6

The Legal Personhood of Artificial Intelligences

Preliminaries: Three Contexts

This chapter applies the Bundle Theory of legal personhood to artificial intelligences (AIs) to see what insights the theory can yield. As has been the case throughout this book, I do not propose to participate directly in the debate over whether AIs should be legal persons, but rather to provide a structure and framework for that debate. My aim is also to expose certain problems that afflict any efforts by proponents of the Orthodox View of legal personhood to elucidate the issues at hand.

The field of artificial intelligence is developing at a breathtaking rate. One relatively recent example is how Google's software was able to beat the best human players in the Chinese game of Go—a feat that was until recently considered to be decades away.[1] Legal and political actors are responding to this change in various ways. The United Nations General Assembly commissioned in 2013 a report on lethal autonomous robots (which could decide to kill without human intervention),[2] and investment banks already employ so-called robot traders.[3] The increasing role of AIs in commerce prompted the Committee on Legal Affairs of the European Parliament to assert in a 2017 report that 'the civil liability for damage caused by robots is a crucial issue which also needs to be analysed and addressed at Union level'.[4] The Parliament called on the Commission to

---

[1] See for instance Matt Reynolds, 'DeepMind's AI beats world's best Go player in latest face-off' *New Scientist* (23 May 2017) <https://www.newscientist.com/article/2132086-deepminds-ai-beats-worlds-best-go-player-in-latest-face-off/>, visited on 13 June 2018.

[2] Christof Heyns, Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, A/HRC/23/47.

[3] I use the word 'artificial intelligence' (as a countable noun) to refer to a human-built entity that can act in ways characteristic of intelligent beings, especially humans. What I mean by 'robot', on the other hand, is a mechanical entity capable of interacting directly with the physical world. Even if robots can also be under the direct control of a human being, I will focus on autonomous robots here. The often-used label 'robot trader' (meaning software designed to buy and sell stock, derivatives, and so on) is not very good because it does not refer to mechanical entities. Phrases such as 'software trader' or 'AI trader' would be more suitable.

[4] European Parliament, Committee on Legal Affairs. Report with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)) 16.

---

explore, analyse and consider the implications of all possible legal solutions, such as […] creating a specific legal status for robots in the long run, so that at least the most sophisticated autonomous robots could be established as having the status of electronic persons responsible for making good any damage they may cause, and possibly applying electronic personality to cases where robots make autonomous decisions or otherwise interact with third parties independently.[5]

Questions surrounding AI legal personhood are thus multifaceted, and a cluster-property understanding of legal personhood is well suited to analysing the question. Let us now distinguish three main contexts that bear on the legal personhood of AIs. The contexts overlap in many ways, but they can usefully be distinguished, and they will provide a structure for this chapter.

First, in the *ultimate-value context* we ask whether AIs are of ultimate value and therefore worthy of receiving some of the protections that legal persons such as human children enjoy. Science fiction is replete with examples of scenarios where some features of a (usually humanlike) robot prompt questions of its morally correct treatment. Fundamental protections are especially important here, but many other—perhaps all other—incidents of legal personhood are relevant as well.

Second, in the *responsibility context* our focus is, unsurprisingly, on the legal responsibility of AIs. Could self-driving cars or autonomous security robots be held criminally or tortiously liable for their actions? The most relevant incident here is—as one might anticipate—onerous legal personhood. In addition, tort liability for an AI would also require that the AI could own property.[6]

Finally, the *commercial context* has to do with AIs' functioning as commercial actors; buying, selling, and so on. The three most relevant incidents of legal personhood here are special rights, ownership, and legal competences.

These three contexts share a certain connection with the three central 'building blocks' of legal personhood that I have introduced in Chapter 4. Passive legal personhood functions through claim-rights, and is closely connected to the question of whether AIs could be ultimately valuable. I have argued that idols and bodies of water are not legal persons—regardless of the pronouncements of any legislator or judiciary—because they are not of ultimate value and therefore cannot hold claim-rights. Some human collectivities, on the other hand, can hold claim-rights because they are a shared project of human beings. In what follows, AIs' claim-rights will be addressed from two different angles. If some AI is of ultimate value, then it follows

---

[5] European Parliament, Committee on Legal Affairs. Report with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)) 17–18.

[6] See Gabriel Hallevy, *When Robots Kill: Artificial Intelligence under Criminal Law* (Northeastern University Press 2013), and Robert van den Hoven van Genderen, 'Do We Need New Legal Personhood in the Age of Robots and AI?' in Marcelo Corrales, Mark Fenwick, and Nikolaus Forgó (eds), *Robotics, AI and the Future of Law* (Springer 2018).

quite straightforwardly that the AI can hold claim-rights. If, on the other hand, an AI is not of ultimate value, it can under certain conditions hold claim-rights as the 'administrator' of a human-defined project.

Only AIs that are of ultimate value could be purely passive legal persons—a legal status comparable to that of an infant or a comatose individual. Such an AI could of course also be endowed with an active legal personhood if it were for instance capable of contracting, administering its property, and so on. If an AI is not of ultimate value, it can only be a legal person in virtue of its capacity to be subjected to legal duties and/or to administer legal platforms through the exercise of competences. Legal duties are here connected to the responsibility context, whereas legal competences relate to the commercial context.

The discussion will be focused on so-called strong AIs, and I will proceed from the assumption that such AIs will sooner or later come into existence. A strong AI is an entity that can in relevant respects act like a human being. As with human collectivities, we can treat AIs as legal persons if they can perform like human individuals in a sufficient number of the relevant legal contexts: ownership, contracting, and so on.[7] For most of our purposes here, it is irrelevant whether the AI can 'really' think or whether it merely acts 'as if' it thinks.[8] Here we need to recall the functional/felt distinction and the intentional stance. Let us take the institution of contracting as an example. An AI does not need to understand the institution in a felt, phenomenal way in order to be able to contract. If an AI's potential business associates can rely on the AI's ability to adjust its future behaviour accordingly when signing a contract, contracting with the AI would be intelligible. The associates can then adopt the intentional stance when dealing with the AI; it is immaterial for this purpose whether the AI can for instance experience mental states pertaining to the contract. It might be that such mental states are required for moral contracts, but certainly not for legal contracts.

I will first address the ultimate-value and sanction contexts. The treatment of these two settings will be relatively brief. Because my theory has most to offer when analysing the commercial context, the preponderance of the chapter will be focused thereon.

---

[7] This 'intentional-stance' or 'pragmatic' approach to AI legal personality is endorsed by many who have written on the topic. See for instance Samir Chopra and Laurence F. White, *A Legal Theory for Autonomous Artificial Agents* (University of Michigan Press 2011) 1–17, and Lawrence B. Solum, 'Legal Personhood for Artificial Intelligences' (1992) 70 *North Carolina Law Review* 1231.

[8] John Searle has famously argued that an entity could be able to process information in a way that allows for it to act as if it comprehends the information, but without actually understanding it. John Searle, 'Minds, Brains and Programs' (1980) 3 *Behavioral and Brain Sciences* 417. For a good summary of the argument and the subsequent counterarguments, see Larry Hauser, 'Chinese Room Argument', *International Encyclopedia of Philosophy* <http://www.iep.utm.edu/chineser/>.

AIs as Ultimately Valuable

The moral status of AIs is a fiercely debated topic. Some philosophers maintain that AIs can never achieve a status of moral considerability—Joanna Bryson argues explicitly that robots should be treated as slaves[9]—whereas many have contended that we will at some point owe moral duties to robots or other types of AIs.[10] Even though the moral debate over the value of AIs is complex, its implications are straightforwardly applicable here. If we assume that some AIs are of ultimate value, then they can hold claim-rights; we can owe duties to them, and our duties do not merely pertain to them. We can therefore conclude that they can be passive legal persons.

We should once again note the difficulty that the Orthodox View has in explaining what is at stake here. Authors often describe the legal personhood of AIs as consisting in the ascription to them of 'rights and duties of their own'.[11] This definition would entail that, if AIs are endowed with any rights or duties whatsoever, they are legal persons. But this conclusion would be as problematic in connection with AIs as it is in connection with animals and slaves. Let us assume that a society becomes concerned about the bad treatment of humanoid household robots because the robots are thought to feel pain. The legislature then enacts a Robot Welfare Act that prohibits certain particularly gratuitous acts of cruelty toward robots. According to the interest theory of rights, such prohibitions endow the robots with rights (assuming the robots are of ultimate value). Are the robots now legal persons? For reasons laid out in Chapter 2, I am sceptical of such a conclusion. Animals hold, and slaves held, similar legal rights, yet animals and slaves are widely—and correctly—classified as legal nonpersons. Thus, we need to distinguish robots as right-holders from robots as legal persons. These humanoid robots would likely qualify as legal persons if they could no longer be owned and if they received wide-ranging fundamental protections (for instance, attempts to shut them down would be classified as attempted homicides).

---

[9] Joanna J. Bryson, 'Robots Should Be Slaves' in Yorick Wilks (ed.), *Close Engagements with Artificial Companions* (John Benjamins Publishing Company 2010).

[10] See for instance John Basl, 'Machines as Moral Patients We Shouldn't Care About (Yet): The Interests and Welfare of Current Machines' (2014) 27 *Philosophy & Technology* 79 and Eric Schwitzgebel and Mara Garza, 'A Defense of the Rights of Artificial Intelligences' (2015) 39 *Midwest Studies in Philosophy* 98. For a more noncommittal view, see David J. Gunkel, 'The Other Question: Can and Should Robots Have Rights?' [2017] *Ethics and Information Technology* 1. Gunkel's account builds, however, on a rather peculiar understanding of the is/ought distinction, as he claims that the question whether robots can hold rights is an 'is' question rather than an 'ought' question. Claims such as 'Robots can hold rights and should therefore hold rights' would then supposedly be problematic from the point of view of Hume's law. However, whether an entity can hold rights is clearly a moral question in itself, and therefore an 'ought' question. Of course, the proposition that AIs should hold rights doesn't follow from the proposition that they can hold rights, even when both are correctly understood as 'ought' questions.

[11] See for instance Ugo Pagallo, *The Laws of Robots* (Springer 2013) 40 and Ugo Pagallo, 'Vital, Sophia, and Co.—The Quest for the Legal Personhood of Robots' (2018) 9 *Information (Switzerland)*.

AIs as Active Legal Persons

Active legal personhood has to do with the capacity to act. We can understand the potential approaches to the acts—or 'acts'—of AIs on a continuum. One end involves treating AIs purely as tools. If we access a website to download illegal material or to buy stock, the personal computer we use for these purposes is obviously not taken to have performed the illegal act or entered the contract. The computer is analogous with, say, a pen used to sign a contract or a gun used to rob a bank.

The other extreme involves treating AIs more or less identically with the way adult human beings of sound mind are treated today. We could imagine a sophisticated AI that owns property in its own name, can contract and sue, and is fully subject to criminal and civil responsibility. The AI starts trading in stock and decides that the best way to maximize its profits is through insider trading, so it starts to acquire insider information illegally. The AI is able to comprehend the legal consequences of its actions; it simply performs an analysis of the potential costs—the risk of getting caught, the sanctions, and so on—and concludes that the expected value of the operation is positive. The AI's attempts are detected, and it has to pay large fines as well as compensation to the parties who have lost money because of the operation. The designers of the AI are not held accountable.

In between these two extremes fall a multitude of scenarios. A rather obvious and often-invoked example is the treating of an AI as a representative of a legal person. Similarly, criminal doctrines could be revised to allow for the criminal liability of AIs, but restricted to some limited cases. In other cases the blame for wrongdoing would fall on the designer or owner of the AI. The Bundle Theory of legal personhood offers many tools for analysing such scenarios. First, the benefits of an incident-based theory are quite obvious here. In addition, two particular notions, introduced in Chapter 4, will be useful: the distinction between independent and dependent legal personhood, and the concept of a legal platform.

Holding AIs Responsible

An often-mentioned example of a 'homicide' committed by a robot is from 1981, when an industrial robot caused the death of an employee at a Japanese motorcycle factory. The employee had entered a restricted safety zone to perform maintenance on the robot but had failed to shut it down properly, which resulted in the robot's pushing him against adjacent machinery. He died instantly.[12] Today, scholars ask whether for instance autonomous combat drones, self-driving cars, or commercial

---

[12] This incident is mentioned in much of the relevant literature, for example Gabriel Hallevy, 'The Criminal Liability of Artificial Intelligence Entities—from Science Fiction to Legal Social Control' (2010) 4 *Akron Intellectual Property Journal* 171, 171–2

AIs might be held morally and/or legally responsible for their actions.[13] I will here mainly focus on the intelligibility of punishing AIs, rather than on whether and when they should be punished.

Prima facie, it might appear that AIs should be 'angels', never acting in ways that would give rise to questions about whether they should be held legally responsible. This result could be achieved in two ways. Designers could be obligated to program their AIs not to engage in certain types of conduct ('Don't kill people, steal, commit fraud', and so on). Alternatively, nominal duties—duties unaccompanied by legal sanctions—would be extended to AIs, and programmers would be required to program AIs to always obey their legal duties, so that it would be impossible for the AIs not to follow its duties. The AIs would, obviously, then have to be able to understand and follow the precepts of law.

According to the just-depicted 'angelic ideal', the legal responsibility of AIs would be unnecessary. The ideal is, however, quite problematic. First, a programmer could simply refuse to follow the ideal. She could create an AI that is capable of understanding legal requirements but does not treat such requirements as overriding or exclusionary reasons-for-action. Rather, the potential legal consequences would be included in the overall calculus of whether to perform some action. Thus, if an AI's goal was to amass as much wealth as possible, it would—like the ideal *Homo economicus* of economic theory—weigh the potential benefits and disadvantages of some illegal course of action (e.g. insider trading), and simply the pick the option with the highest expected value. AI legal responsibility could then be justified on deterrence grounds because it would reduce the expected value of undesirable behaviour.

Second, programming an AI to be able to contravene its legal duties might also have socially beneficial consequences. Acting in breach of a contract may be economically the most efficient option, which will produce the most utility for all parties. Consider a family of four—the Smiths—who have made a reservation at a small hotel that has only three four-person suites. Before the Smiths arrive to check in, a family of twelve—the Franks—come and ask if they could book the whole hotel, as they would all like to spend the night at the same place. The hotel is unable to contact the Smiths, but the owner asks a smaller establishment across the road if they have an available four-person bedroom. The answer is affirmative, and the Smiths are relocated to the other establishment and even given a 30 per cent discount. Everybody wins: the Smiths get a discount, the Franks can stay the night under the same roof, and both hotels get more customers. Regardless, a contractual duty has been breached, as the Smiths (or rather the one who made the reservation) did not consent to this change. It is plausible that the creator of an AI would like it to perform in the manner of the hotel owner.[14]

---

[13] Chopra and White explain, using numerous examples, why AI responsibility could be meaningful—see for instance Chopra and White (n 7) 119–51.

[14] AIs functioning in inherently dangerous spheres of life might also have to be able to 'choose the lesser evil'. Google has already programmed its self-driving cars to be able to

Finally, an AI could occasionally be mistaken about the content of law. Human beings are typically not excused for their ignorance of law, so perhaps the AI should be sanctioned for its breaking the law as well—or at least be required to pay compensation for any harm it has caused by breaking the law.[15]

What kind of sanctions?

Obviously, the type of sanctions to which AIs would be subjected would depend on the details of the legal personhood arrangement. Economic sanctions would be highly relevant with regard to AI responsibility, but such sanctions require that the AI can own property. However, purely onerous legal personhood—as with slaves in the US—would be possible as well. One potential sanction could be disabling the AI if it did not obey the law.[16] We need not assume any self-interest on the part of the AI for such a sanction to work. The AI would simply recognize that its goals would be thwarted if it were disabled, and would therefore avoid conduct likely to elicit such a sanction.

One problem with the responsibility of AIs has to do with their autonomy or 'free will'. AIs have been programmed to act in a certain way—why not, therefore, direct the responsibility at the programmers and/or owners of an AI? [17] This is a multifaceted issue. First, it is easy to overestimate the capacity of programmers to predict an AI's conduct, especially if the AI functions as a neural network that can learn patterns of behaviour independently. Even if such entities cannot be responsible in some thick, moral sense of responsibility, the deterrence rationale of punishment is certainly applicable to them. The kinds of AIs we are imagining are goal-directed, intentional beings, which could take legal sanctions into account. Furthermore, the programmers and/or owners can certainly *also* be held responsible if they have, say, intentionally created AIs that commit crimes. If they have done this unintentionally, they might in fact benefit from the criminal responsibility of their creations just as much as everyone else. A robot inclined to contravene basic legal norms might, after all, just as well determine that its owner is an obstacle to its goals. The responsibility

exceed the speed limits if following the limits would be dangerous. A self-driving car may end up in a situation where it may have to choose between, say, killing a passenger or two pedestrians (see MIT's Moral Machine website, which tests people's intuitions regarding such cases: http://moralmachine.mit.edu/ (accessed 5 October 2018)). A robot police officer might have to determine whether an armed robber will need to be shot, or whether leaving her alive poses a greater threat.

[15] One solution to this issue could be the AI's buying insurance, as Lawrence Solum points out. Solum (n 7) 1245.

[16] Chopra and White also propose the forcible modification of AIs as a similar kind of punishment. Chopra and White (n 7) 168.

[17] Alfred R. Mele has addressed questions relating to whether one is an autonomous agent if one's preferences have been determined by someone else. Alfred R. Mele, *Autonomous Agents: From Self-Control to Autonomy* (Oxford University Press 2001).

arrangement would therefore serve largely the same purposes as the criminal liability of slaves in the antebellum US South. Holding slaves criminally responsible protected both slave owners and third parties from slave crimes.

AIs as Commercial Actors

The tool–full–legal-person continuum introduced above is in commercial contexts often understood as a trifurcation:

(1) AI *qua* tool,
(2) AI *qua* representative,[18]
(3) AI *qua* legal person.

This trifurcation can very crudely be summed up as follows. Approach (1) treats an AI like any other piece of property in the owner's possession. The AI is like a word processor, used to draft a contract. Suppose that, because of a programming error, an AI omits an important clause from a contract. If the AI is treated as a tool, the owner might be able to sue the programmer, but the contract is most likely valid. If, on the other hand, the AI is treated as a representative, the contract might not be valid if the AI has acted outside its authority. Chopra and White mention 'induction errors, where a discretionary agent incorrectly inducts from contracts where the principal has no objections to a contract the principal does object to' as an eventuality that could result in the owner's preferring an agent–principal relationship.[19] Finally, endowing AIs with legal personhood would supposedly mean treating them 'as subjects of legal rights and obligations',[20] as entities with 'rights (and duties) of their own',[21] or as something 'to which the law can ascribe any Hohfeldian jural relation'.[22] It should come as no surprise that I am critical of such definitions of legal personhood.

The tool/representative/legal person trifurcation is, however, insufficient. We should distinguish two conceptual dimensions that underlie different AI legal personhood arrangements: separateness and independency.

Separateness pertains to a particular feature of legal platforms. A human being who founds a one-person corporation is in control of two legal platforms. Such platforms have three features.[23] First, they are named ('Mary' and 'Mary Inc.'); second, the legal positions within each platform are integrated (Mary can end up losing her house because of a contract she has entered); and third, the platforms are separate (Mary

---

[18] The term 'agent' is often used in this context, but it is ambiguous, which is why I will refrain from using it when referring to representatives.
[19] Chopra and White (n 7) 46. See also Pagallo, *The Laws of Robots* (n 11) 99.
[20] Chopra and White (n 7) 153.   [21] Pagallo, *The Laws of Robots* (n 11) 40.
[22] Shawn Bayern, 'The Implications of Modern Business-Entity Law for the Regulation of Autonomous Systems' (2015) 19 *Stanford Technology Law Review* 93, 94f.
[23] See Chapter 4.

Table 6.1 Two dimensions of AI legal personhood

S. Separateness	*1. Unity* AI completely part of owner's platform	*2. Partial separation* AI-controlled legal platform partially separate and revocable	*3. Total separation* AI-controlled legal platform completely separate and irrevocable
I. Independency	*1. Assimilation* Any exercise of competence by AI is treated as having been done by the owner/operator	*2. Dependency* Someone can, for example, retroactively cancel contracts made by AI	*3. Independency* Completely independent in exercise of competences

cannot normally end up personally liable for the debts of Mary Inc.). Legal positions controlled by an AI can be more or less separate from those of the owner of the AI. Two legal platforms are completely separate if, for instance, the debts pertaining to platform A can never be recovered from platform B and vice versa, and if the separation cannot be revoked. One example of partial separation is an AI-controlled corporation that is owned by a natural person. If the latter declares bankruptcy, the corporation's assets can be used to pay the creditors, whereas the creditors of the corporation do not have access to the natural person's assets.

The *independency* dimension pertains in particular to the exercise of competences: independent legal persons may normally exercise their competences without the supervision of anyone else, whereas dependent legal persons are subject to such supervision.[24] This dimension is distinct from the separateness dimension. For instance, in jurisdictions that allow for minors to own property, an adolescent's property is distinct from that of her father. In the absence of any fraudulent transfers aimed at evading the father's creditors, those creditors cannot access the adolescent's funds. Regardless, the father—assuming he is a legal guardian of his daughter—is normally able to exert some level of control over what his child chooses to do with her money. Thus, the two platforms are separate, but the adolescent is regardless not independent in her exercise of competences.

Table 6.1 presents three levels of separateness and independency each. Let us now focus on how these distinctions shed light on the tool/representative/legal-person trifurcation. Consider investment banks that employ AI traders for buying and selling

---

[24] This is partly based on Samir Chopra's and Laurence White's distinction between dependent and independent legal personhood. Samir Chopra and Laurence F White, *A Legal Theory for Autonomous Artificial Agents* (The University of Michigan Press 2011) 160–70. See also the discussion on legal competences in Chapter 4 and Visa A. J. Kurki, 'Legal Competence and Legal Power' in Mark McBride (ed), *New Essays on the Nature of Rights* (Hart Publishing 2017).

stock, derivatives, and so on. Given that the trading proceeds at a superhuman pace, most trades happen completely without human intervention. Regardless, such AIs are not legal persons: any trade they make is made in the name of the bank, and thus pertains to the bank's legal platform. The arrangement falls under S1 and I1: the AI does not have its 'own' funds that would be separate from those of the bank (unity), and the AI is also not treated as a representative of the bank, meaning that the bank is strictly liable for whatever contracts the AI chooses to enter (assimilation).[25] These two features sum up the AIs-as-tools arrangement.

The bank and the AI could also be in a representative–principal relationship. This would result in a somewhat different risk allocation, as the bank would not always be bound by contracts entered by the AI. A representative–principal relation would fall under I2 (dependency) and either S1 (unity) or S2 (partial separation). The latter would depend on whether the AI could be liable to a third party for acting outside the scope of its authority; if the AI could be liable, it would need to be able to own property in its own name.

Now consider a move toward complete separation. Shawn Bayern envisages a rather surprising scenario through which existing US law could already enable an AI to gain control of a limited liability company. I will not be concerned with the doctrinal accuracy of the scenario; it is regardless an interesting thought experiment. Bayern bases his argument on the 'process-agreement equivalence principle', according to which 'a legally enforceable agreement may give legal significance to arbitrary features of the state of any process (such as an algorithm or physical system) by specifying legal conditions satisfied by features of that state'.[26] A contract could, for instance, subject some contractual obligation to the behaviour of a dog or the temperature on a given day. Bayern contends that this principle can be extended to the performance of an algorithm:

Consider, for example, an artificially intelligent algorithm that passes the Turing Test in apparently acting roughly as a human acts. An agreement can, by specifying obligations and conditions, effectively delegate legal rights and decision-making powers to such an algorithm even though that algorithm is not a legal person. An agreement might say, for example, 'Your obligation to perform is discharged if the algorithm indicates X,' where X could be (for an unsophisticated algorithm) a formal output on a computer terminal or (for an artificially intelligent algorithm) something that approaches a description of human understanding and action (like 'that it is satisfied with the arrangement and physically signs a release form').[27]

In addition, Bayern notes that US company law allows unanimous shareholders to change the structure of a limited liability company as they like, even eliminating quintessential corporate bodies such as the board of directors. Assuming that these

---

[25] Pagallo, *The Laws of Robots* (n 11) 98. If the AI enters a very unsatisfactory contract, the bank may however be able to sue the designer of the AI.
[26] Bayern (n 22) 99.   [27] Ibid.

two premises regarding contract and company law are correct, one could form a corporation C, 'signing an "agreement" that specifies that C is to have no board of directors and instead shall take all legal actions determined by *A* (an autonomous system)'.[28]

We could, for instance, imagine a legal arrangement that allows for banks to create subsidiaries run by AIs using this procedure. Such an arrangement might be beneficial for the bank, as it could thereby reduce its risks: the bank would not be responsible for the financial liabilities of the subsidiary. However, Bayern notes that such an AI is not 'an autonomous legal entity', for the bank 'remains a shareholder and can continue to exert control over the entity'.[29] The arrangement would likely fall under S2 (partial separation), and either under I2 (dependency) or I3 (independency). If the subsidiary also had a board that could, say, retroactively cancel some large trades, then I2 would be the appropriate designation. If—for some reason—the bank were to completely 'recuse itself' from meddling in the trades of the AI, then I3 would be the apposite classification. The bank could of course still revoke the whole arrangement, but it would not be able to affect any individual trade.

The scenario just described resembles in many ways the *peculium* institution of ancient Rome.[30] If granted a *peculium* by the owner, a slave could own property, enter into contracts, and so on. The arrangement is somewhat difficult to summarize in contemporary terms: the slave owner did for instance receive the title to whatever the slave acquired and was liable to the slave's creditors, but the master's financial liabilities could not exceed the worth of the *peculium*.[31] The system resembled in many regards a limited-liability company: the master owned a slave, and could create a separate legal platform for that slave. The arrangement could be revoked by the slave owner at will, so a slave could normally keep his *peculium* for only as long as the slave owner perceived this to be in his own interests.

Now, were an AI to control a subsidiary in the described way, this would really only amount to a *peculium*.[32] Moreover—as with a *peculium*—the AI-run subsidiary would best be described as an extension of the owner's legal personhood, rather than as a legal person *tout court*. I am of course not denying that the arrangement would exhibit some features of legal personhood. However, one of the incidents of legal personhood is that one cannot be owned by anyone else, and the significance of that incident can be seen here. The Orthodox View, on the other hand, cannot explain as

---

[28] Ibid. 100.   [29] Ibid. 99.

[30] The similarities between *peculium* and certain aspects of AI legal personality are also noted by Ugo Pagallo in Pagallo, *The Laws of Robots* (n 11).

[31] Richard Gamauf, 'Slaves Doing Business: The Role of Roman Law in the Economy of a Roman Household' (2009) *16 European Review of History: Revue europeenne d'histoire* 331.

[32] Perhaps an even better comparison would be that of X's slave acting as the director of a company owned by X. Such an arrangement would not imply that the slave has become a legal person, even if the slave is in control of a legal entity, because the company is under the complete control of X.

easily why the AI subsidiary arrangement would not amount to 'full' legal person-hood. The AI could, for instance, normally decide on whether to sue over a debt or whether to waive it, thus endowing it with a will-theory right.[33] The AI would also certainly have duties distinct from those of the owner, given the limited liability of the subsidiary. The Orthodox View would therefore attribute legal personhood to the AI.

What has just been stated is not inconsistent with my claim that ordinary corporations—owned and run by a collectivity of human beings—can be legal persons *tout court*, distinct from their owners. First, one-person corporations are not legal persons; rather, they simply provide a new legal platform for their owner. In contrast, corporations with many members are not reducible to any single owner but are rather an exercise of collective intentionality, distinct from that of the participants. This collective nature grounds such corporations' distinct legal personhood. In addition, the arrangement cannot normally be revoked by any single member alone. The AI subsidiary arrangement, on the other hand, is revocable by its sole owner at any given moment.[34] This is why the subsidiary arrangement—though involving the use of incidents of legal personhood—does not endow the AI with legal personhood *tout court*.[35]

However, Bayern devises a method for completely detaching an AI-run corpor-ation from its members. According to the Uniform Limited Liability Company Act, a limited liability company can exist for up to ninety days without any members; this is to account for certain cases involving, say, the death of the only owner.[36] Bayern ar-gues that this ninety-day limit is in fact only a waivable default rule. The sole founder of C could therefore withdraw from C, creating 'a perpetual LLC [limited liability company]—a new legal person—that requires no ongoing intervention from any preexisting legal person in order to maintain its status'.[37]

Let us assume that the scenario described by Bayern is possible. I agree that it would result in legal personhood for the AI. Again, the situation could be compared to a scenario involving slaves. If—by means of a procedure relevantly similar to the one proposed by Bayern—a slave were to gain complete, irrevocable control of a corporation, she would indeed become a legal person, though a very special one: her

---

[33] I am assuming here that AIs can hold will-theory rights.

[34] The AI subsidiary could of course be owned by a group of distinct natural and/or arti-ficial persons, rather than by a single corporation. In this case, the AI would act as a sort of 'CEO', executing the collective intentionality of the owners. Once again, the AI itself would not be a legal person except in a *peculium*/representative sense.

[35] As I have noted above, the legal personality of children is different. First, even though they are subject to supervision in their exercise of competences, their legal platform is com-pletely separate from that of their legal guardians. The child's natural legal entity follows her from the cradle to the grave; a guardian cannot simply choose to subsume the child's platform into his or her own platform.

[36] RULLCA § 701(a)(3).   [37] Bayern (n 22) 101.

legal name would be, say, 'Mary Inc.', and she would for instance be unable to marry. What would happen to the corporation after her death would also be unclear (unless specified in the original shareholder agreement). Regardless, I see no reason to deny that this would constitute legal personhood for her; similarly, the arrangement depicted by Bayern would result in legal personhood for the AI in question. However, the AI arrangement would be somewhat different because the AI is not of ultimate value. This issue is intertwined with the question of whether AIs that are not ultimately valuable can hold claim-rights.

AIs and Claim-Rights Redux

I have maintained that ultimate value is normally a precondition for holding claim-rights. We can hold duties towards adults, infants, and animals because they are of ultimate value. However, I noted in Chapter 5 that human collectivities are a special case: their interests, though explicable in terms of human individuals, are not reducible to the individual interests of the members. I employed Raimo Tuomela's insight about the 'for-groupness' of the products of a group: if Mary buys some service from a group of friends, the agreed-upon payment is not owed to any individual member of the group but rather to the group itself. This is because the final recipient of the money is often not settled—the group may for instance decide to use the money to invest in its project, to divide it among the members, or to donate it to charity. But we should note here that the project itself is 'owned' by the members: they have committed to the project because they find it meaningful, and they likely have a say in how the products of the group project are to be used. Thus, their interests *qua* group members and *qua* individuals are in many ways intertwined. However, not all collective projects are necessarily 'owned' by the members in a similar way, and such projects will provide a way of understanding the claim-rights of AIs that are not ultimately valuable.

Consider again the foundation—introduced in Chapter 4—whose purpose is to preserve an old manuscript. Such a foundation can hold claim-rights. Now, let us say that all of the board members of the foundation are uninterested in its goals, with no personal stake or interest in whether the manuscript is in fact preserved. Regardless, we can say that when a board member carries out her tasks, it is in her interests (as an administrator of the foundation) that no one interfere with her work. One of the distinguishing features of such administrator-interests is that they can typically be transferred from one individual to another along with the duties and competences of the administrator.

Now suppose that all the workings of the aforementioned foundation were taken care of by a single AI. Even if the AI did not have any 'personal' interests that warrant the ascription of claim-rights, we could ascribe interests to it *qua* administrator of the foundation. The AI *qua* administrator would have an interest in preserving

the manuscript and in administering the assets of the foundation towards this purpose. Preserving the text was considered important enough by a human individual or a number of human individuals to warrant establishing a foundation. It is these interests that establish the claim-rights held by whoever or whatever—human or AI—that acts as the administrator of the foundation. The foundation's and the AI's interests are also clearly distinct. Imagine that, according to the rules of the foundation, the AI representing the foundation is replaced every five years (because of improvements in technology). As the old AI would not represent the foundation anymore, it would no longer be able to hold claim-rights.

The aforementioned applies not only to foundations but also to business corporations. An AI could control a business and promote its prescribed goals without having a 'personal' stake in the matter.[38] Consider, for instance, a real estate tycoon who wants to invest in companies in the Blackacre District in order to increase the value of his land there. However, he is an unpopular man and everyone else refuses to deal with him, hoping to prevent him from achieving too much influence in the area. As a solution, he creates the corporation AI Inc., invests a significant amount of money in it, and defines as its purpose that of promoting economic activity in the Blackacre District. He then 'sets the corporation loose' under the control of an AI, following the procedure described by Bayern above, in order to make clear that any deals made with AI Inc. do not benefit the tycoon directly. AI Inc. is now in many ways analogous to the foundation above: its purpose has been determined by the tycoon but is regardless now distinct from his personal interests. The interests of AI Inc. align with those of the tycoon, but only as much as with anyone else who owns land in the Blackacre District. The tycoon's and AI Inc.'s claim-rights can be distinguished. Let us say that Elisabeth has the duty to pay $5,000 to AI Inc.'s bank account because of her contract with AI Inc. The only party that can feasibly be said to hold the claim-right correlative to Elisabeth's duty is AI Inc. itself. Only AI Inc.'s goals are immediately thwarted if the duty is not fulfilled.[39]

To sum up, AIs can hold claim-rights as administrators of legal platforms with goals set by human beings. The correlative duties are not borne towards the AIs as 'private individuals' but rather as representatives of the legal platform they are administering. However, this analysis raises one further issue. In Bayern's scenario, an AI could gain control over a legal platform, and then—perhaps because of a programming error—use it for a purpose that is clearly detrimental for the creator of the AI, as well as for everyone else. Thus, it would no longer fulfil a project determined by a human being or a collectivity. Let us suppose that the arrangement would regardless not be revoked. Would the AI still be able to hold claim-rights? As long as the legal

---

[38] The typical purpose of a corporation is, of course, to generate profits for the shareholders, but in this case there are no shareholders so the purpose has to be something else.
[39] As with the claim-rights of collectivities, Bentham's test can be applied here. It will trim off beneficiaries such as the tycoon and any other beneficiaries in the Blackacre District.

system recognized for instance any contracts entered by the AI, treating it as a claim-right-holder would be the most intelligible option. I argued in Chapter 4 that idols cannot be legal persons because they are not of ultimate value; the relevant legal plat-form should rather be attributed to the administrator of the idol. But such options are not available here: no one else would qualify as the administrator or 'guardian' of AI Inc., as the shareholders and the executive board would have been removed. In addition, the legal institution of contracting would be quite incomprehensible if some contractual duties were not held towards anyone. Thus, the AI described here should still be classified as a legal person. However, the legal system would likely be committing a moral error when allowing such AIs to enjoy the protections of legal personhood.

Conclusion

The legal personhood of AIs is a topic that, in fact, covers numerous underlying issues. There are a staggering number of different possible legal personhood ar-rangements for AIs. The two-dimension analysis—distinguishing separateness and independency—serves to categorize some of these arrangements. It is, however, mostly restricted to the commercial setting.

It should be stressed that endowing an AI with the incidents of legal personhood that enable it to function as an independent commercial actor does not bespeak any acceptance of the notion that AIs are endowed with ultimate value. The legal per-sonhood of an AI can rather serve various purposes that might have nothing to do with the AI itself, such as economic efficiency or risk allocation. Of course, if some AIs ever become sentient, many of the questions addressed in this chapter will have to be reconsidered.