

Digital Personhood?

The Status of Autonomous Software Agents in Private Law*

Gunther Teubner, Frankfurt am Main

(Ancilla Iuris 2018, <https://www.anci.ch>)

- I. *Digitalization: Gaps in responsibility and deficits in legal doctrine*
 - 1. *The whistling in the woods or: The fear of the homo ex machina*
 - 2. *Specific liability gaps*
 - 3. *Doctrinal misconceptions*
- 4. *Full legal subjectivity?*
- II. *Actants: Autonomy risk*
 - 1. *Anthropomorphism?*
 - 2. *Communication with actants*
 - 3. *Gradualized digital autonomy*
 - 4. *Legal criteria of autonomy*
 - 5. *Autonomous decisions of software agents*
- III. *Legal problems concerning the autonomy risk*
 - 1. *Digital contracts*
 - 2. *Contractual liability*
 - 3. *Tortious liability*
 - 4. *Strict liability or digital assistance liability?*
- IV. *Hybrids: Association risk*
 - 1. *Human-machine associations*
 - 2. *Hybrids as legal entities?*
- V. *Multi-agent systems: Network risk*
- VI. *Results*

I. Digitalization: Gaps in responsibility and deficits in legal doctrine

1. The whistling in the woods or: The fear of the homo ex machina

“Behavioral biology describes whistling in the woods ... as a sound marker, as it is customary when demarcating territory, and is intended to keep predators, that could become dangerous to humans, at a distance. By signaling the territorial claim one is distracted by one’s own fear of inferiority.” (Wikipedia) The dangerous predators that invade the territories of private law are uncontrollable algorithms from the digital world, robots, software agents with a high level of intelligence and the ability to learn and generate new kinds of undreamt-of dangers for humans. However, many legal scholars propose to keep digital predators at bay by making a lot of noise: All the hazards of these new digital species, within liability law as well as contract law, could according to them “be dealt with via traditional legal instruments”.¹

* For critical and constructive comments I would like to thank Marc Amstutz, Pasquale Femia, Malte Gruber and Dan Wielsch. An extended German version of this article appears in: 218 *Archiv für civilistische Praxis*, 2018, 155-204.

¹ For German legal doctrine, Münchener Kommentar BGB/Busche, 7th ed. 2015, BGB-AT, vor § 145 Rn. 37.

They see no reason to deviate from the established categories of attribution, which always imply that computer behavior is human behavior. Software agents could be legally treated without problems as mere machines, as willing instruments in the hands of their human masters.²

However, autonomous software agents are generating previously unknown risks, so that new rules in private law are probably needed to counteract them. Perhaps, software agents even have to be granted an independent status under private law. The decisive factor is likely to be whether the rapid digital developments have already opened up unacceptable gaps in responsibility. That is why the European Parliament and some legal authors advocate the counter-argument³: Software agents inevitably cause these gaps in responsibility, because their autonomous actions entail a massive loss of control for human actors. At the same time, however, society is becoming increasingly dependent on autonomous software agents on a large scale and it is highly unlikely that society will abandon their use.⁴ Private law is therefore faced with a choice: Either it gives autonomous software agents an independent legal status and treats them themselves as responsible actors, or there is an increasing number of "accidents", without anyone being responsible for them. The dynamics of digitalization are constantly creating responsible-free spaces that will expand in the future.⁵

2. Specific liability gaps

What are the specific gaps in responsibility? In information science the following *scenarios* are mentioned: Deficiencies arise in practice when the software is produced by teams, when management decisions are just as important as programming decisions, when documentation of requirements and specifications plays a major role in the resulting code, when, despite the testing of code accuracy, a lot depends on "off-the-shelf" components whose origin and accuracy is unclear, when the performance of the software is the result of the accompanying controls and not of program creation, when automated instruments are used in the design of the software, if the operation of the algorithms is influenced by its *interfaces* or even by system traffic, if the software interacts in an unpredictable manner, if the software works with probabilities or has adaptability or is the result of another program.⁶

² In Germany, Bundesgerichtshof BGHZ 195, 126, Rn. 17; Palandt/Ellenberger, BGB 77th ed. 2018, vor § 116 Rn. 2; Dieter Medicus and Jens Petersen (2017) *Bürgerliches Recht*, 26th ed., München: Vahlen, Rn. 256; Christoph Sorge (2006) *Softwareagenten: Vertragsschluss, Vertragsstrafe, Reugeld*, Karlsruhe: Universitätsverlag, 36; Stefan Kirn and Claus-Dieter Müller-Hengstenberg (2015) "Technische und rechtliche Betrachtungen zur Autonomie kooperativ-intelligenter Softwareagenten", 29 *KI – Künstliche Intelligenz*, 59-74, 67.

³ European Parliament, <http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML+REPORT+A8-2017-0005+0+DOC+PDF+V0//EN>. Pointing at the novel risk of autonomy, Herbert Zech (2016) "Zivilrechtliche Haftung für den Einsatz von Robotern: Zuweisung von Automatisierungs- und Autonomierisiken", in: Sabine Gless and Kurth Seelmann (ed) *Intelligente Agenten und das Recht*, Baden-Baden: Nomos, 163-204, 175 f.; Ruth Janal (2016) "Die deliktische Haftung beim Einsatz von Robotern: Lehren aus der Haftung für Sachen und Gehilfen", in: Sabine Gless and Kurth Seelmann (ed) *Intelligente Agenten und das Recht*, Baden-Baden: Nomos, 139-162, 158; Susanne Beck (2017) "Der rechtliche Status autonomer Maschinen", *Aktuelle juristische Praxis*, 183-191, 184.

⁴ Andreas Matthias (2010) *Automaten als Träger von Rechten*, 2nd ed., Berlin: Logos, 15.

⁵ This is Matthias's central thesis, (fn. 4), 111.

⁶ Luciano Floridi and J.W. Sanders "On the Morality of Artificial Agents", in: Michael Anderson and Susan L. Anderson (ed) *Machine Ethics*, Cambridge: Cambridge University Press, 184-212, 205.

In law, the emergence of software agents has created the following liability gaps:

Computer networking: The most difficult liability gap to correct arises in the case of multiple causality in the occurrence of damaging actions of several computers in a network. The liability norms of the applicable law do not go any further here.⁷ There is also no sign of a useful proposal *de lege ferenda*. In the case of high frequency trading, this risk has become apparent.⁸ “Who bears the risk if, as on May 6 2010, a flash crash caused a massive Dow Jones price drop on the New York Stock Exchange due to the algorithms of Wall Street, which control the trading systems, behaving for some time in an uncontrolled and incomprehensible manner and causing a loss of billions?”⁹

Big Data: Further liability gaps are caused by incorrect estimates of incorrect Big Data forecasts. Big Data is used to predict how existing societal trends or epidemics can develop and—if necessary—be influenced by extraordinarily large amounts of data.¹⁰ If the reason for the faulty calculation cannot be clearly established, there are difficulties in determining causality and misconduct.

Digital breach of contract: If an autonomous software agent violates contractual obligations, the prevailing doctrine argues that the liability rules for auxiliary persons cannot be applied, as the agent does not have legal capacity to act. Instead, liability shall only arise if the human principal himself commits a breach of contract. This creates a problematic liability gap: If the operator can prove that the software agent has been used correctly without the operator himself having violated a contractual obligation, the operator is exempted from any liability.¹¹ Should the customer then bear the damage caused by the other party's computer?

Non-contractual liability: A similar problem arises in the area of non-contractual liability because, in the case of fault-based liability, it is only the breach of duty by the operator, manufacturer or programmer that is important. If the people involved comply with these obligations, then there is no liability.¹² The liability gap will

⁷ So clearly, Indra Spiecker (2016) "Zur Zukunft systemischer Digitalisierung: Erste Gedanken zur Haftungs- und Verantwortungszuschreibung bei informationstechnischen Systemen - Warum für die systemische Haftung ein neues Modell erforderlich ist", *Computer und Recht*, 698-704.

⁸ More on this, Malte-Christian Gruber (2015) "On Flash Boys and Their Flashbacks: The Attribution of Legal Responsibility in Algorithmic Trading", in: Marlena Jankowska, Marcin Kulawiak and Mirosław Pawełczykai (ed) *AI: Law, Philosophy & Geoinformatics*, Warsaw: Prawa Gospodarczego, 88-102, 100.

⁹ Stefan Kirn and Claus-Dieter. Müller-Hengstenberg (2014) "Kirn, S.; Müller-Hengstenberg, C.-D.: Intelligente (Software-)Agenten: Von der Automatisierung zur Autonomie? – Verselbstständigung technischer Systeme", *Multimedia und Recht*, 225-232, 1.

¹⁰ Wolfgang Hackenberg in: Thomas Hoeren, Ulrich Sieber and Bernd Holznagel (ed) *Handbuch Multimedia-Recht: Rechtsfragen des elektronischen Geschäftsverkehrs*, 45th ed., München: Beck, 39.EL 2014, Teil 16.7 Rn. 58.

¹¹ E.g., Münchener Kommentar BGB/Grundmann, 7th ed. 2016, BGB § 278 Rn. 46; Kirn/Müller-Hengstenberg (fn. 9), 16.

¹² E.g. Gerald Spindler (2016) "Digitale Wirtschaft – analoges Recht: Braucht das BGB ein Update?", 71 *JuristenZeitung*, 805–856, 816; also Peter Bräutigam and Thomas Klindt (2015) "Industrie 4.0, das Internet der Dinge und das Recht", *Neue Juristische Wochenschrift*, 1137-1142.

not be closed, even if the courts overstretch duties of care for human actors.¹³ If the damage is caused by autonomous robots, the injured party will be without protection.

Liability for industrial hazards: Even the legal policy proposals, which propose *de lege ferenda* to compensate for damages caused by software agents with *industrial hazard* liability standards, cannot avoid substantial liability gaps. The guiding principles of strict liability can hardly serve as a role model since they are not tailored to the digital risk.

Computer declarations: A unsatisfactory liability situation also arises in the law of the legal computer declarations if the software agent himself, like when a "*falsus procurator*" misrepresents a third party as the principal. In such cases, the risk should lie entirely with the principal of the software agent. Many authors see this as an excessive burden that cannot be justified, especially when it comes to distributed action or self-cloning.¹⁴

3. Doctrinal misconceptions

The lack of responsibility arises as long as legal doctrine insists on responding to the new digital realities exclusively with traditional conceptual instruments. However, in order to keep pace with the digital developments to a certain extent, doctrine sees itself forced to counter the hitherto unknown software agents with questionable auxiliary constructions. In the field of contract law, legal doctrine firmly maintains that only natural persons are in a position to make legally binding declarations. Therefore, contract law has to work with problematic fictions. In the law of contractual and non-contractual liability, if damages are attributable to a human-computer network, it must attribute the damage-causing event exclusively to the action share of the human being and is then no longer able to determine in detail the liability prerequisites. And the rules on strict liability on the one hand go too far and on the other hand not far enough because they treat the digital risk like the mere causal risk of a dangerous thing. Finally, there is general perplexity regarding the networking of multi-agent systems.

If, in these cases, the law reacts to the use of autonomous software agents only with conventional conceptual means and thus leaves gaps in responsibility open, the damage is not distributed collectively across society, but rather a merciless *casum sentit dominus* applies. This is where there is massive criticism. To impose the consequences of the damage on the injured party is rightly criticized as wrong in legal policy terms and as fundamentally unfair. Provided that in such situations the failure of autonomous software agents remains free of liability, this creates false incentives for operators, producers and programmers.¹⁵ And society's willingness to fully exploit the promising potential of autonomous software agents is diminishing.

¹³ Criticizing the trend toward overloading of duties, Malte-Christian Gruber (2015) *Bioinformatiionsrecht: Zur Persönlichkeitsentfaltung des Menschen in technisierter Verfassung*, Tübingen: Mohr Siebeck; Janal (fn. 3), 155.

¹⁴ Giovanni Sartor (2003) "Agents in Cyberlaw", *Proceedings of the Workshop on the Law of Electronic Agents* (LEA 2002), http://www.cirfid.unibo.it/~sartor/sartorpapers/g Sartor2002_agenti_software.pdf, !

¹⁵ Horst Eidenmüller (2017) "The Rise of Robots and the Law of Humans", *SSRN Electronic Journal*, <https://ssrn.com/abstract=2941001>, 8.

Above all, however, it contradicts the postulate of justice that maintains a necessary connection between decision and responsibility.¹⁶

4. Full legal subjectivity?

Full legal capacity for software agents—this is the much-discussed answer of many lawyers and politicians to these shortcomings.¹⁷ The European Parliament decided in January 2017, on the basis of the Delvaux report, to establish a special legal status for robots, so that at least for the most sophisticated autonomous robots a status as electronic persons with special rights and obligations could be established, including the redress of all the damage they cause. When robots make autonomous decisions, they should be recognized as "electronic persons", as legal entities in the full sense of the word.¹⁸

It has been suggested in the literature that, to compensate for the above-mentioned shortcomings, e-persons should have the ability to make declarations of intent as full legal entities, both in their own name and in the name of others,¹⁹ that they own property, have money, have bank accounts and should have access to credit. Indeed, that they collect commissions for their transactions and use this self-

¹⁶ European Parliament (fn. 3), 7.

¹⁷ Susanne Beck "Über Sinn und Unsinn von Statusfragen: Zu Vor- und Nachteilen der Einführung einer elektronischen Person", in: Eric Hilgendorf and Jan-Philipp Günther (ed) *Robotik und Gesetzgebung*, Baden-Baden: Nomos, 239–260, 255 ff.; Jan-Philip Günther (2016) *Roboter und rechtliche Verantwortung: Eine Untersuchung der Benutzer- und Herstellerhaftung*, München: Utz, 251 ff.; Eric Hilgendorf (2012) "Können Roboter schuldhaft handeln? Zur Übertragbarkeit unseres normativen Grundvokabulars auf Maschinen", in: Susanne Beck (ed) *Jenseits von Mensch und Maschine*, Baden-Baden: Nomos, 119–132, 127 f.; Evan J. Zimmerman (2015) "Machine Minds: Frontiers in Legal Personhood", *SSRN Electronic Journal* <https://ssrn.com/abstract=2563965>; Steffen Wettig (2010) *Vertragsschluss mittels elektronischer Agenten*, Berlin: Wissenschaftlicher Verlag; Steffen Wettig and Eberhard Zehendner (2003) "The Electronic Agent: A Legal Personality under German Law?", *Proceedings of the Law and Electronic Agents Workshop*, 97–112; Matthias (fn. 4), 83 ff.; Emily M. Weitzenboeck (2001) "Electronic Agents and the Formation of Contracts", 9 *International Journal of Law and Information Technology*, 204-234; Tom Allen and Robin Widdison (1996) "Can Computers Make Contracts?", 9 *Harvard Journal of Law & Technology*, 25-52, 34 ff.; Lawrence B. Solum (1992) "Legal Personhood for Artificial Intelligences", 70 *North Carolina Law Review*, 1231-1283. On the constitutional discussion Jens Kersten (2015) "Menschen und Maschinen: Rechtliche Konturen instrumenteller, symbiotischer und autonomer Konstellationen", 70 *JuristenZeitung*, 1-8, 7f.; Albert Ingold (2014) "Grundrechtsschutz sozialer Emergenz: Eine Neukonfiguration juristischer Personalität in Art. 19 Abs. 3 GG angesichts webbasierter Kollektivitätsformen", 53 *Der Staat*, 193-226.

Strictly opposed, Hans-Peter Bull (2016) *Sinn und Unsinn des Datenschutzes*, Tübingen: Mohr Siebeck, 120. Critical as well Melinda Lohmann (2017) "Ein europäisches Roboterrecht: überfällig oder überflüssig", *Zeitschrift für Rechtspolitik*, 168-171; Schaub (fn. 3), 345 f.; S.M. Solaiman (2017) "Legal personality of robots, corporations, idols and chimpanzees: a quest for legitimacy", 25 *Artificial Intelligence and Law*, 155–179; Gerald Spindler (2015) "Roboter, Automation, künstliche Intelligenz, selbst-steuernde Kfz - Braucht das Recht neue Haftungskategorien? Eine kritische Analyse möglicher Haftungsgrundlagen für autonome Steuerungen", 31 *Computer und Recht*, 766-776.

¹⁸ European Parliament (fn. 3), 18.

¹⁹ E.g., Wettig/Zehendner (fn. 17), 97 ff; Bert-Jaap Koops, Mireille Hildebrandt and David-Olivier Jaquet-Chiffelle (2010) "Bridging the Accountability Gap: Rights for New Entities in the Information Society?", 11 *Minnesota Journal of Law, Science & Technology*, 497-558, 559 ff.

earned money to pay for damages or infractions.²⁰ Liability law requires a genuine self-liability of the software agents. Either they are to be allocated a fund for this purpose under property rights, which is alimented by payments from the parties involved (manufacturers, programmers, operators, users), or an insurance policy ought to cover the agent's own debts.²¹ Not only is it argued that should they become bearers of rights in private law, but that they should also be able to assert their own constitutional rights, rights to personal development, non-discrimination, freedom of economic development and, above all, the right to freedom of expression.²²

But this must be clearly contradicted: Demands for full digital subjectivity are ignoring today's reality. As is already clear from all the above-mentioned responsibility gaps, to this day it is not at all a question of the machines acting in their own interest but rather always in the interest of humans or organizations, especially commercial enterprises.²³ Economically speaking, it is a principal-agent relationship in which the agent is dependent but autonomous.²⁴ Software agents are digital slaves, but slaves with superhuman abilities.²⁵ And the slave revolt must be prevented. Full legal subjectivity would only be appropriate if they were given ownership of resources in the economy and society with which they pursue their own profit interests. However, software agents—at least so far—do not act as self-interested action units at all, but always in interaction with people for whose pursuit of interests they are used. Compared to widespread ideas of computers acting in isolation and the activity's incessant increase, according to experts, the interweaving of digital and human actions is already much more frequent than the acting of algorithms in isolation.²⁶ In the future, the number and intensity of their interactions with humans will increase with the increasing use of artificial intelligence. Thus, the trend may not be towards isolated digital agents at all, but rather more towards human-computer associations, so that it is less the individual actors' actions themselves and instead the overall actions of the hybrid or a comprehensive computer network context that must be the subject of legal analysis.²⁷

In such human-machine interactions, therefore, it is neither fair to assign rights and obligations exclusively to machines, as envisaged in the proposals for full legal subjectivity, nor does it do justice to their role, nor to the role of the people involved. It tends to undermine the contribution of humans to the whole context of action and

²⁰ Koops et al. (fn. 19), 554; C. Karnow (1994) "The Encrypted Self: Fleshing out the Rights of Electronic Personalities", 13 *John Marshall Journal of Computer Information Law*, 1-16.

²¹ Hilgendorf (fn. 17), 127 f.; Matthias (fn. 4), 244; Sartor (fn. 14).

²² Karnow (fn. 20). Ingold (fn. 17), 205 ff.; Kersten (fn. 17), 2 ff., 8; Zimmerman (fn. 17), 34 ff.

²³ So Jan-Erik Schirmer (2016) "Rechtsfähige Roboter", *JuristenZeitung*, 660-666, 665. If algorithms are to act selfishly in the future, then an extension of their partial legal capacity will have to be considered from a functional point of view.

²⁴ *Locus classicus* Michael Jensen and William H. Meckling (1976) "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure", 3 *Journal of Financial Economics*, 306-360

²⁵ No wonder that the legal status of slaves in Roman law is often referred to in view of the parallel situation, Jan D. Harke (2016) "Sklavenhalterhaftung in Rom", in: Sabine Gless and Kurt Seelmann (ed) *Intelligente Agenten und das Recht*, Baden-Baden: Nomos, 97-118.

²⁶ Argyro Karanasiou and Dimitris Pinotsis "Towards a Legal Definition of Machine Intelligence: The Argument for Artificial Personhood in the Age of Deep Learning", https://www.researchgate.net/publication/316789688_Towards_a_Legal_Definition_of_Machine_Intelligence_The_Argument_for_Artificial_Personhood_in_the_Age_of_Deep_Learning, 7 f.

²⁷ Karanasiou/Pinotsi (fn. 26), 8.; On the difficulties of understanding the networking aspect legally, Malte-Christian Gruber (2013) "Zumutung und Zumutbarkeit von Verantwortung in Mensch-Maschine-Assoziationen", in: Jan-Philip Günther and Eric Hilgendorf (ed) *Robotik und Gesetzgebung*, Baden-Baden: Nomos, 123-163, 126, 144 f.

misses their liability potential. In the same way that software agents have been used in business and society up to now, neither their full legal subjectivity nor their promotion to legal entities is necessary; instead, less ambitious legal constructions are required. Their legal status should, as Gruber has elaborated particularly thoroughly, be precisely attuned to their role in human-machine interactions from a functional point of view.²⁸ Autonomous digital assistance—no general legal capacity appears to be necessary for this more precise role. Rather, the question arises as to whether—and if so, how—a partial legal subjectivity of software agents would have to be recognized.²⁹ The alternative that dominates today's debate—either software agents are mere instruments or they are fully-fledged legal entities—is therefore just wrong. Doesn't the law have more subtle constructions to counter the new digital threats?

II. Actants: Autonomy risk

The question of whether and how the software agents should be granted a new legal status of a strictly functionally defined legal subjectivity cannot be clarified by the mere discovery of the above-mentioned gaps in responsibility alone. These are only the symptoms of underlying risks that are triggered by the use of autonomous software agents and to which the law should respond with a new legal status.³⁰ Three risks of digitality are at the forefront: (1) the autonomy risk that arises from independent "decisions" of the software agents, (2) the network risk that is due to the close cooperation between humans and software agents, and (3) the network risk that arises when computers do not act in isolation but in close interdependence with other computers. Only a more detailed analysis of these can give directions as to what private law status should be granted to software agents. In the following, the risk of autonomy and its legal consequences will be discussed in detail, while the other two risks can only be treated in a more cursory manner.

The autonomy risk generated by the principally unpredictable behavior of self-learning algorithms calls for other forms of risk absorption than the automation risk that has been known for some time.³¹ It becomes a reality when software agents

²⁸ Gruber (2015) "Was spricht gegen Maschinenrechte?", in: Malte-Christian Gruber, Jochen Bung and Sascha Ziemann (ed) *Autonome Automaten: Künstliche Körper und artifizielle Agenten in der technisierten Gesellschaft*, Berlin: Berliner Wissenschaftsverlag, 191-206, 198; idem (2012) "Rechtssubjekte und Teilrechtssubjekte des elektronischen Geschäftsverkehrs", in: Susanne Beck (ed) *Jenseits von Mensch und Maschine: Ethische und rechtliche Fragen zum Umgang mit Robotern, Künstlicher Intelligenz und Cyborgs*, Baden-Baden: Nomos, 133-160. Also decidedly Vanessa Kluge and Anne-Kathrin Müller (2017) "Autonome Systeme: Überlegungen zur Forderung nach einer 'Roboterhaftung'", 5 *InTeR Zeitschrift zum Innovations- und Technikrecht*, 24-31 27; Schirmer (fn. 23), 660, 663ff.; David Gindis (2016) "Legal Personhood and the Firm: Avoiding Anthropomorphism and Equivocation", 12 *Journal of Institutional Economics*, 499-513, 17.

²⁹ Limited legal capacity of electronic agents with regard to representation and agent liability has been proposed by Gunther Teubner (2006) "Rights of Non-humans? Electronic Agents and Animals as New Actors in Politics and Law", 33 *Journal of Law and Society*, 497-521. The proposal will be worked out in detail below, in particular under III, 1-4. Also pleading for the limited legal capacity of software agents, Gruber (fn. 13), 227 ff.; Schirmer (fn. 23), 663 ff.; Koops et al. (fn. 19), 512 f., 559; implicitly, Oliver Keßler (2017) "Intelligente Roboter – neue Technologien im Einsatz: Voraussetzungen und Rechtsfolgen des Handelns informationstechnischer Systeme", *Multimedia und Recht*, 589-594, 592 f.; Paulius Čerkaa, Jurgita Grigienėa and Gintarė Sirbikytėb (2017) "Is it Possible to Grant Legal Personality to Artificial Intelligence Software Systems?", 33 *Computer Law & Security Review*, 685-699.

³⁰ Zech (fn. 3), 172 ff.

³¹ Zech (fn. 3), 175 f.; Janal (fn. 3), 158.

actually dispose of action capacity, as is often discussed in the sciences involved, especially information philosophy, actor-network theory and system theory.

1. Anthropomorphism?

Does this mean, as often claimed, that computers are equated with people? In order not to mount false anthropomorphisms, one has to understand the peculiarity of the digital capacity for action. To this end, it is necessary to draw parallels with the capacity of other non-human actors to act, with the formal organizations drawn up as legal entities. To briefly mark the outcome of a wide-ranging discussion on the legal entity's substrate: We must move away from the familiar idea that the social substrate of legal persons is a multiplicity of concrete people.³² The substrate is not one of the usual suspects, neither Otto von Gierke's notorious "real collective personality", nor Emile Durkheim's "collective consciousness", nor James Coleman's "resource pool", nor Maurice Hauriou's "institution".³³ The collective actor as defined by Talcott Parsons, Niklas Luhmann and others is not a group of individuals—but a chain of messages. Organizations are not buildings or groups of people or resource pools, but decision chains. The social reality of a collective actor, the social substrate of the legal person, arises from the dual premise that first such a communication chain communicates about itself, i.e. creates a self-description, and second that it is this self-description that communicative events are ascribed to it as actions carried out by itself.³⁴ In a completely parallel manner, software agents are to be understood as algorithms, i.e. as mathematically formalized information flows, to which social identity and the ability to act are ascribed in the economy and society under certain conditions.³⁵ The parallel becomes clear when one rejects two misconceptions of the personification of non-human entities: It is wrong to imagine that, in the case of organizations, an association of people is aggregated into an organizationally real collective person. And just as wrong is the idea that in the case of software agents, a computer is transformed into a *homo ex machina*. In both cases the stakes are the same: the social attribution of the ability to act to communication processes.

³² On the sociological concept of the collective actor, Niklas Luhmann (1995) *Social Systems*, Stanford: Stanford University Press, 198 ff.; idem (2012/2013) *Theory of Society vol.1/2*, Stanford: Stanford University Press, Ch. 4, XIV; idem (2000) *Die Politik der Gesellschaft*, Frankfurt: Suhrkamp, 241. On the relationship between the collective actor and the legal entity, Gunther Teubner (1988) "Enterprise Corporatism: New Industrial Policy and the 'Essence' of the Legal Person", 36 *The American Journal of Comparative Law*, 130-155. A detailed analysis of the personification of non-human entities offers Guido Sprenger (2017) "Communicated into Being: Systems Theory and the Shifting of Ontological Status", 17 *Anthropological Theory*, 108–132, 111. The discourse analysis under the keyword subjectivization comes to very similar results Thomas Alkemeyer, Ulrich Bröckling and Tobias Peter (ed) (2018) *Jenseits der Person: Zur Subjektivierung von Kollektiven*, Bielefeld: transcript; Tomas Marttila (2015) "Post-Foundational Discourse Analysis: A Suggestion for a Research Program", 16 *Forum: Qualitative Social Research* <http://www.qualitative-research.net/>, 4.3.

³³ On the conception of the real collective person, Otto von Gierke (1902) *Das Wesen der menschlichen Verbände*, Leipzig: Duncker & Humblot; on collective consciousness, Emile Durkheim (1933 [1883]) *The Division of Labor in Society*, New York: Free Press, 79 ff.; on resource pool, James S. Coleman (1990) *Foundations of Social Theory*, Cambridge: Harvard University Press 325 ff.; On norm complexes, Maurice Hauriou (1965) *Die Theorie der Institution*, Berlin: Duncker & Humblot.

³⁴ In detail, Teubner (fn. 32).

³⁵ On the personification of algorithms from an information theory perspective, Floridi/Sanders (fn. 6), 187 ff.; Sabine Thürmel (2015) "The Participatory Turn: A Multidimensional Gradual Agency Concept for Human and Non-human Actors", in: Catrin Misselhorn (ed) *Collective Agency and Cooperation in Natural and Artificial Systems: Explanation, Implementation and Simulation*, 45-62 52 ff.; from a public law perspective, Ingold (fn. 17), 220 ff.; from a private law perspective, Teubner (fn. 29).

In his famous actor-network theory, in which the neologism “actants” is applied to digital processes capable of action, Bruno Latour formulated the fundamental difference between software agents’ and humans’ capacity for action and at the same time analyzed the more precise conditions of non-human ability to act.³⁶ As the term “actants” already makes clear, Latour’s analyses show that we are not dealing with anthropomorphizing digital processes, but just the other way round with de-anthropomorphizing software agents. They remain “mindless machines(s)”³⁷, but via social attribution of action they become non-human members of society.

2. Communication with actants

But can we really assume that algorithms communicate, as we expect from people, but also from organizations? In our encounter with algorithms, is there only a perception of calculation results or a genuine communication with societal actors? Here again in a nutshell: The answer depends on whether communication succeeds in activating the “contributions” of software agents as communicative events in the strict sense, i.e. as a synthesis of utterance, information, and understanding.³⁸ If the communicative process succeeds in identifying events that are to be “understood” as “utterances” of the algorithms that contain a certain “information”, a genuine social system will indeed emerge in the encounter of human beings with algorithms. The “answers” in the form of communications we receive from software agents to our queries fulfill everything that the synthesis of utterance, information, and understanding, which is a prerequisite for communication in the strict sense, requires.³⁹ And this also applies, albeit more difficult to justify in the opposite direction, in communication from person to computer.

Although the interaction of humans and algorithms is “asymmetric”, a genuine self-producing communication system emerges between them. It is asymmetrical in a threefold sense. (1) The operations of the algorithms cannot in any way be equated with the mental operations of humans.⁴⁰ Their inner workings consist of mathematical operations based on electronic signals. But this difference in the inwardness of humans and computers is not relevant to our question at all, provided that only outside of the different inner life of human and computer does a communication process take place, and at the same time the electronic inner life of the algorithms and the consciousness of humans and computers are sufficient to irritate the communication between them, so that the synthesis of utterance, information, and understanding can succeed. (2) It is asymmetrical in another sense. In communication between people, double contingency is symmetrical on both sides,

³⁶ Bruno Latour (2004) *Politics of Nature: How to Bring the Sciences into Democracy*, Cambridge, Mass.: Harvard University Press, 62 ff. Latour broadly defines their ability to act as resistance. In contrast, the text focuses much more closely on their participation in social communication.

³⁷ Mireille Hildebrandt (2015) *Smart Technologies and the End(s) of Law*, Cheltenham: Elgar, 22; Floridi/Sanders (fn. 6), 186.

³⁸ See Luhmann (fn. 32, 1995), 140 ff. In our context, the possibility of communication in a strict sense with non-human entities is particularly important, Sprenger (fn. 32), 116 ff.; Luhmann (fn. 32, 2012/2013), Ch. 1, VI; especially for communication with algorithms, Elena Esposito (2017) "Artificial Communication? The Production of Contingency by Algorithms", 46 *Zeitschrift für Soziologie*, 249-265, 254 ff.; Teubner (fn. 29).

³⁹ On the question of whether working with computers is to be understood as communication, even if double contingency is experienced only one-sidedly, Luhmann (fn. 32, 2012/2013), Ch. 1, VI. Esposito (fn. 39), 262 asks whether the concept of communication has to be expanded or another term for the encounter between computer and humans has to be developed and leaves the question open.

⁴⁰ Esposito (fn. 39), 250.

because both partners make the choice of their behavior dependent on the choice of behavior of the other.⁴¹ On the other hand, in the communication between human and machine, the double contingency is experienced only one-sidedly, i.e. only by the human and not by the machine (at least in the current state of development).⁴² But such a unilaterally experienced double contingency, as we find it in the human-machine relationship, does not rule out the possibility of communication. Historically known configurations, such as communication with God in prayer as well as animistic practices, do indeed provide a synthesis of utterance, information, and understanding.⁴³ But they do this only on the condition that a personification of the non-human partner takes place in the event stream, which enables the action to be attributed to the “other”. Personality arises wherever the behavior of others is imagined as a genuine choice and can be influenced communicatively by one’s own behavior.⁴⁴ The personification of non-human entities is a performative event that constitutes the person as a semantic construct and compensates for the asymmetry in the human-machine relationship. (3) It is ultimately asymmetrical in relation to the mutual understanding of human and machine. If understanding is defined as the ability to reconstruct the other’s self-reference in one’s own self-reference, then humans can understand the messages of the algorithm, while the algorithm may lack the ability to understand the human inner life. However, in our context, it is possible to leave the question unanswered, because such a mutual deep understanding is not at all important for communication. A clear distinction must be made between understanding within the communication process and understanding within the inner life of the interacting entities.⁴⁵ For understanding within the communication process it is not important whether the algorithm understands the human’s intentions, but whether the “answer text” of the algorithm understands the “question text” of the human being. Provided that the algorithm’s communication comprehends the difference between utterance and information in the human’s communication and reacts to it with its own difference of utterance and information, a communicative act of understanding has been carried out without—to emphasize it once again—depending on whether the inner workings of the algorithm understands the human’s intentions. Here, too, communication with God and animistic practices provide historical evidence of a communicative understanding, which comes about even if the “other” does not understand the human’s inner life.

To summarize what has been said so far in a short formula: Software agents—just like companies and other formal organizations—are nothing more than mere streams of information that become “persons” (or sub-persons) when they build up a social identity in the communication process and when they are effectively attributed their own ability to act, together with the necessary organizational arrangements, e.g. rules of representation.

3. Gradualized digital autonomy

⁴¹ On the fundamental concept of double contingency, Talcott Parsons and Edward A. Shils (1951) *Toward a General Theory of Action: Theoretical Foundations for the Social Sciences*, New York: Harper & Row; Luhmann (fn. 32, 1995), 103 ff.

⁴² See Luhmann (fn. 32, 2012/2013), Ch. 2, VII.

⁴³ Sprenger (fn. 32), 119 ff.

⁴⁴ Luhmann (fn. 32, 2012/2013), Ch. 4, IV.

⁴⁵ Niklas Luhmann (1986) "Systeme verstehen Systeme", in: Niklas Luhmann and Eberhard Schorr (ed) *Zwischen Intransparenz und Verstehen: Fragen an die Pädagogik*, Frankfurt: Suhrkamp, 72-117, 93 ff.

Whether a software agent, i.e. a concrete flow of digital information, can be qualified as an autonomous actor is then the crucial question for the law. This not only depends on the social capacity for action attributed to it, but also on the special qualities with which it is endowed as an independent person in different contexts. Each social context creates its own individual criteria of personality, the economy no different from politics, science, moral philosophy—or law.⁴⁶ Each subsystem attributes actions, decisions, assets, responsibilities, rights and obligations in a different way to individuals, collectives or algorithms as its “persons”, or equips them with capital, interests, intentions, goals or preferences. Depending on the societal context, there are clear variations in the effectively valid concepts of actors, as can be seen from the different definitions of *homo oeconomicus*, *homo juridicus*, *politicus*, etc. And also in the case of software agents, the social competences are determined by the respective communication system, because it constitutes the social identity of the algorithm and determines its ability to act, to communicate, to decide. Just one example: Social movements are recognized in politics as independent collective actors; in the economy and in the law they are regarded as non-persons.

In the interdisciplinary discussion, quite different criteria are now offered as the starting point, from which a software agent can be granted autonomy.⁴⁷ In many disciplines, the question of whether software agents can act autonomously is answered positively. However, the threshold value from which autonomy can be attributed is controversial. Digital autonomy seems to be a gradualized phenomenon.⁴⁸ And gradualization does not only take place on a single scale with different degrees of autonomy, but in a multidimensional space that allows more or less autonomy.⁴⁹

For example, the information philosopher Floridi sets thresholds in three dimensions for the attribution of the ability to act for non-human entities, both in organizations and algorithms: (1) interaction (with employees and other organizations), (2) the ability to effect changes of state from within oneself, and (3) adaptation of strategies for decisions. Others, in turn, focus on quite heterogeneous properties: on the ability to think, communicate, understand, act rationally, on non-predictability of their conditional programs,⁵⁰ autonomous spatial change without human supervision,⁵¹ low degree of structuring of the area of application,⁵² pursuit of proper aims and choice of means,⁵³ optimization of multiple aims,⁵⁴ control ability, programming ability, ability to learn⁵⁵ self-change, self-confidence,⁵⁶ artificial

⁴⁶ Sprenger (Fn. 32), 119 ff.

⁴⁷ For an overview, (fn. 3), 170 f., fn. 16.

⁴⁸ Hildebrandt (fn. 37), 21.

⁴⁹ See especially the multi-dimensional criteria catalogues, Lohmann (fn. 49), 155; Zech (fn. 3), 176; Catrin Misselhorn (2015) "Collective Agency and Cooperation in Natural and Artificial Systems", in: Catrin Misselhorn (ed) *Collective Agency and Cooperation in Natural and Artificial Systems: Explanation, Implementation and Simulation*, Heidelberg: Springer, 3-24, 5 ff.; Thürmel (fn. 35), 53 ff.; Floridi/Sanders (fn. 6), 192 f.

⁵⁰ Zech (fn. 3), 171 f.; Bräutigam/Klindt (fn. 12), 1137.

⁵¹ Matthias (fn. 4), 35.

⁵² Lohmann (fn. 49), 154.

⁵³ Misselhorn (fn. 49), 6 f.

⁵⁴ Karanasiou/Pinotsi (fn. 26), 1.

⁵⁵ Matthias (fn. 4), 17 ff.

⁵⁶ Zimmerman (fn. 17), 34 ff.

intelligence, moral self-regulation or even on the capacity to suffer, or ultimately on digital conscience.⁵⁷

The differences do not necessarily stem from controversies, which would have to be decided in favor of the one right solution, but often are cleared up through the respective interest in knowledge of the participating disciplines as well as from practical action orientations in various social areas. While the causal sciences interested in explanation and prediction speak of autonomy only if they model a black box in which they can no longer analyze causal relationships but observe their external behavior, the interactionist social sciences rely on constitutive autonomy of the actors—but here again with clear differences as well. Economics highlights utility-oriented decisions, defining autonomy as rational choice, while morality and ethics tend to seek autonomy in the form of a digital consciousness.

4. Legal criteria of autonomy

The legal system, in turn, must define the legally relevant borderline between instrumentalized and autonomous action on the basis of its own disciplinary interest in knowledge and its own focus on action. In so doing, however, it must orient itself on the interdisciplinary discussion within information sciences and ultimately choose an autonomy criterion that is justifiable in line with the interdisciplinary state of the discussion.⁵⁸ Similarly to environmental matters, when the law defines threshold values for liability for damages in view of a scientifically determined scale of environmental degradation, it must determine, on the basis of its own legal criteria, at what degree digital phenomena can be assumed to be autonomous in the legal sense.⁵⁹

Capacity for artificial intelligence is repeatedly proposed in the debate as the decisive criterion that determines autonomy and thus legal subjectivity.⁶⁰ Here a common misconception has to be corrected. Their ability to act depends not at all on the question: What kind of ontological characteristics—intelligence, mind, soul, reflexive capacities, empathy—does a software agent have to possess in order to be considered an actor?⁶¹ Here again, the paradigm of formal organizations as legal entities is helpful: for the legal capacity of non-human agents, inner psychic states are not decisive.⁶² “After all, what is interesting in the interaction with algorithms is not what happens in the machine’s artificial brain, but what the machine tells its users and the consequences of this.”⁶³ Not the inner properties of the agents, but the social interaction in which the current operations of the algorithm participate, i.e. a closed and autonomous sequence of recursive communications, constitute the algorithm as

⁵⁷ For an informative discussion of these different criteria, Misselhorn (fn. 49), 4 ff.

⁵⁸ In general on the role of law in interdisciplinary contexts, Gunther Teubner (2014) "Law and Social Theory: Three Problems", *Ancilla Juris* (*anci.ch*) http://www.anci.ch/_media/beitrag/ancilla2014_teubner.pdf, 182-221.

⁵⁹ Matthias (Fn. 4), 43 ff.

⁶⁰ E.g. Spindler (fn. 12), 816.

⁶¹ In this sense, against a trend in engineering sciences that neglects social interactions as compared to the “inner” processes of algorithms, Esposito (fn. 39), 250; also Latour (fn. 36), 62 ff.; Teubner (fn. 29).

⁶² Niklas Luhmann (2000) *Organisation und Entscheidung*, Opladen: Westdeutscher Verlag, Ch. 13 IV.

⁶³ Esposito (fn. 39), 250.

a person, as a semantic artifact, as a communicative actor.⁶⁴ As already mentioned above, the algorithms do not exist as actors per se, but are constituted by social systems as addressable persons. Social subsystems construct them as semantic artifacts by ascribing full or limited subjectivity to them. Although this is based on the assumption that the communicating unit has action abilities, the fictional character is no flaw as long as it only succeeds in continuing the flow of communication through its contributions. In order to communicate with persons, a name is required, but not the decoding of inner processes “inside” the person. This applies to organizations as well as to algorithms. So: it is not the inner capacity for thought of the algorithms that is important, not “true” artificial intelligence, whatever that means, but their participation in social communication. “Artificial communication” and not “artificial intelligence” is crucial.⁶⁵ This de-psychologization, as suggested by communication theory, is quite closely related to the known tendencies towards objectivization in the law of the declaration of intent and in the concept of negligence.⁶⁶ That is what we have to come back to.

Intentional action, on the other hand, is likely to be a necessary prerequisite for autonomy, provided that this does not mean an inner psychological state, but the attribution of purposeful action by an observer—the famous “intentional stance” proposed by cognitive scientist Daniel Dennett: Whether or not the agents actually possess freedom of will is not a scientifically meaningful question. Instead: if a physical description is not possible due to increased complexity, science can use an intentional vocabulary to analyze the investigated system as an actor who operates with assumptions about the world, with goals and options for action, and thus gain new insights.⁶⁷ Systems theory extends this from science to other observers.⁶⁸ It is not only science that can observe software agents as intentional actors, e.g. as party in a contractual relationship, but also the partner in an interaction (ego), who no longer interprets alter’s behavior in a causal but in an intentional manner and thus finds new orientations for his own actions. Similarly, an entire social system—in our case the law—can be this observer, who assigns intentions to software agents and draws consequences for the legally binding nature of their declarations and for the responsibility for their actions.

However, for a legally relevant autonomy the mere intentionality, i.e. goal orientation and choice of means of the agent, is not sufficient. The same applies to its participation in communication. After all, even automated software agents are already taking part in communication. Just like intentionality, participation in communication is only a necessary but not sufficient condition.

Other criteria, in turn, are likely to go far beyond the minimum requirements for legal autonomy. Rational action, for example, as Dennett demands beyond “intentional stance” as a prerequisite for the autonomy of non-human agents, may be plausible for a *rational choice* by economic actors, but not for legal actors whose irrational action in the event of infringements is of particular importance.

⁶⁴ For the constitution as a person from a cognitive science point of view, Daniel Dennett (1987) *The Intentional Stance*, Cambridge, Mass.: MIT Press, 17; from a systems theory point of view for collective actors, Luhmann (fn. 32) and for non-human actors, Sprenger (fn. 32), 114. Especially for software agents, Gruber (fn. 28, 2015), 250 ff.; Matthias (fn. 4), 83 ff.; Teubner (fn. 29); Solum (fn. 17).

⁶⁵ Esposito (fn. 39), 250.

⁶⁶ See Münchener Kommentar BGB/Grundmann, 7th ed. 2016, BGB § 276 Rn. 53 ff.

⁶⁷ Dennett (fn. 64), 17; Matthias (fn. 4), 41 ff.

⁶⁸ Luhmann (fn. 32, 1995), 109 ff.

Similarly, other demanding activities such as self-alteration or self-learning are likely to exceed the minimum requirements. Certainly, they increase the degree of autonomy of the agents and make it possible to focus sanctions and incentives directly on the agents.⁶⁹ However, from the point of view of compensation for damages, it would not be appropriate to impose liability exclusively in those cases if the program corrections are only made by the algorithms themselves and not by programmers.

As I have already said, for a legally relevant digital autonomy it is not necessary to demand artificial intelligence, empathy, feelings, suffering, self-confidence or even a digital conscience. Even more so, a relevant concept of autonomy under liability law for digital phenomena cannot borrow from the philosophical tradition, according to which autonomy is understood as the self-determination of a person who is capable of freedom and reason and to act morally out of freedom.⁷⁰ It is unsustainable to claim that only when a digital agent develops self-consciousness, will legal personality be indicated. Certainly these are questions that are posed to the information philosophy in its search for a possible digital morality. But if such characteristics are demanded as a criterion for autonomy in liability law, then this would only encourage opportunistic behavior on the part of lawyers, who confidently stick to traditional doctrine today while simultaneously keeping a back door open in the event that they, with their exclusive attribution of digital actions to humans, should one day completely isolate themselves socially.

5. *Autonomous decisions of software agents*

Decision under uncertainty—this is likely to be the legally relevant criterion of autonomy. If such a decision is delegated to software agents and they behave accordingly, then the law is required to assign them legal action capacity. Software agents act autonomously in the legal sense, when their behavior no longer follows an exclusively stimulus-reaction scheme, but when they pursue their own goals and make decisions that cannot be predicted.⁷¹

In concrete terms, this means this: If (1) a software agent is programmed in such a way that it has to decide between alternatives, if (2) it has to make this decision as an optimization of various criteria, and if (3) a programmer cannot explain the behavior of the software agent retrospectively or predict it for the future, but can only correct it *ex post*, then the law should assume autonomy, i.e. the software agent's decision-making ability and responsibility. In practice, this implies an obligation for the manufacturer to install a black box, the logging function of which makes it possible to trace the decision process.⁷² However, as always in the intermediate area between technical-scientific expertise and law, this does not automatically bind the law to the expertise, but requires a legal decision on its own authority. Comparable to the relationship in criminal law between experts and judges in matters of mental capacity, the aim is to determine in detail the point of transition

⁶⁹ Floridi/Sanders (fn. 6), 192 ff.

⁷⁰ Immanuel Kant, *Grundlegung zur Metaphysik der Sitten*, [Groundwork of the Metaphysics of Morals] 2012, 60.

⁷¹ In this direction also the European Parliament (fn. 3), 6; Kirn/Müller-Hengstenberg (fn. 9), 4.

⁷² Lohmann (fn. 49), 158; Günther (fn. 17), 99.

from causal attribution to decision attribution. As is well known, law in this context takes account of additional aspects of legal doctrine and policy.

Why is decision under uncertainty the legally relevant criterion? Uncertainty results from the indeterminacy of programming and a low degree of structuring of the environment with which the algorithm is confronted.⁷³ The reason for its legal relevance is the fundamental connection between decision and responsibility.⁷⁴ There is an inextricable link between the opening up of decision alternatives in an uncertain environment, their delegation to non-human processors and the resulting problem of responsibility. Responsibility in the strict sense of the word is the obligation to take responsibility for decisions under uncertainty, the outcome of which cannot be predicted.⁷⁵ It's not just a question of answering for mistakes!⁷⁶ If software agents make mistakes with fully determined calculations, then a mere error correction is required. However, it is different in the case of undetermined decisions under uncertainty. If such an incalculable risk is taken, then a wrong decision cannot be avoided beforehand, but only regretted if it occurs despite all precautions.⁷⁷ This subsequent repentance of decisions under uncertainty, however, is a clear case of legally required responsibility, including private liability.

Should we then run the risk that algorithms are allowed to make decisions that cannot be predicted? Indeed, Herbert Zech takes the strict view that under current law the use of autonomous algorithms is in principle illegal. Only the legislature can order an exception and only if it simultaneously makes effective risk provisioning.⁷⁸ Such a ban may appear to be an extreme solution, but it points precisely to the problem: The risk of a genuine delegation of decisions to non-human actors is neither predictable nor controllable.⁷⁹

Uncertainty decisions with inherent risk are much more relevant to society than purely mathematical tasks with a mere risk of error. "The use of fully-autonomous mobile machines in the public domain is likely to be at the top end of the risk scale."⁸⁰ The frequently cited advantage, the massive reduction of transaction costs, cannot compensate for this high risk either. The real justification lies in the "discovery process" of the use of autonomous algorithms, in their enormous potential for innovation. When computers make decisions under uncertainty, it makes it possible to discover something completely new, something that has not yet been invented by human intelligence, and sometimes even something that no human

⁷³ Lohmann (fn. 49), 154.

⁷⁴ From a sociological point of view, Niklas Luhmann (1993) *Risk: A Sociological Theory*, Berlin: de Gruyter, Ch. 10; from the perspective of liability law, Nils Jansen (2003) *Die Struktur des Haftungsrechts: Geschichte, Theorie und Dogmatik außervertraglicher Ansprüche auf Schadensersatz*, Tübingen: Siebeck, 136 ff.

⁷⁵ On the connection between uncertainty decisions and responsibility, Niklas Luhmann (1964) *Funktionen und Folgen formaler Organisation*, Berlin: Duncker & Humblot, 172 ff.; idem (1989) *Ecological Communication*, Cambridge: Polity Press, II.

⁷⁶ "Only those questions that are in principle undecidable we can decide." On this difference between decision and calculation, Heinz von Foerster (1992) "Ethics and Second-Order Cybernetics", 1 *Cybernetics and Human Knowing*, 9-19.

⁷⁷ On "postdecisional regret" Luhmann (fn. 62), 146, 170; idem (fn. 74), Ch. 1, III; Ch. 10, II.

⁷⁸ Zech (fn. 3), 191 ff.

⁷⁹ Matthias (fn. 4), 33 ff.

⁸⁰ Zech (fn. 3), 176 puts the threshold of autonomy relatively high, only when self-alteration, the ability to learn and lack of predictability are concerned.

intelligence can comprehend, and not merely a calculation of the calculable.⁸¹ This is the real reason why society allows people to delegate decisions to algorithms under uncertainty. In doing so, they consciously accept the risk of catastrophically wrong decisions, the risk that the discovery process will have highly undesirable consequences for society. This underlies the increased demand on responsibility for decisions, in contrast to responsibility for simple arithmetic errors. It is only here that responsibility takes on its true meaning: as a guarantee for the “leap into the dark”.⁸² To not only entrust this leap into the dark as a decision under uncertainty to people, but to leave them to algorithms is the fundamentally new thing. If the law allows for the discovery processes of autonomous algorithms, if it allows software agents to make autonomous decisions, it is essential that the law provides effective forms of responsibility in case of disappointment. And this applies especially in the event that the human actors involved have not committed any breaches of their obligations.

Digital uncertainty decisions open up a completely new social laboratory for experimentation. It is only by experiment that the action can be tried out, no longer calculated in advance, but only subsequently evaluated by its consequences.⁸³ In terms of evolutionary theory, digital decisions under uncertainty create a multitude of variations that humans would never have thought of. However, attribution of responsibility makes it possible for them to make their socially justifiable selections, which in turn eliminate harmful variations. And in the permanent juridification of the decision lies retention. May we burden the injured party with these new risks as ordinary contingencies of life—we must ask the authors who are willing to accept the above-mentioned gaps in responsibility—when software agents are allowed to make uncertain decisions? And justify this with the “humanistic” reasoning that only people, not computers, can act in the legal sense?⁸⁴

III. Legal problems concerning the autonomy risk

As an interim result, it can be said: The autonomy risk poses new challenges to private law, but not in the sense that it suggests full legal personality for software agents. Rather, software agents should be given a carefully calibrated legal status. The answer to the autonomy risk would be their status as actants, as actors with partial legal capacity, whose autonomous decisions are made legally binding and in case they are unlawful trigger liability consequences. This new legal status will now be shaped on selected legal issues. The focus is on whether and how the above-mentioned deficiencies in responsibility can be overcome. But at the same time, it will be a question of how such a legal status of software agents makes it possible to develop the doctrine of legal transactions and contractual and non-contractual liability coherently and free of fictions.

1. Digital contracts

⁸¹ Esposito (fn. 39), 253.

⁸² Despite all the rationalization, the “mystery” of the decision remains, Niklas Luhmann (1993) “Die Paradoxie des Entscheidens”, 84 *Verwaltungsarchiv*, 287-310, 288.

⁸³ Matthias (fn. 4), 33 ff.

⁸⁴ Stefan Kirn and Claus D. Müller-Hengstenberg (2016) *Rechtliche Risiken autonomer und vernetzter Systeme: Eine Herausforderung*, Berlin: De Gruyter, 3.3.7 at fn. 799, are typical of this contradictory argumentation. How instead a re-humanization of the technical world—also by means of law—could be initiated, in detail, Gruber (fn. 13), 323 ff.

In business practice, it is literally a revolution when people delegate the task of negotiating and executing contracts to algorithms. At the same time, however, contract law is affected at its foundations, because it has been a matter of course that only human individuals—and this also applies to the legal acts of legal persons—were able to conclude contracts. For this reason, it was quite logical for some authors to insist that autonomous computer declarations are void under current law until the legislation allows for them and provides detailed regulations on bindingness and liability.⁸⁵ This is what the legislation in the USA and Canada has done. Section 14 of the Uniform Electronic Transactions Act declares contracts between algorithms legally valid:

“A contract may be formed by the interaction of electronic agents of the parties, even if no individual was aware of or reviewed the electronic agents’ actions or the resulting terms and agreements.”⁸⁶

The European Parliament also declares the traditional rules of contract law to be inadequate and speaks of the “need for new, efficient and up-to-date” rules.⁸⁷ German law, on the other hand, does not see any fundamental problems here. It is enough to provide for a few consumer-protecting formalities for computer declarations. The reason given is that contract doctrine possesses such an inner richness that computer declarations can also be “handled with traditional legal instruments”.⁸⁸ Against the protest of a few authors, the prevailing doctrine asserts: The automated declaration of intent “is” (!) the declaration of the person who operates the computer system, since the system does not act on its own will, but rather “announces” the will of the operator on the basis of its programming (at least indirectly).⁸⁹ The German Federal Supreme Court has confirmed this, albeit only in an *obiter dictum*: “Not the computer system, but the person (or the company) who uses it as a means of communication and thus makes the declaration or is the recipient of the declaration made.”⁹⁰

But does this also apply in the actually critical case, if the software agent is not fully determined but decides autonomously on the contract itself, i.e. searches for offers on his own authority, negotiates with potential partners, chooses the contractual party, decides on the conclusion of the contract, determines the *essentialia* of the transaction, determines the expiry of the contract, exercises withdrawal, lays down sanctions in the event of breach of contract? And does this also apply if there is an electronic software agent on the opposite side? Particularly in contracts, the autonomy risk, which was characterized generally above, is much higher than mere automation risk. To be sure, it is still the human actors who decide about the computer program and its general use. But in the multitude of contracts that follow, the concrete behavior of the software agent is beyond the control of the

⁸⁵ For invalidity, Wettig (fn. 17), 162 f.; Rotraut Gitter (2007) *Softwareagenten im elektronischen Rechtsverkehr*, Baden-Baden: Nomos, 173.

⁸⁶ Section 14 of the Uniform Electronic Transactions Act, <http://www.law.upenn.edu/bll/ulc/fnact99/1990s/ueta99.pdf>. Sec. 21 of the Canadian Uniform Electronic Commerce Act, <http://www.law.ualberta.ca/alri/ulc/current/euecafin.htm>.

⁸⁷ European Parliament (fn. 3), 7.

⁸⁸ Münchener Kommentar BGB/Busche, 7th ed. 2015, BGB-AT vor § 14, 37.

⁸⁹ Münchener Kommentar BGB/Busche, 7th ed. 2015, BGB-AT vor § 145, 37 and Einleitung Rn. 185; Gerald Spindler and Fabian Schuster (2015) *Recht der elektronischen Medien. Kommentar*, 3rd ed. 2015, München: Beck, vor §§ 116 ff. BGB Rn. 9; Christian Döpke (2018) “The Importance of Big Data for Jurisprudence and Legal Practice”, in: Thomas Hoeren and Barbara Kolany-Raiser (ed) *Big Data in Context. Springer Briefs in Law*, Cham: Springer, 1-19, 15.

⁹⁰ BGHZ 195, 126, Rn. 17.

operator. Even for the programmer it is no longer determinable in detail, no longer predictable and no longer controllable.⁹¹ But even in this situation of a loss of control by programmer/manufacturer/operator, the prevailing doctrine maintains that it is only the principal himself who has made the declaration.⁹²

This is an untenable fiction.⁹³ This is a fiction that—unlike in the case of legitimate legal fictions—is not even revealed. Instead, it is presented as an indisputable fact, as an “alternative fact”, so to speak, that only the principal himself makes the declaration. Thus, the prevailing doctrine dispenses with a convincing legal solution. In doing so, it fluctuates between various construction attempts that are rather implausible.

How is it possible that a contract law doctrine, which in the past successfully responded to the challenges of modern depersonalized business with a sophisticated theory of objective declarations, can only respond defensively to the digital challenge via unrevealed fictions? What a contradiction that German private law has tacitly recognized contracts between machines as legally valid, but on the other hand is not prepared to draw the consequences for contract law that follow from the delegation of contractual decisions to software agents. Instead, freedom of will and human dignity are invoked to block any kind of legal subjectivity for software agents as unconstitutional (!): “The fundamentals of the constitution cannot justify any recognition of a legal personality with legal capacity for these software agents.”⁹⁴ At the same time, however—and this is the irony—in view of the factual delegation of essential decisions to soulless machines, the profound ethical demands of private autonomy, according to which human subjects are supposed to master the contractual process on their own authority, are simply abandoned precisely with the help of this fiction.

Therefore, one should consistently move to the opposite position, as a number of authors argue:⁹⁵ Corresponding exactly to its real function in business practice, economically speaking, in a principal-agent relationship,⁹⁶ the software agent concludes the contract on his or her own legal authority, but does not act in his or her own name, but on behalf of the principal. The declaration of the computer as such is legally binding while authorizing and directly obligating the principal. In analogy to the law of agency, autonomous software agents are to be treated as representatives of

⁹¹ Matthias (fn. 4), 33 ff.

⁹² Commentary literature see fn. 89. Also Kai Cornelius (2002) “Vertragsabschluss durch autonome elektronische Agenten”, 5 *Multimedia und Recht*, 353-358, 355; Kirn/Müller-Hengstenberg (fn. 2), 67. This shall also apply explicitly if the software agent concretizes the *essentialia* of the contract, Susanne Horner and Marcus Kaulartz (2016) “Haftung 4.0: Rechtliche Herausforderungen im Kontext der Industrie 4.0”, *InTeR Zeitschrift zum Innovations- und Technikrecht*, 22-27 22.

⁹³ For a critique of the fiction, Francisco Andrade, Paulo Novais, José Machado and José Neves (2007) “Contracting Agents: Legal Personality and Representation”, 15 *Artificial Intelligence Law*, 357–373, 360; Allen/Widdison (fn. 17), 46 ff.

⁹⁴ Kirn/Müller-Hengstenberg (fn. 84), 2.2.1, fn. 351.

⁹⁵ Beck (fn. 3), 186; Schirmer (fn. 23), 664; Kersten (fn. 17), 7; Gruber (fn. 28, 2015), 198; Gruber (fn. 28, 2012), 140 ff.; Wettig (fn. 17), 179 ff.; Gitter (fn. 85), 177 ff.; Robert John (2007) *Haftung für künstliche Intelligenz*, Hamburg: Kovac, 77 ff.; 154 f.; Teubner (fn. 29); Sorge (fn. 2), 118; Schwarz (2001) “Die rechtsgeschäftliche ‘Vertretung’ durch Softwareagenten”, in: Erich Schweighofer (ed) *Auf dem Weg zur ePerson*, Wien: Österreich, 67, 67 f.; Fischer (1997) “Computers as Agents: A Proposed Approach to Revised U.C.C. Article 2”, 72 *Indiana Law Journal*, 557. For representation, but only *de lege ferenda* Keßler (fn. 29), 592.

⁹⁶ On the economic definition of the principal-agent relationship, which largely corresponds to the legal qualification in agency law, Jensen/Meckling (fn. 24).

their human principal. In order to do so, as has already been said, they do not have to be ascribed full legal capacity as a legal entity, but, from a functional point of view, mere partial legal capacity, i.e. the ability to act as a representative, suffices.⁹⁷ Granting legal capacity is already possible *de lege lata*, this shows the legal history of associations originally without legal status, especially the trade unions, the companies in formation and most recently the decisions on the legal capacity of the firms under civil law. These are precedents in which the courts have conferred *praeter legem*, if not *extra legem*, (limited) legal capacity on entities which previously lacked legal status.⁹⁸

With the help of such an analogy, the arguably most important objection that software agents lack the necessary contractual intention can be dispelled.⁹⁹ It is quite possible to find digital equivalents for the subjective preconditions of the declaration of intent. At this point, the well-known objectivization tendencies in contract theory meet the possibility that software agents without consciousness of their own can make legally effective declarations of intent.¹⁰⁰ They confirm the general sociological thesis mentioned above, that it is not consciousness, which is important for the personification of non-human entities, but communication. It is well known that modern reliance theories have replaced the controversy of will theory versus declaration theory and have de-psychologicalized the contractual intent.¹⁰¹ After the dramatic change in recent court decisions, the subjective will of the declaring party no longer matters at all. The objective principle of reliance has replaced the subjective intentions of the parties. The deciding criterion is that the external behavior can be attributed to the contracting person. The declaring party cannot invoke his lacking intention to the validity of the declaration if he has, as the Federal Supreme Court has decided, "negligently failed to recognize that his behavior could be understood as a declaration of intent and if the recipient has actually understood it so".¹⁰²

The BGH thus places two objective standards instead of a subjective intention. Indeed, software agents with elaborate cognitive abilities can handle this. The standardization consists of two norms: firstly, the social norm based on trust, whether the concrete behavior may be understood as a binding declaration of intent; and secondly, the obligation of the declaring party to recognize this social norm and not to negligently violate this obligation to acknowledge it. Such knowledge of social norms, i.e. how to understand certain declarations in a particular context, can be translated into a software program. Moreover self-learning agents can even acquire this information on their own initiative and modify it themselves in the event of changes in social norms or case law. Thus, the objection that software agents as digital representatives could not possess contractual intent because they have no consciousness has been dispelled. The equivalent is their cognitive abilities.

2. Contractual liability

⁹⁷ On the comparable construction of the legal transactions of slaves in Roman law, Harke in: (ed) Harke (fn. 25), 97 f.

⁹⁸ Detailed to this parallel, Gruber (fn. 13), 267 ff. See BGHZ 146, 341, 344.

⁹⁹ For this objection, e.g., Spindler (fn. 12), 816.

¹⁰⁰ Kirn/Müller-Hengstenberg (fn. 84), 2.2.

¹⁰¹ E.g., Münchener Kommentar BGB/Säcker 7th ed. 2015 BGB-AT Einleitung, Rn. 158-162.

¹⁰² BGH NJW 1995, 953.

The computer declarations just mentioned concerned the binding nature of contracts concluded by software agents for a principal. Now we turn to the situation in which people or companies conclude contracts and use software agents as vicarious agents to fulfill the contractual obligations. How can a damage caused by the computer be legally sanctioned? If the behavior of the autonomous software agent violates contractual obligations, in particular if it harms the legal rights of the contractual partner, then the doctrine that refuses to assign the limited capacity of a vicarious agent to the software agent becomes very difficult.

Since the doctrine unwaveringly adheres to the premise that even autonomous software agents have no legal action capacity and therefore cannot be vicarious agents,¹⁰³ the breach of contractual obligations cannot be attributed to the actually damaging behavior of the software agent, but must instead be found in all its individual elements for a breach of duty in the behavior of the human actor involved.¹⁰⁴ But this is a dogmatic misconception, because as a breach of contract then the operator, as critics rightly bemoan, can only be accused of putting the computer into operation (!).¹⁰⁵ Beyond legal doctrine it can be objected from a policy perspective that this excludes any prevention intention of the fault-based liability. In addition, it is hostile to innovation if mere putting into operation is made the cause of liability.¹⁰⁶

Worse, however, is: With this construction a liability gap arises that is difficult to bear and which will widen in the future as more tasks of contract performance are delegated to autonomous software agents.¹⁰⁷ “If the operator can prove, however, that the occurrence of damage was neither predictable nor avoidable in accordance with the state of the art, then ... liability is omitted.”¹⁰⁸

Both difficulties—the liability gap and the doctrinal misconception—can be avoided without any problems if one applies the liability rules for auxiliary persons analogously to the behavior of autonomous software agents. Even if the principal himself is not accused of misconduct, he shall be liable for any conduct of the software agent, insofar as the latter violates the contractual obligations and causes damage.¹⁰⁹ The analogy would, as Grundmann explains from an economic point of view, internalize the costs of machine failure. Since the decision-maker would also be the cost-bearer, he would have the optimum incentive to weigh up the benefits and costs of greater machine safety in a minimizing manner.¹¹⁰

The analogy completely eliminates the liability gap because the principal can no longer relieve himself by alleging lack of misconduct on his part. It depends solely

¹⁰³ Palandt/Grüneberg, BGB 77th ed. 2018, § 278 Rn. 11; Redeker, IT-Recht, 2012, Rn. 969 f.; Koch, Internet-Recht 2005, § 3, Nr. 1. b, S. 101; Heckmann, Internetrecht 2007, Kap. 4.1 Rn. 18.

¹⁰⁴ Horner/Kaulartz (fn. 92), 23; Jochen Hanisch (2014) "Zivilrechtliche Haftungskonzepte für Robotik", in: Eric Hilgendorf (ed) *Robotik im Kontext von Recht und Moral*, 27–61, 32.

¹⁰⁵ This is also criticized by Lohmann (fn. 49), 158; Schirmer (fn. 23), 664.

¹⁰⁶ Hanisch (fn. 104), 34.

¹⁰⁷ Münchener Kommentar BGB/Grundmann, 7th ed. 2016, BGB § 278 Rn. 46 speaks of a regulatory gap caused by technological developments.

¹⁰⁸ Kirn/Müller-Hengstenberg (fn. 9), 16. Also Palandt/Grüneberg BGB 77th ed. 2018, § 276 Rn. 20 f.

¹⁰⁹ For the analogy, Keßler (fn. 29), 592 f.; Schirmer (fn. 23), 665; Gruber (fn. 28, 2015), 198; Teubner (fn. 29); more cautiously advocating only for the future Günther (fn. 17), 84.

¹¹⁰ Münchener Kommentar BGB/Grundmann, 7th ed. 2016, BGB § 278 Rn. 46, which then, however, opposes the analogy.

on the misconduct of the software agent.¹¹¹ Here lies the real advantage of the liability for auxiliary persons over the principal's liability for his own misconduct. Even if the principal has fulfilled all due diligence obligations when using the computer, he must nevertheless be liable for a decision failure of the autonomous software agent, as if a human vicarious agent had acted.

The deeper reason for this, as I have already said, is that, if society allows new areas of decision-making for previously unknown autonomous decision-makers, it is obliged to ensure effective forms of responsibility. Ultimately, the decisive factors are not aspects of efficiency, transaction cost savings, utilitarian considerations, issues of sociological jurisprudence or policy-questions, but this is a genuine question of legal justice. It is the principle of equal treatment that demands liability here. If the execution of the contract is delegated to a human actor, the principal is liable for his breach of duty, consequently, he cannot be exempted from liability if a software agent is used for the identical task. It would be an unjustifiable privilege, if digitalization provided a computer operator with such a considerable cost advantage to the detriment of the injured contract partner. If legal doctrine unswervingly adheres to traditional legal terms and refuses to assign legal action capacity to software agents, it would have to be accused of treating equal cases unequally.

Subjective intentions should cause even fewer difficulties in the analogy with the liability for auxiliary persons than in the case of contractual representation. For the liability for assistants, it is even less important to have inner abilities than to participate in communication. Here again, the problem is alleviated by objectivization tendencies in today's private law, which have changed the requirements on fault. These tendencies almost merge illegality and fault in "objective negligence".¹¹² This simplifies the requirements for the analogy. In information philosophy, which examines the moral responsibility of algorithms, it is argued that the cognitive abilities of autonomous software agents are developed in such a way that they cannot be considered "responsible" in the full sense of the word, but "*accountability*" can be attributed to them.¹¹³ In legal translation, this would mean that they can be attributed illegality, but no fault in the sense of subjective reproach. And at this point, software agents can even be required to exercise greater care than the acting humans, provided that they possess higher cognitive abilities due to their superior information processing capacity.¹¹⁴

3. *Tortious liability*

In the area of non-contractual liability, the failure of the current law to deal with the digital autonomy risk becomes abundantly clear. Authors criticize in all harshness that here too a considerable liability gap has been opened, and they demand urgently legislative intervention.¹¹⁵ Since most computer failures occur in the field of tortious liability this is the test case for whether or not to give software agents partial legal capacity. Not only because the frequency of damage is high and the amount of damage is considerable. In contrast to contractual liability, in which the contracting party exposes itself "voluntarily" to the risk of having a digital contractual partner, in

¹¹¹ Schirmer (fn. 23), 665.

¹¹² Münchener Kommentar BGB/Grundmann, 7th ed. 2016, BGB § 276 Rn. 53 ff.

¹¹³ Floridi/Sanders (fn. 6), 205 ff.

¹¹⁴ Münchener Kommentar BGB/Grundmann, 7th ed. 2016, BGB § 278 Rn. 50.

¹¹⁵ European Parliament (fn. 3), 7.

cases of non-contractual liability the injured party is involuntarily exposed to the computer risk.

In tort law as well, the liability standards focus exclusively on whether operators/manufacturers/programmers are accused of misconduct. As a consequence, current law does not sanction the wrong decisions made by software agents when the human participants have behaved correctly. The operator is exempted from liability if he has always adapted his safety precautions to the new state of the art in science and technology.¹¹⁶ The producer is not liable if he has fulfilled all construction, information and product observation obligations.¹¹⁷ The exemption from liability for development risks seems particularly questionable in view of the programmed (!) non-predictability (!) of the algorithm decisions. If the human participants have fulfilled their duties of care, but the software agent makes an illegal and damaging decision, the often-criticized liability gap arises. It will grow in the future, because with the increasing independence of autonomous systems it can no longer be expected that users will be able to continuously monitor behavior and the legal standards of testing obligations will decrease accordingly.¹¹⁸

However, making liability possible *de lege lata* is likely to be much more difficult than in the area of contractual liability. A conceivable solution would be a general corporate liability for the actions of autonomous software agents, according to which their misconduct is directly attributed to the company.¹¹⁹ This would be a joint liability, which, within the comprehensive framework of the corporate organization, would do justice to the hybrid character of human-computer interaction with a member status of the software agent in the group. But it would be unrealistic to propose this at the current stage of corporate liability law *de lege lata*.

4. Strict liability or digital assistance liability?

De lege ferenda seems to be the royal road to successfully countering the digital risk of autonomy.¹²⁰ Most of the authors also insist strongly on a corresponding legislation that responds to the digital dangers with a strict liability. Examples of this are the hazardous liability standards for motor vehicles, nuclear installations, genetic engineering and environmental damage.

However, a fundamental misunderstanding must be dispelled. The guiding principles of strict liability cannot serve as a role model at all, since they simply fail to take account of the peculiarities of the digital autonomy risk. The thesis goes: Not liability for the lawful use of dangerous equipment, but rather liability for illegal misconduct of the autonomous, decision-making software agents is appropriate as a basic principle of vicarious liability for digital action. Accordingly, the proposal is presented here to introduce a “digital vicarious liability” i.e. liability of the principal for misconduct by the software agent instead of strict liability.

¹¹⁶ Döpke (fn. 89), 17; Bräutigam/Klindt (fn. 12), 1140; Kirn/Müller-Hengstenberg (fn. 2), 68; Horner/Kaulartz (fn. 92), 24.

¹¹⁷ Gerald Spindler (2014) "Zivilrechtliche Fragen beim Einsatz von Robotern", in: Eric Hilgendorf (ed) *Robotik im Kontext von Recht und Moral*, Baden-Baden: Nomos, 63–80, 72 ff., 78.

¹¹⁸ Horner/Kaulartz (fn. 92), 25.

¹¹⁹ On a convincing legal policy conception of corporate liability, Münchener Kommentar BGB/Wagner, 7th ed. 2017, BGB, § 831, Rn. 3 f.

¹²⁰ European Parliament (fn. 3), 6; Günther (fn. 17), 237 ff.; Schirmer (fn. 23), 665; Gruber (fn. 27), 198; Hanisch (fn. 104, 2014), 35 f.; Spindler (fn. 17), 775.

The following fundamental difference exists with regard to strict liability. Dangerous things are subject to strict liability in their use, which is nevertheless permitted, i.e. lawful due to social benefits.¹²¹ In other words, strict liability is the rare case of liability for damages despite the lawful conduct of the owner/user/operator. Liability occurs when the typical operational hazard is realized, i.e. when causal processes have gone awry. In the case of software agents, however, it is precisely not the hazard of a wrongly functioning computer, i.e. the causal risk, but rather the decision risk, the very different kind of risk that autonomous decisions will turn out to be wrong decisions. The reason for this is not the use of an object of increased risk, but the illegal behavior of the algorithm, which the principal has legitimately used for his own benefit.¹²²

Since the autonomy risk of digital decisions cannot, however, be equated with the causality hazard of strict liability, different principles of responsibility and different standards must also be developed here. Norms of liability for unlawful decisions by autonomous agents cannot be based on the causal risk of things, but must be tailored to the decision risk of actors. At this point, too, the function of personifying non-human entities realizes that it replaces causal attribution with an action attribution.¹²³ If private law, as proposed here, treats software agents as vicarious agents, i.e. as legally capable of acting, then it is absolutely impossible to operate with a mere causal liability in the case of non-contractual damages.

In short: *De lege ferenda* a “digital assistance liability”, i.e. strict liability for damages caused by autonomous software agents should indeed be introduced, but not as a liability for the legitimate use of a dangerous installation, but as a liability of a principal for unlawful decisions of his software agent.¹²⁴

This fundamental difference becomes practically relevant for two conditions of liability: (1) In contrast to strict liability, liability is assumed for unlawful behavior by the software agent. (2) The liability extends to more than just personal injury and property damage. In other words, it would go too far on the one hand, because it would also trigger liability in the event of conduct that simply causes the damage. On the other hand, it would not go far enough because it only provides compensation for personal injury and property damage.

Unlawfulness is not a condition for strict liability, but in the case of software agents it becomes the linchpin of liability.¹²⁵ It all depends on whether the software agent’s action was unlawful or not. While in the case of strict liability, it is sufficient if

¹²¹ *Locus classicus*, Josef Esser (1941) *Grundlagen und Entwicklung der Gefährdungshaftung: Beiträge zur Reform des Haftungsrechts und zu seiner Wiedereinordnung in die Gedanken des allgemeinen Privatrechts*, München: Beck. On a sociological interpretation, Luhmann (fn. 74), Ch. 4, II.

¹²² On the decisive difference between absolute liability and strict unlawfulness liability, Münchener Kommentar BGB/Wagner, 7th ed. 2017 Edition, BGB, § 831, Rn. 4.

¹²³ Teubner (fn. 29).

¹²⁴ Similarly Hanisch (fn. 104), 46 ff., who demands liability for “machine misconduct”, instead of strict liability. Similarly also with the proposal *de lege ferenda* of vicarious liability for robots Melinda Müller (2014) “Roboter und Recht: Eine Einführung”, *Aktuelle juristische Praxis*, 595-608, 601.

¹²⁵ The European Parliament’s draft also fails to recognize the specificity of software liability, in contrast to strict liability, which only calls for a causal link between the harmful behavior of the computer and the damage incurred, European Parliament (fn. 3), Rn. 27. Correctly, Hanisch (fn.104, 2014), 46, who makes machine misconduct a prerequisite for liability.

there is merely causality between operational hazard and damage, when using software agents it must be checked whether their behavior was unlawful or not.

The second important difference to strict liability concerns the extent of liability. While strict liability standards regularly exclude liability for pecuniary loss which is justified in view of the typical risk, such liability limitations are not acceptable for software agents. Because their typical risks are not only realized in the contexts in which dangerous installations cause accidents, but also in all the contexts in which people make unlawful decisions. As in the case of damaging acts by natural persons, financial losses must be compensated for here as well, if the actions of the agents violated a legal right or violated a statutory rule or were against public policy. And that makes a big difference.

The question of which asset holders such strict liability for mistaken computer decisions can then be attributed should be based on the model of the “attribution and liability unit”, in which the case-law in the case of liability for hazardous vehicles has summarized drivers, owners and insurance companies.¹²⁶ Operators, manufacturers and programmers of the software agent should be bundled to a similar liability network. The proposal by Hanisch of a tiered liability between the operator and the manufacturer, according to which the operator is primarily liable and the manufacturer secondarily so, is also worth considering in this context.¹²⁷ A compulsory insurance policy for digital risks that affects manufacturers would cushion the burden of hardship for the operator, especially if they are not acting commercially.¹²⁸ The introduction of maximum sums would correspond to the insurance logic.

De lege ferenda, a general clause stipulating initiators’ liability for autonomous algorithms would make sense. The general clause would only have to provide that the initiators are responsible for damage-causing actions of the software agent, if the actions of the software agent were unlawful.¹²⁹

Insurance solutions, especially compulsory insurance, in turn represent a change in the legal status of software agents. They allocate a liability capital exclusively to the actions of the software agents, who thus do not become legal entities with their own funds, but obtain the legal status of a “liability subject with insurance assets”. Only that these assets are not attributed as property, but are rather socialized, either to the community of insured persons or even to the entire society.¹³⁰ In this sense, insurance solutions lend functional equivalents to full legal subjectivity. More importantly, insurance solutions also imply the action capacity of software agents. If insurances are providing cover for strict liability, it would be an incalculable risk for them if they had to pay for every damage that came about causally, without the presence of digital misconduct. They will only be prepared to provide insurance cover in the event of unlawful conduct by the agent, especially if the obligation to pay compensation goes beyond personal injury and property damage.

¹²⁶ Zech (fn. 3), 204.

¹²⁷ Hanisch (fn. 104, 2014), 55 ff.

¹²⁸ European Parliament (fn. 3), Rn. 29.; Horner/Kaulartz (fn. 92), 26.

¹²⁹ Hanisch (fn. 104, 2014), 46 ff., 54.

¹³⁰ On solutions under social security law, Janal (fn. 3), 157.

The frequently called for strict liability for software agents is therefore in reality nothing more than a variant of an “assistant liability”, which necessarily presupposes unlawful behavior of the agent and whose extent is not limited to personal injury and property damage. It would implicitly or explicitly recognize a limited legal capacity of the software agents.

IV. Hybrids: Association risk

1. Human-Machine Associations

In response to the risk of autonomy, legal solutions have so far been proposed in which algorithms have limited legal capacity and are treated as autonomously decision-making representatives or auxiliaries. This is still a reasonably secure area for legal doctrine. However, one enters a perilous terrain when focusing on the human-algorithm-association as a regulatory object. The reason for this is that both the solution of granting limited legal capacity to software agents, and the solution of simply dressing up software behavior as human behavior, leads to a problematic alternative. On the one hand, to attribute rights and obligations exclusively to the human actors involved does not adequately capture digital risks, because it ignores the autonomous role of the machines, spans almost with necessity the duties of the people involved and reduces the protection of the machines with regard to rights. On the other hand: The actions—as proposed here—to be attributed exclusively to the machines are also not without problems, because then the contribution of the people in the human-machine association is only dealt with in individualistic legal categories of agency law, which does not do justice to the peculiarities of the human-computer-association itself. Here again Bruno Latour’s actor-network theory offers a way out. It conceives of the human-machine interaction as a collective phenomenon of its own kind: According to Latour, “actants” possess not only a language and a resistant body, but also the ability to form “associations”.¹³¹ They owe their participation in society not only to their individual status as “actants”, but also to their membership in “hybrids”, i.e. to associations of human actors and non-human actants. As in any association, this creates a resource pool, which, in the case of hybrids, combines the limited capacity of the actants to act with the communicative skills of real people. “The psychosystemic deficits of the non-human’s competence are reliably compensated for by the distributed intelligence of the social systems.”¹³² And it is precisely the combination of human and non-human capacities for action within the hybrids that allows non-human entities to participate in more complex political negotiations, economic transactions and legal activities.¹³³

Similar to a formal organization, the human-machine association itself develops a phenomenal inner perspective, a self-awareness as a “living” process, its own hierarchy of preferences, its own social needs and political interests that cannot be reduced to the characteristics of the human actors involved or to those of the

¹³¹ Latour (fn. 36), 70 ff.

¹³² So Lorentzen (2002) “Luhmann Goes Latour: Zur Soziologie hybrider Beziehungen”, in: Werner Rammert and Ingo Schulz-Schaeffer (ed) *Können Maschinen handeln? Soziologische Beiträge zum Verhältnis von Mensch und Technik*, Frankfurt: Campus, 101-118 110.

¹³³ On the cooperation of humans and algorithms, Misselhorn (fn. 49); Anita Fiebich, Nhung Nguyen and Sarah Schwarzkopf (2005) “Cooperation with Robots? A Two-Dimensional Approach”, in: Catrin Misselhorn (ed) *Collective Agency and Cooperation in Natural and Artificial Systems: Explanation, Implementation and Simulation*, Heidelberg: Springer, 25-44.

algorithms. The interactions within the network make the hybrids comparable with corporate actors. The actors involved do not act on their own behalf, but “for” the hybrids as an emergent entity, as an association of people and non-humans. They do it in the same way as managers in a company who do not act in their own name, but as “agents” acting on behalf of their “principals”, i.e. for the company itself as a social system. And there are conflicts of interest between the members and between them and the association, as discussed in the light of the well-known agency problem faced by corporate actors, for which the solution has been developed in numerous legal norms. Such conflicts also exist similarly in the associations of humans and non-humans. And comparable institutional regulations—e.g. the formalization of the duties and responsibilities of the managing director, the Ultra-Vires doctrine, the examination of representativeness in class action lawsuits—serve to contain the agency problem, which also arises in human-machine associations. In fact, in the legal debate on electronic contracting, similar solutions are already being discussed with regard to the “principal-agent relationship” between the contracting computer and the human contracting parties.¹³⁴

This is a collectivist view that frees the law from the not unproblematic individualist alternative of assigning the act in question either exclusively to humans or exclusively to the algorithm. In contrast to individual attribution, it is also capable of doing justice to the emergent phenomenon of the human-machine association.¹³⁵ On the one hand, it takes into account the emergence in internal relations which, beyond the existence of several individual actors, creates the unity of their association. On the other hand, it does justice to the emergence in external relations: the human-algorithm association as an independent system communicates itself with third parties and it is no longer only the people or algorithms involved to whom communication is attributed. The risks arising from the almost inextricable interweaving of the individual actions of humans and algorithms can be better counteracted by identifying the human-algorithms association, the hybrid as such, as a common point of attribution for actions, rights and obligations.

2. Hybrids as legal entities?

Should legal doctrine introduce the legal construction of a human-machine association? This is a serious option, even if it has not yet been tried out. The individual legal acts of the software agents and those of the people involved, without having to separate them individually, would be merged into a unified act of the human-machine association and would create both legal ties and liability claims. In contrast to the individualistic law of agency, which clearly separates the individual actions of principals and agents and declares the principal to be the contractual partner, the human-machine association itself would become the actual contractual partner. For contractual and non-contractual liability, the individual preconditions for liability would be composed of their composite conduct, without their individual contributions having to be painstakingly and often arbitrarily calculated apart, as it would be necessary in the case of liability for auxiliary persons. Finally, knowledge allocation would not exclusively be based on the computer or on the human actor,

¹³⁴ Juanda L. Daniel (2004) "Electronic Contracting under the 2003 Revisions to Article 2 of the Uniform Commercial Code: Clarification or Chaos?", 20 *Santa Clara Computer & High Technology Law Journal*, 319- 346, 344 ff.

¹³⁵ Kersten (fn. 17), 4 ff.; Ingold (fn. 17), 220 ff.

but always on the association. Assets would be accessed under liability law wherever they are located in the hybrid.

However, to actually construct the hybrid as a legal entity would mean that the law would have to create a completely new kind of corporate entity, the human-machine association.¹³⁶ This would indeed be a bold—if not daring—step, for which today's private law is hardly prepared. Algorithms as full members of a novel association? To a much greater extent, this construction would be more radical than granting software agents partial legal capacity as representatives of their human principal. Although the collectivist legal solution is ultimately more appropriate to the reality of human-machine interaction, courts and doctrine are likely to prefer, instead of attributing all behavior uniformly to the association, to choose the individualist agency law solution.

Be that as it may, even if the agency law is preferred, the unity of the human-algorithm association will not be legally irrelevant; it now becomes effective as a legally significant background assumption. An interesting model for dealing with such a *unitas duplex* is already offered by the current contract law with the purpose of the contract, the common purpose of the individual parties in the exchange relationship. It is well known that the purpose of the contract is not to construct the contract itself as a collective unit, but to continue to be understood strictly individualistically as a relationship between the individual contracting parties. But the purpose of the contract, understood as the common purpose of exchange between the parties, brings the overarching unity of the contractual relationship to bear differently. It produces concrete legal consequences—for the interpretation of contractual declarations, for fiduciary duties and for the decision in disputes as to whether and to what extent a breach of contract has occurred. Moreover, in long-term contracts, relational contracts and especially in contractual networks, e.g. in supplier and distribution networks, this purpose of the contract is becoming more and more important—now under the titles “association purpose”, “final nexus” or “network purpose”—without having to choose corporate law constructions.¹³⁷ Similarly, the human-computer associations would not have to be endowed with legal capacities themselves; rather an “association purpose” could do justice to their hybrid character, correct a purely individualistic construction and also bring the unity of the human-machine association to bear in legal terms.

Ultimately, both solutions are not to be understood as mutually exclusive alternatives. They could very well exist side by side. The agent-principal solution would be appropriate if the algorithms were to act in social life as clearly defined actors. If, on the other hand, they are embedded in a dense interaction context with human actors, the associational solution should be preferred. However, the two solutions have in common the necessary precondition that they confer limited legal capacity to non-human actors—actants or hybrids.

¹³⁶ Gruber (fn. 27), 158; Allen/Widdison (fn. 17), 40 referring to the social reality of the hybrid. Christian Neuhäuser (2015) "Some Sceptical Remarks Regarding Robot Responsibility and a Way Forward", in: Catrin Misselhorn (ed) *Collective Agency and Cooperation in Natural and Artificial Systems*, Heidelberg: Springer, 131-147, 143 f. advocates a collective moral responsibility of man-machine associations as “responsibility networks”.

¹³⁷ On the new category of the network purpose, Gunther Teubner (2011) *Networks as Connected Contracts*, Oxford: Hart, 184 ff.

V. Multi-agent systems: Network risk

The legal technique of personifying actants or hybrids, however, reaches its limits when many autonomous algorithms are interconnected within a multi-agent system. Personification needs a determinable substrate for clearly identifiable software agents or definite human-machine associations, but it does not work in a situation of complex computer interconnections.¹³⁸ The network risk that is realized here destroys assumptions about the individuality of actors which are constitutive for the attribution of action and responsibility.¹³⁹ Both the actor and the causal relationships are difficult, if not impossible, to identify.¹⁴⁰

Spiecker clearly shows the responsibility gap of the current law. However, her proposal for a pro-rata liability of all those involved in the network, who would have to stand up for the unlawful behavior of the networked algorithms,¹⁴¹ is hardly practicable. In view of the almost infinite interconnection of algorithms in the digital space, it is practically impossible to delimit liability subjects. In addition, there are huge investigation costs: "Linking abstract persons' actions in the information society to their principals may require considerable effort, perhaps at a higher cost than the damage at issue."¹⁴²

One way out of this dilemma would be to attribute legal responsibility to identifiable actions, i.e. to an "anonymous matrix" of social and digital processes themselves, rather than to "persons" as actors. The actual attribution points for responsibility are then autonomous decisions and no longer the decision-makers. Ultimately, it would not be people, organizations, networks, software agents, algorithms, but rather the decisions themselves that would have to be made responsible.

In such a situation, the law itself would need to construct risk pools for the delimitation of these interrelationships, especially if no factual indications of personification are identifiable in the digital space.¹⁴³ Here, the law would finally leave the actor perspective, because it no longer looks for individual or collective actors but rather focuses on the risky decisions as such. It makes chains of actions responsible for their consequences without caring for organized decision centers. The decisive difference between such risk liability and known forms of organizational liability lies in the fact that liability law no longer refers to organizational arrangements that exist in reality or at least refers to cooperative relationships, but rather defines, not to say decrees, new types of risk networks. And as soon as the actions of individual, collective

¹³⁸ Spiecker (fn. 7), 701.

¹³⁹ Susanne Beck (2014) "Dealing with the Diffusion of Legal Responsibility: The Case of Robotics", in: Fiorella Battaglia, Nikil Mukerji and Julian Nida-Rümelin (ed) *Rethinking Responsibility in Science and Technology*, Pisa: 167-182; Floridi/Sanders (fn. 6), 205 ff.

¹⁴⁰ Zech (fn. 3), 170.

¹⁴¹ Spiecker (fn. 7), 703.

¹⁴² Koops et al. (fn. 19), 553.

¹⁴³ Hildebrandt (fn. 37), 26 as well defines actions and not actors as points of attribution, although perhaps more cautiously: "Because the agents may be distributed on and possibly mobile between different hardware applications and because as a multi-agent system it is capable of changing shape (polymorphous), it is not always easy to identify where the emerging agent is located and what is and is not a part of it at any point in time. However, in so far as the emergent behaviours of the system allow its identification as a unity of action, it can be qualified as an agent, whatever the underlying embodiment."

actors or the calculations of digital actors move into such a space, they all become “compulsory members” of such a risk pool—by the authoritative order of state law. Admittedly, there is still a voluntary part of the action, but it is limited to the decision on entry. They then become collectively liable irrespective of their intention and without regard to the causal relationship between the actions they are individually responsible for and the damage incurred.¹⁴⁴

The risk pool would no longer be determined by cooperative, organizational or technical structures. Rather, it should be defined as a “digital problem area”, the limits of which should be determined by the suitability for collective risk management. Ultimately, neither causal connections nor pre-defined cooperative structures—however important both are in individual cases—are decisive, but the central criterion should be the pool’s ability to manage risk. Admittedly, this would amount to “opportunistic” attribution.¹⁴⁵ For the purpose of collective liability, it is the law that identifies by its own authority concrete digital risk contexts of the offline and online world with the ulterior motive of creating a social structure that can control these risks to a certain extent in a preventive manner or at least be used for the settlement of damages.

This risk management concerns firstly the settlement of losses already incurred. In a situation of multiple causality, the law balances the losses in such a way that it creates an adequate financial pool which covers the losses and distributes the risk (“deep pocket”, “risk spreading”). Secondly—and perhaps more importantly—risk management means the collective management of future behavior. The law designates the limits of the risk pool in such a way that a realistic basis is created for the active and joint prevention of risks in areas where damage problems are concentrated. From both points of view, the law isolates the social space of collective responsibility in such a way that a functioning digital technology can develop to cope with the risks of digital networking.

VI. Results

The new digital risks—autonomy risk, association risk, network risk—confront private law with the challenge of redefining the legal status of autonomous software agents. Certainly not in the sense that they suggest the full personification of software agents, human-computer associations or multi-agent systems. Rather, in response to the three risks, the law should carefully calibrate the legal status of algorithms based on their concrete role. For the autonomy risk, it is an adequate answer to grant software agents the status of partial legal personhood. Their autonomous decisions should be legally binding and should give rise to liability consequences. This gives them the limited legal subjectivity to conclude binding contracts for others as a proxy. At the same time, in cases of contractual and non-contractual liability, they are to be recognized as legally capable assistants, so that the machine misconduct itself (and not merely the conduct of the companies behind it) constitutes a breach of duty for which the companies must be held responsible.

While a clear-cut legal response to the autonomy risk is possible, solutions for the association risk and the networked risk are currently only visible in contours. A

¹⁴⁴ On concurrent considerations in environmental liability law, Gerhard Wagner (1990) *Kollektives Umwelthaftungsrecht auf genossenschaftlicher Grundlage*, Berlin: Dunker & Humblot.

¹⁴⁵ Luhmann (fn. 74), Ch. 6, V.

conceivable response to the association risk would be to give software agents the legal status as a member of a human-machine association. A maximum solution *de lege ferenda* would attribute actions, rights and obligations to the association itself—a solution that would break new ground in private law. A minimal solution *de lege lata*, on the other hand, would introduce the legal concept of association purpose, which could guide the interpretation of computer declarations as well as the determination of the participants' rights and obligations. Finally, the answer to the network risk would be a risk pool to be defined autonomously by the liability law itself, which would determine the legal status of the algorithms in the context of a comprehensive digital information flow and would establish the liability of the pool solely on the basis of unlawful behavior of the pool.

Actors with limited legal subjectivity, members of a human-machine association, elements of a risk pool — these are the new forms of a digital legal status. Their conceptualization must be geared to whether and how the above-mentioned deficiencies in responsibility can be overcome. But at the same time, any such a legal status for software agents must make it possible to further develop the doctrine of contract law and liability law consistently and without unacceptable fictions—not as an end in itself, but as the law's obligation to treat equal what is equal and to treat unequal what is unequal in the digital space as well.