Societies, it became obvious that both legal science and psychology agree that, for now, whenever a human being is involved, they are at the focal point. Human beings are the only entities of Hybrid Societies which are considered (a) legal subjects and (b) full moral agents, making them the only actors who can be fully responsible, in contrast to artificial entities such as robots. Both disciplines have similar

definitions of robots. For robots, a smaller-scale grading of their application areas, for example industrial or social settings, is necessary. Robots, same as machines, are both not capable of responsibility as of now due to their inability of making fully autonomous decisions. Legal science as well as psychology agree that machines and robots are mere tools, which can be either based on algorithms or have simple supportive value for human beings. However, there are also noticeable differences between both disciplines: whereas psychology does not necessarily make a difference in terms of responsibility between human beings, cyborgs, or hybrids—they are inherently human, therefore the same conditions apply, legal science discerns between the levels of technology incorporated. As can be derived from the evolution of artificial beings, questions of responsibility are linked to autonomy, which presupposes the (theoretical) ability of decision-making. This applies to both legal science and psychology, but with different areas of application: from a legal point of view, even a non-viable entity, provided it has only a certain degree of autonomy, can bear co-responsibility in a legal sense. From a psychological viewpoint, autonomy is a precursor of (moral) agency and refers to the ability of making self-reliant decisions based on a predefined set of rules.

Technical implementations of decision-making along these rules, and, if applicable, straying from this set of rules, is extraordinarily difficult. This would require complex decision-making algorithms. Dignum et al. proposed an approach via deontic logics, implemented as a mixture of top-down explicit design and bottom-up derivation [84], e.g., based on reinforcement learning. These hybrid approaches for implementing moral decision-making into AMAs have been proposed by Allen et al. as a way to account for problems other approaches of top-down and bottom-up decision-making might face. Top-down would require an ethical framework which is implemented [127], e.g., prima facie duties, deontology, or divine command ethics. Bottom-up approaches are dependent on the data used for training purposes, for example associative learning machines, or using the virtue ethics framework [139].

Talking about Hybrid Societies presupposes the interaction of humans and EDTs are taking place on a comparable level. This means that ideas of society and legislation, which are ascribed to human beings, need to be applied to EDTs as well. While this is currently not the case [150], scholars are anticipating different future scenarios where EDTs might be capable of agency [20, 35, 47, 84, 151]. Some researchers go even further and expect robots to be able to self-reproduce which lends a whole new meaning to the term 'Hybrid Societies' [152]. If this happens, this poses problems insofar as that the system of rights and duties, which applies to human beings, will need to be extended to EDTs. We argue that, while granting rights to EDTs might be possible under certain conditions [112], imposing duties on EDTs is impossible due to practical capacities, e.g., financial property [32]. This follows the same line of argumentation as animal rights: animals are granted certain rights, but no duties are imposed [153]. Therefore, full agency in the sense of agency of human beings is not possible for EDTs. Imagine wanting to switch off for example a robot which has comparable rights to human beings or animals, this would not be allowed according to criminal law, since it would be seen as assault [112, 154].

5.2 Responsibility

Before we define capacities necessary for responsibility, we need to take a closer look at the concept of agency. There are divergent characteristics in psychology and legal science. In psychology, (moral) agency refers to capacities such as communication, thought, memory, or intentionality, which are necessary preconditions for deliberate actions. In legal science, agency refers to the ability to cause legal effects by virtue of one's own action. As can be seen, agency in a psychological sense encompasses a larger construct which includes many different aspects, whereas the definition in a legal sense is more straightforward and less extensive. One factor closely related to taking responsibility is the Level of Autonomy (LoA; e.g. [87]), on which an actor operates. The more autonomous, the more responsibility can be assumed. Therefore, responsibility cannot be assumed from EDTs of a basic LoA. Capacities necessary for responsibility are currently only applicable for human beings, regardless of whether legal or psychological capacities are being referred to. Capacities which were regularly deemed necessary for responsibility are included in the following simplified model (Fig. 5).

From a legal point of view, an entity who acts responsibly necessarily has to possess fundamental abilities: The ability to be the bearer of rights and obligations (legal capacity), and the ability to act consciously (capacity to act) and thus, to act in legal transactions. Only in this way it can be recognized as a legal subject acting in its own authority and be regarded as the agent of action or omission. This is closely related to the respective level of autonomy—the more autonomous, independent or self-reliant a legal subject can act, the more likely it is to accept the duty to assume responsibility. Finally, a possible consequence is the assumption not only of theoretical but also of practical responsibility—in the sense of liability—which can be assumed only by those who are capable of bearing this duty. To affirm the causality of responsibility, which in law connects fact and legal consequence, the addressee of a norm need to be able to explain the technology.

From a psychological perspective, two main foci can be identified: those referring to capacities within users (trust), and those of capacities within the responsible artificial agent

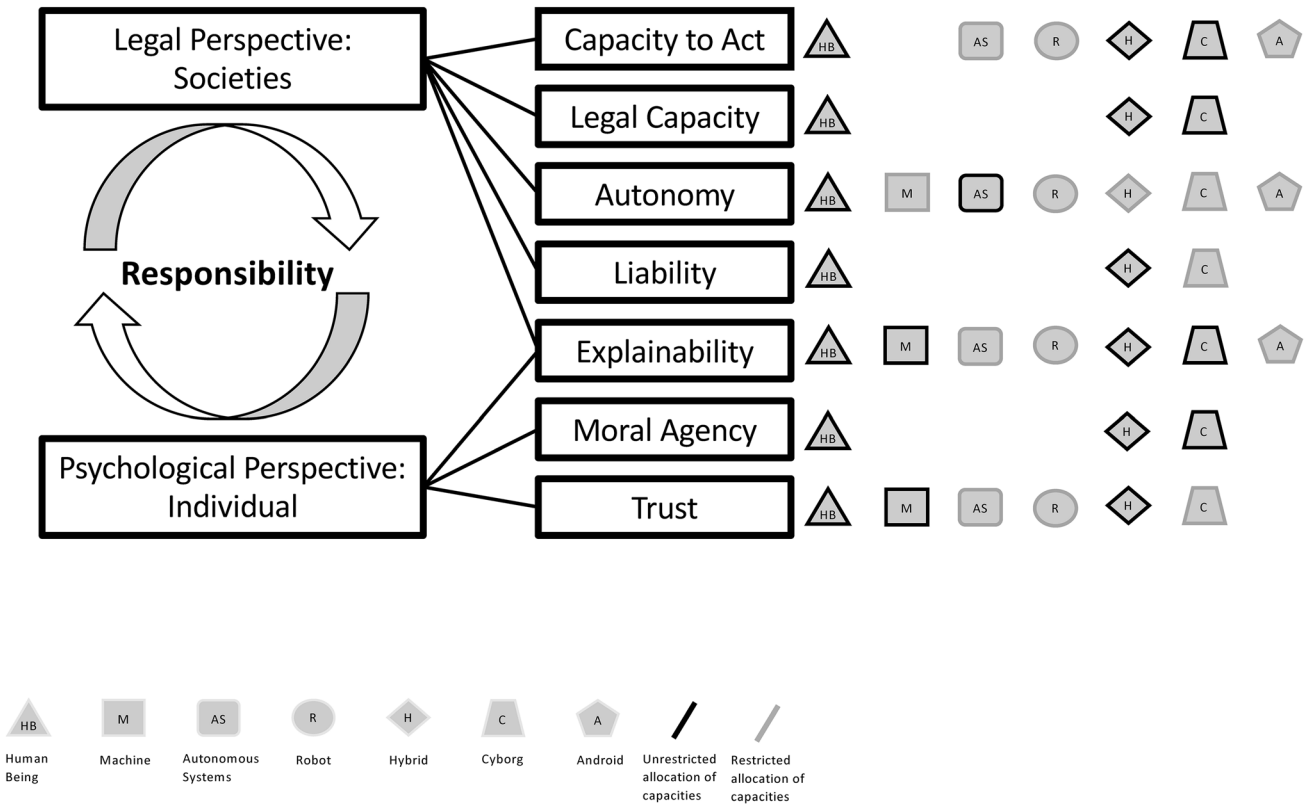

Fig. 5 Preliminary model for responsibility from a legal and psychological perspective

(explainability and moral agency). Trust and explainability refer to how responsible artificial agents need to relate to human beings. First, humans need to trust technology they use, otherwise, either overreliance or disuse might be the undesired results. Second, decisions made by responsible artificial agents need to be understandable, that is, explainable for users, regardless of who the user is, including elderly or children. A third capacity, moral agency, is required for responsibility. Moral agency includes various capacities, such as autonomy, intentionality, or sentience, among others, as was illustrated earlier.

From a legal as well as psychological viewpoint, vulnerable groups such as the elderly or children, require explicit attention. Therefore, as far as the design of responsible artificial agents is concerned, special focus has to be laid on explainability and privacy issues: details need to be explained in a way which is equally understandable for laypeople and experts.

Even though, by defining capacities necessary for responsibility, we assume that EDTs could be members of Hybrid Societies in the future, we need to point out some limitations: For one, the willingness of the users to hand over decision-making, and therefore responsibility for an action, to EDTs needs to be taken into account. For example, concerning AI, people are losing trust in these technologies due to media coverage such as the headline 'The foundations of AI are riddled with errors' [155]. Of course, trials such as Microsoft's Tay, a chatbot which was 'released into the wild' of Twitter and 'learned' racism, sexism, and antisemitism with lightning speed due to the information presented to them by human users, do not enhance trust in AI [88]. Aforementioned headline alluded to the problem of faulty recognitions of pictures by AI due to biased training data [35, 156, 157]. These two examples show that algorithms can only be as good as the data they are trained on. To try and counter these negative examples, the idea of Artificial Intelligence for Social Good (AI4SG) is being followed by different initiatives [135]. They proposed principles and factors to be considered when working on and with AI [135]. The aim of these factors is to ensure that the design, development, and deployment of AI systems happens in such a way, that they prevent, mitigate, or resolve problems which might harm human life and/or the natural world and/or enable socially preferable and/or environmentally sustainable developments [135]. Five ethical principles were proposed, four of them the traditional bioethics principles (i.e., justice, autonomy, beneficience, and non-maleficience) [95] and a new enabling principle for AI, explicability. In 2020, these principles were broadened by seven ethical factors: (1) falsifiability and incremental deployment, (2) safeguards against

the manipulation of predictors, (3) receiver-contextualized intervention, (4) receiver-contextualized explanation and transparent purpose, (5) privacy protection and data subject consent, (6) situational fairness, and (7) human-friendly semanticization [135].

To ensure that developments in AI are for societal good, three aspects were identified: accountability, responsibility, and transparency (ART; [86]). To account for these aspects, three control systems were proposed: human-in-the-loop, regulated environment, and the AI system as the ethical agent AMA, that is, ethics by design. Apart from the willingness to use EDTs, cultural backgrounds of users could play a role in the acceptance of EDTs [143]. For example, different cultures prefer different ethical decisions and eventual outcomes (e.g., [158]). Therefore, one global framework is not feasible.

On a European level, the legislator tries to solve the problem of the lack of responsibility by regulating just a part of the product, e.g., the implemented AI. Different from previous approach of regulating product responsibility (see New Legislative Framework[11]) [159–161], the AI Act [70] regulates the different levels of risk of AI itself. It therefore regulates just one part of an EDT, for example, instead of regulation the responsibility of the actor.

6 Conclusion

We identified seven capacities in total which need to be met by actors in societies for responsibility. From a legal perspective, responsible actors need to be able to be the bearer of rights and obligations (legal capacity) and to act consciously (capacity to act). This includes the capacity to contract, capacity in tort, and responsibility for the breach of legal obligations arising from liabilities entered into. Furthermore, they need to have a sufficient level of autonomy and be capable of bearing the duty of liability, including liability for an act and financial liability. From a psychological perspective, two foci of capacities necessary for responsibility can be identified: those within the user (trust) and those within the agent (moral agency and explainability). Moral agency is the ability of discerning morally relevant information, making moral judgments based on this information, and acting upon it. It includes other important aspect, such as, for example, autonomy and sentience. Trust is influential on the acceptance and subsequent use of artificial agents, since inappropriate levels could lead to misuse, or disuse. Explainability refers to the ability of a system to explain their decisions to lay users in a globally understandable way. It includes factors such as traceability and transparency. Explainability is also immensely important for the understanding of the creation of legal concepts for artificial agents.

As of now, it is unclear whether and to which extent artificial entities will have these capacities, and subsequently, whether they can be responsible in the same sense as human beings are. The integration of psychological and legal viewpoints showed that from both perspectives human beings need to be considered the most important beings in Hybrid Societies, whose well-being is of paramount importance. Both disciplines agree that human beings are currently the only actors in Hybrid Societies which possess all the capacities necessary for responsibility. Future developments toward possible responsible artificial agents will rely more on technological advancements and less on theoretical considerations. However, based on the conceptual clarification, further steps can now be taken to develop a concept of responsibility in Hybrid Societies. We propose that the integration of sociological perspectives into the model of responsibility should be considered.

Funding Open Access funding enabled and organized by Projekt DEAL. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 416228727 – SFB 1410.

Declarations

Conflict of interest On behalf of all authors, the corresponding authors state that there is no conflict of interest.

References

1. Gesmann-Nuissl, D.: Künstliche Intelligenz – den ersten Schritt vor dem zweiten tun! Zeitschrift zum Innovations- und Technikrecht (InTeR) **3**, 105–106 (2018)
2. Jentzsch, S., Schramowski, P., Rothkopf, C., Kersting, K.: The moral choice machine: Semantics derived automatically from language corpora contain human-like moral choices. In: Proceedings of the 2nd AAAI/ACM Conference on AI, Ethics, and Society. ACM, New York (2019) https://doi.org/10.1145/3306618.3314267

[11] Consisting of Regulation (EC) 765/2008; Decision 768/2008; Regulation (EU) 2019/1020 [159, 160, 161].

3. Beck, S.: Roboter, Cyborgs und das Recht – von der Fiktion zur Realität. In: Spranger, T.M., Dederer, H.-G., Herdegen, M., Müller-Terpitz, R. (eds.) Aktuelle Herausforderungen der Life Sciences, pp. 95–120. Lit Verlag, Berlin (2010)
4. Radbruch, G.: Legal Philosophy. Quelle & Meyer, Leipzig (1932)
5. v. Savigny, E.: Grundkurs im wissenschaftlichen Definieren. Deutscher Taschenbuchverlag, München (1976)
6. Hilgendorf, E.: Können Roboter schuldhaft handeln? Zur Übertragbarkeit unseres normativen Grundvokabulars auf Maschinen. In: Beck, S. (ed.) Jenseits von Mensch und Maschine, pp. 119–132. Nomos, Baden-Baden (2012)
7. Abbott, R.: The reasonable computer: disrupting the paradigm of tort liability. George Washington Law Rev. **86**, 1 (2018). https://doi.org/10.2139/ssrn.2877380
8. Johnson, N., Zhao, G., Hunsader, E., Qi, H., Johnson, N., Meng, J., Tivnan, B.: Abrupt rise of new machine ecology beyond human response time. Sci. Rep. **3**, 2627 (2013). https://doi.org/10.1038/srep02627
9. Scherer, M.U.: Regulating artificial intelligence systems: risks, challenges, competencies, and strategies. Harvard J. Law Technol. **29**(2), 354–400 (2016). https://doi.org/10.2139/ssrn.2609777
10. McCarthy, J.: What is Artificial Intelligence? http://jmc.stanford.edu/articles/whatisai.html (2007). Accessed 31 Jan 2022
11. Kant, I.: Metaphysik der Sitten / 1, Metaphysische Anfangsgründe der Rechtslehre (1797). In: The Philosophical Library 360. Unchanged eBook of the 4th, revisited and improved Ed. Felix Meiner Verlag, Hamburg (2018)
12. Gruber, M.-C.: Rechtssubjekte und Teilrechtssubjekte des elektronischen Geschäftsverkehrs. In: Beck, S. (ed.) Jenseits von Mensch und Maschine, pp. 133–160. Baden-Baden, Nomos (2012)
13. Matthias, A.: Automaten als Träger von Rechten, vol. 46. Logos Verlag, Berlin (2008)
14. John, R.: Haftung für Künstliche Intelligenz. Rechtliche Beurteilung des Einsatzes intelligenter Softwareagenten im E-Commerce, vol. 376. Verlag Dr Kovac, Hamburg (2007)
15. Lehmann, M.: Der Begriff der Rechtsfähigkeit. In: AcP 207, 225 ff (2007)
16. Moor, J.H.: The nature, importance, and difficulty of machine ethics. IEEE Intell. Syst. **21**(4), 18–21 (2006). https://doi.org/10.1109/MIS.2006.80
17. Hakli, R., Mäkelä, P.: Robots, autonomy, and responsibility. Front. Artif. Intell. Appl. **290**, 145–154 (2016). https://doi.org/10.3233/978-1-61499-708-5-145
18. Cave, S., Nyrup, R., Vold, K., Weller, A.: Motivations and risks of machine ethics. Proc. IEEE **107**(3), 562–574 (2019). https://doi.org/10.1109/JPROC.2018.2865996
19. Bauer, W.A.: Virtuous vs. Utilitarian artificial moral agents. AI Soc. **35**(1), 263–271 (2020). https://doi.org/10.1007/s00146-018-0871-3
20. Gunkel, D.J.: The Machine Question. Critical Perspectives on AI, Robots, and Ethics. Institute of Technology, Massachusetts (2012)
21. Hildebrandt, M.: The artificial intelligence of European Union law, 74–77. German Law J. (2020). https://doi.org/10.1017/glj.2019.99
22. Johnson, D.G.: Computer systems: Moral entities but not moral agents. Ethics Inf. Technol. **8**, 195–204 (2006). https://doi.org/10.1007/s10676-006-9111-5
23. Hegel, 1805–07, as cited in Gunkel (2018)
24. Gunkel, D.J.: The other question: can and should robots have rights? Ethics Inf. Technol. **20**, 87–99 (2018). https://doi.org/10.1007/s10676-017-9442-4
25. Merriam-Webster (n.d.). Machine. https://www.merriam-webster.com/dictionary/machine (2021). Accessed 8 Apr 2021
26. Anderson, S. L., Anderson, M.: The Consequences for Human Beings of Creating Ethical Robots. In: Proceedings of the 2007 AAAI Workshop Human Implications of Human-Robot Interaction, vol. 5, pp. 1 (2007)
27. Decker, M.: Ein Abbild des Menschen: Humanoide Roboter. In: Bölker, M., Guthmann, M., Hesse, W. (eds.) Information und Menschenbild, pp. 41–62. Springer, Hamburg (2010)
28. Beck, S.: In: Ebers, M., Heinze, C., Krügel, T., Steinrötter, B. (eds.) Künstliche Intelligenz und Robotik, Sec. 7. Beck, München (2020)
29. Vladeck, D.C.: Machines without principals: liability rules and artificial intelligence. Washington Law Rev. **89**, 117–150 (2014)
30. European Union. Council Directive 85/374/EEC of 25 July 1985 on the approximation of the laws, regulations and administrative provisions of the Member States concerning liability for defective products
31. European Union. Communication of the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions "Building a European Data Economy", COM/2017/09 final
32. Wagner, G: Robot Liability (2018). https://doi.org/10.2139/ssrn.3198764
33. Bostrom, N.: When machines outsmart humans. Futures **35**(7), 759–764 (2003)
34. Caliskan, A., Bryson, J.J., Narayanan, A.: Semantics derived automatically from language corpora contain human-like biases. Science **356**(6334), 183–186 (2017). https://doi.org/10.1126/science.aal4230
35. Bryson, J., Winfield, A.: Standardizing ethical design for artificial intelligence and autonomous systems. Computer **50**(5), 116–119 (2017). https://doi.org/10.1109/MC.2017.154
36. Bigman, Y.E., Waytz, A., Alterovitz, R., Gray, K.: Holding robots responsible: the elements of machine morality. Trends Cogn. Sci. **23**(5), 365–368 (2019). https://doi.org/10.1016/j.tics.2019.02.008
37. Gray, H.M., Gray, K., Wegner, D.M.: Dimensions of mind perception. Science **315**(5812), 619–619 (2007). https://doi.org/10.1126/science.1134475
38. Čapek, Karel, 1890–1938. R.U.R. (Rossum's universal robots). Penguin Books, London, New York (2004)
39. VDI-Guideline 2860. 1990–2005. Assembly and handling; handling functions, handling units; terminology, definitions and symbols
40. ISO 8373: 1994, revised by ISO 8373: Manipulating industrial robots – Vocabulary (2012)
41. Robot Institute of America: Robot Institute of America Worldwide Robotics Survey and Directory. Society of Manufacturing Engineer, Dearborn (1982)
42. Trevelyan, J.: Redefining robotics for the new millennium. Int. J. Robot. Res. **18**(12), 1211–1223 (1999). https://doi.org/10.1177/02783649922067816
43. Christaller, T., Decker, M., Gilsbach, J.M., Hirzinger, G., Lauterbach, K., Schweighofer, E., Schweitzer, D., Sturma, D.: Robotik – Perspektiven für menschliches Handeln in der zukünftigen Gesellschaft. Springer, Heidelberg (2003)
44. Bekey, G.A.: Autonomous Robots: from Biological Inspiration to Implementation and Control. Cambridge, MA (2005)
45. Müller, M.F.: Roboter und Recht. Eine Einführung. AJP/PJA **5**(2014), 595–608 (2014)
46. Balkin, J.M.: The path of robotics law. California Law Rev. Circuit **6**, 45–60 (2015)
47. Calo, R.: Robotics and the lesson of cyberlaw. California Law Rev. **103**, 513–565 (2015)
48. Robotic Industries Association. *Robot Terms and Definitions*. Robotics Online. https://www.robotics.org/product-catal

og-detail.cfm/Robotic-Industries-Association/Robot-Terms-and-Definitions/productid/2953. Accessed 2021
49. Bartneck, C., Forlizzi, J.: A design-centred framework for social human-robot interaction. RO-MAN 2004. In: 13th IEEE International Workshop on Robot and Human Interactive Communication (IEEE Catalog No.04TH8759), pp. 591–594 (2004). https://doi.org/10.1109/ROMAN.2004.1374827
50. Elder, A.: "How could you even ask that?": Moral Considerability. Uncertain. Vulnerability Soc. Robot. (2020). https://doi.org/10.25779/G8A3-F992
51. Naneva, S., Sarda Gou, M., Webb, T.L., Prescott, T.J.: A systematic review of attitudes, anxiety, acceptance, and trust towards social robots. Int. J. Soc. Robot. **12**, 1179–1200 (2020). https://doi.org/10.1007/s12369-020-00659-4
52. Zimmerli, W.C.: Der Mensch wird uns erhalten bleiben. Digitalisierte Welt und die Zukunft des Humanismus. Forschung Lehre **9**(2000), 455–457 (2000)
53. Xanke, L., Bärenz, E.: Künstliche Intelligenz in Literatur und Film – Fiktion oder Realität? J. N. Front. Spatial Concepts. **4**, 36–43 (2012)
54. Brunhöber, B.: Individuelle Autonomie und Technik im Körper. In: Beck, S. (ed.) Jenseits von Mensch und Maschine, pp. 77–104. Nomos, Baden-Baden (2012)
55. Spreen, D: Cyborgs und andere Techno-Körper. Ein Essay im Grenzbereich von Bios und Techne. EDFC e.V., Passau (1998)
56. Schmaucks, D.: Kulturethologische Aspekte in Stanislaw Lems „Summa technologiae". Ein Brückenschlag zwischen Kulturethologie und Futurologie. Matreier Gespräche – Schriftenreihe der Forschungsgemeinschaft Wilheminenberg, pp. 213–230 (2007)
57. Faßler, M.: Hybridität: Welche Realität wie? In: Christaller, T., Wehner, J. (eds.) Autonome Maschinen, pp. 268–288. Westdt. Verlag, Wiesbaden (2003)
58. Heilinger, J.-C., Müller, O.: Der Cyborg. Anthropologische und ethische Überlegungen. In: Manzeschke, A., Karsch, F. (eds). Roboter, Computer und Hybride. Was ereignet sich zwischen Menschen und Maschinen, vol. 5, pp. 47–66. Nomos Verlagsgesellschaft (2016)
59. Meyer, B., Asbrock, F.: Disabled or cyborg? How bionics affect stereotypes toward people with physical disabilities. Front. Psychol. **9**, 2251 (2018). https://doi.org/10.3389/fpsyg.2018.02251
60. Beck, S.: Brauchen wir ein Roboterrecht? Ausgewählte juristische Fragen zum Zusammenleben von Menschen und Robotern. In: Japanisch-Deutsches Zentrum (eds.) Mensch-Roboter-Interaktionen aus interkultureller Perspektive. Japan und Deutschland im Vergleich, pp 124–146. Berlin (2014)
61. MacDorman, K.F., Ishiguro, H.: The uncanny advantage of using androids in cognitive and social science research. Interact. Stud. **7**(3), 297–337 (2006). https://doi.org/10.1075/is.7.3.03mac
62. Mori, M., MacDorman, K., Kageki, N.: The uncanny valley [from the field]. IEEE Robot. Autom. Mag. **19**(2), 98–100 (2012). https://doi.org/10.1109/MRA.2012.2192811
63. Hanson, D.: Exploring the Aesthetic Range for Humanoid Robots. In: Proc ICCS/Cog-Sci-2006 long Symp Towar Soc Mech android Sci, pp. 39–42 (2006). https://www.researchgate.net/publication/228356164_Exploring_the_aesthetic_range_for_humanoid_robots
64. MacDorman, K.F.: Masahiro Mori und das unheimliche Tal: Eine Retrospektive. Zenodo (2019). https://doi.org/10.5281/ZENODO.3226274
65. Rosenthal-von der Pütten, A.M., Krämer, N.C., Becker-Asano, C., Ogawa, K., Nishio, S., Ishiguro, H.: The Uncanny in the wild. Analysis of unscripted human-android interaction in the field. Int. J. Soc. Robot. **6**(1), 67–83 (2014). https://doi.org/10.1007/s12369-013-0198-7
66. von der Pütten, A. M., Krämer, N. C.: A survey on robot appearances. In: Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction – HRI, vol. 12, pp. 267–268 (2012). https://doi.org/10.1145/2157689.2157787
67. McCarthy, J., Minsky, M.L., Rochester, N., Shannon, C.E.: A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. 3 (1995)
68. Kaulartz, M., Braegelmann, T.: Rechtshandbuch Artificial Intelligence und Machine Learning. Nomos/Beck, München (2020)
69. Plessner, H.: Levels of Organic Life and the Human—An Introduction to Philosophical Anthropology. de Gruyter & Co, Berlin (1928)
70. Russel, S., Norvig, R.: Artificial Intelligence: A Modern Approach. Global Edition. 4th ed. Pearson Education (2021)
71. Deutsche Normungsroadmap Künstliche Intelligenz. https://www.din.de/resource/blob/772438/6b5ac6680543eff9fe372603514be3e6/normungsroadmap-ki-data.pdf. Accessed 2020
72. European Union. Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative Act, COM/2021/206 final
73. Gesmann-Nuissl, D.; Meyer, S. Black Hole instead of Black Box? - The Double Opaqueness of Recommender Systems on Gaming Platforms and its Legal Implications. In. Collected Volume on Conference on Recommender Systems: Legal and Ethical Issues. (in appearance). Bonn. (2022)
74. American Psychological Association. (n.d.-b). Intelligence. In: APA dictionary of psychology https://dictionary.apa.org/intelligence. Accessed 11 Nov 2021
75. Hoffmann, C.H., Hahn, B.: Decentered ethics in the machine era and guidance for AI regulation. AI Soc. (2019). https://doi.org/10.1007/s00146-019-00920-z
76. Rabinowitz, N.C., Perbet, F., Song, H.F., Zhang, C., Eslami, S.M.A., Botvinick, M.: Machine Theory of Mind. (2018) http://arxiv.org/abs/1802.07740
77. Cuzzolin, F., Morelli, A., Cîrstea, B., Sahakian, B.J.: Knowing me, knowing you: theory of mind in AI. Psychol. Med. **50**(7), 1057–1061 (2020). https://doi.org/10.1017/S0033291720000835
78. Ryan, M.: In AI we trust: ethics, artificial intelligence, and reliability. Sci. Eng. Ethics (2020). https://doi.org/10.1007/s11948-020-00228-y
79. Kremnitzer, M., Ghanayim, K.: Die Strafbarkeit von Unternehmen. Zeitschrift für die gesamte Strafrechtswissenschaft **13**(3), 539–564 (2009). https://doi.org/10.1515/zstw.2001.113.3.539
80. Merriam-Webster (n.d.). Responsibility. https://www.merriam-webster.com/dictionary/responsibility (2021). Accessed 8 Apr 2021
81. Chinen, M.A.: The co-evolution of autonomous machines and legal responsibility. Virginia J. Law Technol. **20**(02), 338–393 (2016)
82. Beck, S.: Die Diffusion strafrechtlicher Verantwortlichkeit durch Digitalisierung und Lernende Systeme. Zeitschrift für Internationale Strafrechtsdogmatik **2**, 41–50 (2020)
83. Asaro, P.M.: Determinism, machine agency, and responsibility. Politica & Societa **2**, 265–292 (2014)
84. Behdadi, D., Munthe, C.: A normative approach to artificial moral agency. Mind. Mach. **30**(2), 195–218 (2020). https://doi.org/10.1007/s11023-020-09525-8
85. Coeckelbergh, M.: Artificial intelligence, responsibility attribution, and a relational justification of explainability. Sci. Eng. Ethics (2019). https://doi.org/10.1007/s11948-019-00146-8
86. Dignum, V., Baldoni, M., Baroglio, C., Caon, M., Chatila, R., Dennis, L., Génova, G., Haim, G., Kließ, M. S., Lopez-Sanchez, M., Micalizio, R., Pavón, J., Slavkovik, M., Smakman, M., van Steenbergen, M., Tedeschi, S., van der Toree, L., Villata, S., de

Wildt, T.: Ethics by Design: Necessity or Curse? In: Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, pp. 60–66 (2018). https://doi.org/10.1145/3278721.3278745
87. Floridi, L., Sanders, J.W.: On the morality of artificial agents. Mind. Mach. **14**, 349–379 (2004)
88. Suárez-Gonzalo, S.: Tay is you. The attribution of responsibility in the algorithmic culture. Observatorio OBS* **13**, 2 (2019). https://doi.org/10.15847/obsOBS13220191432
89. Willemsen, P.: Direct and derivative moral responsibility: an overlooked distinction in experimental philosophy [Preprint]. PsyArXiv (2020). https://doi.org/10.31234/osf.io/bz38e
90. Etschmaier, M.M., Lee, G.: Defining the paradigm of a highly automated system that protects against human failures and terrorist acts and application to aircraft systems. Int. J. Comput. Appl. **23**, 1 (2016)
91. Palandt/Grüneberg, C., Ellenberger, J.: Bürgerliches Gesetzbuch. C.H. Beck, München (2021)
92. Berger, M.: Treu und Glauben und vorvertragliche Aufklärungspflichten im US-amerikanischen und deutschen Recht. Cuviller Verlag, Göttingen (2003)
93. Kelsen, H.: Reine Rechtslehre. Franz Deuticke, Leipzig (1934)
94. Aydin, T.: Gustav Radbruch, Hans Kelsen und der Nationalsozialismus. Zwischen Recht, Unrecht und Nicht-Recht. Nomos, Baden-Baden (2020)
95. Beauchamp, T.L., Childress, J.F.: Principles of Biomedical Ethics, 7th edn. Oxford University Press, Oxford (2009)
96. Hoerster N.: In: Jordan, S., Nimtz, C. (eds.) Lexikon Philosophie: Hundert Grundbegriffe, pp. 80–84. Reclam, Stuttgart (2009)
97. Neuhäuser, C.: Künstliche Intelligenz und ihr moralischer Standpunkt. In: Beck, S. (ed.) Jenseits von Mensch und Maschine, pp. 23–42. Nomos, Baden-Baden (2012)
98. American Psychological Association. (n.d.-c). Morality. In: APA dictionary of psychology. https://dictionary.apa.org/morality (2021). Accessed 1 Dec 2021
99. American Psychological Association. (n.d.-a). Ethics. In: APA dictionary of psychology. https://dictionary.apa.org/ethics. Accessed 1 Dec 2021
100. Doris, J., Stich, S., Phillips, J., Walmsley, L.: Moral Psychology: Empirical Approaches. The Stanford Encyclopedia of Philosophy (2020). https://plato.stanford.edu/archives/spr2020/entries/moral-psych-emp
101. Kohlberg, L.: The Psychology of Moral Development: The Nature and Validity of Moral Stages, vol. 2. Haper & Row, Manhatten (1984)
102. Rest, J.R.: Moral Development. Advances in Theory and Research. Praeger, New York (1986)
103. Rest, J.R., Narvaez, D., Bebeau, M.J., Thoma, S.J.: DIT2: devising and testing a revised instrument of moral judgment. J. Educ. Psychol. **91**(4), 644–659 (1999)
104. Hannah, S.T., Avolio, B.J., May, D.R.: Moral Maturation and moral conation. A capacity approach to explaining moral thought and action. Acad. Manage. Rev. **36**(4), 663–685 (2011)
105. Strobel, A., Grass, J., Pohling, R., Strobel, A.: Need for Cognition as a moral capacity. Personality Individ. Differ. **117**, 42–51 (2017)
106. Eigenstetter, M., Strobel, A., Stumpf, S.: Diagnostik ethischer Kompetenz. In: Kaiser, S., Kozica, A. (eds.) Ethik in Personalmanagement: zentrale Konzepte, Ansätze und Fragestellungen. Hampp, München (2012)
107. Pohling, R., Bzdok, D., Eigenstetter, M., Stumpf, S., Strobel, A.: What is ethical competence? The role of empathy, personal values, and the Five-Facgtor Model of Personality in ethical decision making. J. Bus. Ethics **137**(3), 449–474 (2016). https://doi.org/10.1007/s10551-015-2569-5
108. Danaher, J.: The rise of the robots and the crisis of moral patiency. AI Soc. **34**(1), 129–136 (2019). https://doi.org/10.1007/s00146-017-0773-9
109. Johansson, L.: The functional morality of robots. Int. J. Technoethics **1**(4), 65–73 (2010). https://doi.org/10.4018/jte.2010100105
110. Kirn, S. Müller-Hengstenberg, C.D.: Intelligente (Software) Agenten: Eine neue Herausforderung für die Gesellschaft und unser Rechtssystem. FZID Discussion Paper No. 86–2014 (2014)
111. Klement, J.: Verantwortung. Mohr Siebeck, Tübingen (2006)
112. Darling, K.: Extending legal protection to social robots: The effects of anthropomorphism, empathy, and violent behaviour towards robotic objects. In: Ryan, C., Froomkin, A.M., Kerr, I. (eds.) Robot Law, pp. 213–231. Edward Elgar Publishing, Cheltenham (2016)
113. Gaede, K.: Künstliche Intelligenz – Rechte und Strafen für Roboter. Nomos, Baden-Baden (2019)
114. Gellers, J.C.: Rights for Robots: Artificial Intelligence, Animal and Environmental Law. Routledge, Milton Park (2021)
115. Maia Alexandre, F.: The legal status of artificially intelligent robots: personhood. Tax. Control. (2017). https://doi.org/10.2139/ssrn.2985466
116. Schirmer, J.-E.: Rechtsfähige Roboter? JuristenZeitung **71**(13), 660–666 (2016)
117. Reed, C., Kennedy, E., Silva, S.: Responsibility, Autonomy and Accountability: Legal Liability for Machine Learning. Queen Mary School of Law Legal Studies Research Paper No 243/2016. (2016)
118. Sheriff, K.: Defining Autonomy in the Context of Tort Liability: Is Machine Learning Indicative of Robotic Responsibility. (2015). https://doi.org/10.2139/ssrn.2735945
119. Haagen, C.: Verantwortung für Künstliche Intelligenz. Nomos, Baden-Baden (2021)
120. Pasquale, F.A.: Toward a Fourth Law of Robotics: Preserving Attribution, Responsibility, and Explainability in an Algorithmic Society. Ohio State Law Journal, 78. University of Maryland Legal Studies Research Paper No. 2017.21. (2017)
121. Chopra, S., White, L.F.: A Legal Theory for Autonomous Artificial Agents. The University of Michigan Press, Ann Arbor (2011)
122. Creifelds/Fuchs, J. Aichberger, T., Groh, G. Schmidt, A.: Rechtswörterbuch. Handlungsfähigkeit. Rechtsfähigkeit. Gefährdungshaftung, 26. (2021)
123. Zippelius, R.: Das Wesen des Rechts. Eine Einführung in die Rechtstheorie **6**, 2 (2012)
124. Frankfurt, H.G.: Freedom of the will and the concept of a person. J. Philos. **68**(1), 5–20 (1971)
125. Albers, M.: Informationelle Selbstbestimmung. Nomos, Baden-Baden (2005)
126. European Union. European Parliament, Resolution of 16 February 2017 with recommendations to the Commission on Civil Law. Rules on Robotics, 2018/C 252/25 (2018)
127. Allen, C., Varner, G., Zinser, J.: Prolegomena to any future artificial moral agent. J. Exp. Theor. Artif. Intell. **12**(3), 251–261 (2000). https://doi.org/10.1080/09528130050111428
128. Mabaso, B.A.: Computationally rational agents can be moral agents. Ethics Inf. Technol. (2020). https://doi.org/10.1007/s10676-020-09527-1
129. Misselhorn, C.: Artificial morality. Concepts, issues and challenges. Society **55**(2), 161–169 (2018). https://doi.org/10.1007/s12115-018-0229-y
130. Tigard, D.W.: There is no techno-responsibility gap. Philos. Technol. **34**(3), 589–607 (2021). https://doi.org/10.1007/s13347-020-00414-7
131. Hancock, P.A., Billings, D.R., Schaefer, K.E., Chen, J.Y.C., de Visser, E.J., Parasuraman, R.: A meta-analysis of factors

affecting trust in human-robot interaction. Hum. Factors **53**(5), 517–527 (2011). https://doi.org/10.1177/0018720811417254
132. Ivanov, S., Kuyumdzhiev, M., Webster, C.: Automation fears: drivers and solutions. Technol. Soc. **63**, 101431 (2020). https://doi.org/10.1016/j.techsoc.2020.101431
133. De Visser, E.J., Monfort, S.S., McKendrick, R., Smith, M.A.B., McKnight, P.E., Krueger, F., Parasuraman, R.: Almost human. Anthropomorphism increases trust resilience in cognitive agents. J. Exp. Psychol.: Appl. **22**(3), 331–349 (2016). https://doi.org/10.1037/xap0000092
134. Bostrom, N., Yudkowsky, E.: Ethics of artificial intelligence. In: Frankish, K., Ramsey, W. (eds.) The Ethics of Artificial Intelligence, p. 21. Cambridge University Press, Cambridge (2014)
135. Floridi, L., Cowls, J., King, T.C., Taddeo, M.: How to design ai for social good: seven essential factors. Sci. Eng. Ethics **26**(3), 1771–1796 (2020). https://doi.org/10.1007/s11948-020-00213-5
136. Miller, T.: Explanation in Artificial Intelligence: Insights from the Social Sciences. [Cs] (2018). http://arxiv.org/abs/1706.07269
137. Vanderelst, D., Willems, J.: Can we agree on what robots should be allowed to do? An exercise in rule selection for ethical care robots. Int. J. Soc. Robot. (2019). https://doi.org/10.1007/s12369-019-00612-0
138. Bandura, A.: Social cognitive theory of moral thought and action. In: Kurtines, W.M., Gewirtz, J.L. (eds.) Handbook of Moral Behavior and Development. Lawrence Erlbaum Associates, New York (1991)
139. Gordon, J.-S.: Building moral robots: ethical pitfalls and challenges. Sci. Eng. Ethics **26**(1), 141–157 (2020). https://doi.org/10.1007/s11948-019-00084-5
140. Tolmeijer, S., Kneer, M., Sarasua, C., Christen, M., Bernstein, A.: Implementations in machine ethics: a survey. ACM Comput. Surv. **56**(6), 1–38 (2020)
141. Winfield, A.F., Michael, K., Pitt, J., Evers, V.: Machine ethics: the design and governance of ethical AI and autonomous systems [scanning the issue]. Proc. IEEE **107**(3), 509–517 (2019). https://doi.org/10.1109/JPROC.2019.2900622
142. Pazzanese, C.: Great promise but potential for peril: Ethical concerns mount as AI takes bigger decision-making role in more industries. The Harvard Gazette. https://news.harvard.edu/gazette/story/2020/10/ethical-concerns-mount-as-ai-takes-bigger-decision-making-role/ (2020). Accessed 26 Oct 2020
143. Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J.-F., Rahwan, I.: The moral machine experiment. Nature **563**(7729), 59–64 (2018). https://doi.org/10.1038/s41586-018-0637-6
144. Frank, D.-A., Chrysochou, P., Mitkidis, P., Ariely, D.: Human decision-making biases in the moral dilemmas of autonomous vehicles. Sci. Rep. **9**(1), 13080 (2019). https://doi.org/10.1038/s41598-019-49411-7
145. Bonnefon, J.-F., Shariff, A., Rahwan, I.: The social dilemma of autonomous vehicles. Science **352**(6293), 1573–1576 (2016). https://doi.org/10.1126/science.aaf2654
146. Malle, B. F., Scheutz, M., Arnold, T., Voiklis, J., Cusimano, C.: Sacrifice one for the good of many: people apply different moral norms to human and robot agents. In: Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction – HRI, 15, pp. 117–124 (2015). https://doi.org/10.1145/2696454.2696458
147. Gamez, P., Shank, D.B., Arnold, C., North, M.: Artificial virtue: the machine question and perceptions of moral character in artificial moral agents. AI Soc. (2020). https://doi.org/10.1007/s00146-020-00977-1
148. Kim, B., Wen, R., Zhu, Q., Williams, T., Phillips, E.: Robots as Moral Advisors: The Effects of Deontological, Virtue, and Confucian Role Ethics on Encouraging Honest Behavior. In: Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction, pp. 10–18 (2021). https://doi.org/10.1145/3434074.3446908
149. Wen, R.: Toward Hybrid Relational-Normative Models of Robot Cognition. In: Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction, pp. 568–570 (2021). https://doi.org/10.1145/3434074.3446353
150. Mandl, S., Bretschneider, M., Meyer, S., Gesmann-Nuissl, D., Asbrock, F., Meyer, B., Strobel, A.: Embodied digital technologies: first insights in social and legal perception of robots and users of prostheses. Front. Robot. AI **9**, 787970 (2022). https://doi.org/10.3389/frobt.2022.787970
151. Bertolini, A.: Robots as products: the case for a realistic analysis of robotic applications and liability rules. Law Innov. Techno. **5**(2), 214–247 (2013). https://doi.org/10.5235/17579961.5.2.214
152. Eiben, A.E. In Vivo Veritas: Towards the Evolution of Things. In: B. Filipic, T. Bartz-Beielstein, J. Branke, and J. Smith (eds.) Proceedings of the 13th International Conference on Parallel Problem Solving from Nature (PPSN 2014), pp. 24–39. (2014)
153. Johnson, D.G., Verdicchio, M.: Why robots should not be treated like animals. Ethics Inf. Technol. **20**(2), 291–301 (2018). https://doi.org/10.1007/s10676-018-9481-5
154. Bartneck, C., Hoek, M. v. d., Mubin, O., & Mahmud, A. A.: "Daisy, Daisy, Give me your answer do!" - Switching off a robot. In: Proceedings of the 2nd ACM/IEEE International Conference on Human-Robot Interaction, Washington DC, pp. 217–222. (2007). https://doi.org/10.1145/1228716.1228746
155. Knight, W.: The foundations of AI are riddled with errors. Wired.Com. https://www.wired.com/story/foundations-ai-riddled-errors/ Accessed 31 Mar 2021
156. Grother, P., Ngan, M., Hanaoka, K.: Face recognition vendor test part 3: Demographic effects (NIST IR 8280; p. NIST IR 8280). National Institute of Standards and Technology (2019) https://doi.org/10.6028/NIST.IR.8280
157. Kleinberg, J., Ludwig, J., Mullainathan, S., Sunstein, C.R.: Discrimination in the age of algorithms. J. Legal Anal. **10**, 113–174 (2018)
158. Graham, J., Meindl, P., Beall, E., Johnson, K.M., Zhang, L.: Cultural differences in moral judgment and behavior, across and within societies. Curr. Opin. Psychol. **8**, 125–130 (2016). https://doi.org/10.1016/j.copsyc.2015.09.007
159. European Union. Regulation (EC) No 765/2008 of the European Parliament and of the Council of 9 July 2008 setting out the requirements for accreditation and market surveillance relating to the marketing of products and repealing Regulation (EEC) No 339/93
160. European Union. Decision No 768/2008/EC of the European Parliament and of the Council of 9 July 2008 on a common framework for the marketing of products, and repealing Council Decision 93/465/EEC
161. European Union. Regulation (EU) 2019/1020 of the European Parliament and of the Council of 20 June 2019 on market surveillance and compliance of products and amending Directive 2004/42/EC and Regulations (EC) No 765/2008 and (EU) No 305/2011