Abracadabra! Law, Language, and Agency in the Digital Real

Jason Grant Allen*

Abstract

This article explores the question of automation – and the apparent 'agency' of automation systems – as a problem for legal theory. It frames this inquiry by drawing certain parallels between magic, law and computer science that turn on the role of language in constituting 'worlds' or domains of reality as well as (just) presenting information. With a focus on current developments in distributed ledger technology and AI, it sets out an initial, programmatic sketch of an emerging approach to the digitalisation of law and legal reality, which could be termed 'digital realism'. This draws on various conceptual resources including Oxonian language philosophy, social ontology, institutional legal theory and Scandinavian Legal Realism. Together, these permit a view that explains the interaction of human and technological components in a complex and increasingly blended social fabric.

* Associate Professor of Law; Director, Centre for Digital Law (CDL), jgallen@smu.edu.sg.

Introduction: Finding Law's Magic

Science fiction writer and futurist Arthur C Clarke's 'Third Law' is that any sufficiently advanced technology is indistinguishable from magic.[1] Indeed, many of the technologies we take for granted today would only have made sense to our grandparents by reference to powers that operate above nature, outside the normal rules that govern actions in the physical universe. (Imagine a voice-activated 'smart' lock that was set to operate only when a certain speaker says: 'Open Sesame!') Of course, we know that computers are not magical, but it is instructive as well as entertaining to take a step back and ask what 'magic' is meant to be, how it works, and what the concept of magic might teach us about the interaction of law and digital technology. In this article, I explore both 'Code-Driven Law' and 'Text Driven Law' through the lens of the anthropology of magic, in particular the idea that a person can effect changes in the world through the proper incantation of words. This speaks to some of the central themes of the COHUBICOL Project[2] insofar as it examines the effect of digital documents that *appear*, at least, to perform acts in the world – including acts-in-the-law such as executing a contract.

Magic is often concerned with 'automation'. In Goethe's poem *The Sorcerer's Apprentice*, a young wizard-in-training summons supernatural help to perform his chores while the master is away.[3] A broom and a pile of rags are animated into an automaton, which (or, to beg the question: who) fetches pails of water to fill the young wizard's bath.[4] We will return to the hapless apprentice in due course, but it is helpful first to look at what is implicitly claimed in stories about magic like this one. For classical anthropologists, the premise of magic is that words have *effects in the world*.[5] D.A. Westbrook explains:

> Following Malinowski and Tambiah… 'magic' is an instrumental practice, a practice that presupposes a world upon which the magician can, by following the right procedures, effect his will. The magician calls upon powers to which the world must bend. Thus magic assumes a rationality, *a logic or procedure, or grammar*, even though such logic may not be understood, or even clearly discerned, by the uninitiated.[6]

On this view, magic constitutes the recruitment of supernatural powers to intervene causally in brute nature. These powers are recruited through *language*: incantations of words can hurt and heal bodies, open doors, animate objects, control wild animals and violate the ordinary laws of nature in various other ways. For example, in the Jewish magical tradition, the Divine Name written on a piece of paper is said to animate a *golem* – which literally means 'lump of shapeless matter' – into a humanoid, capable of directed action.[7] As the tradition goes, a *golem* is capable of many things, but not of the full range of human action – in particular, a *golem* is not gifted with the capacity for speech, reflecting both the limited capacity of its human creator to channel the powers of the ultimate Creator. (Conveniently and importantly, this limitation avoids the situation where one *golem* produces another.)

The central idea in these approaches to the concept of magic is that words play a constructive, constitutive role in the world. One self-identified practicing magician (with a handy training in semiotics) explains a possible Kabbalistic origin for the familiar, but obscure, word 'abracadabra', which features in the title to this paper:

---

[1] Arthur C Clarke, *Profiles of the Future: An Inquiry Into the Limits of the Possible* (Harper & Row 1973) 14, 21, 36.

[2] https://www.cohubicol.com/.

[3] Goethe's poem, and some of the problems illustrated in it, were discussed by Norbert Wiener, 'Some Moral and Technical Consequences of Automation' (1960) 131(3410) Science 1355.

[4] Many of us are familiar with Walt Disney's classical interpretation of the traditional story in the animated film *Fantasia*. Goethe was probably inspired by an ancient Greek author, Lucian: See Gerard Alston Reichling, 'A Demotic Folk-Tale the Basis of Goethe's "Der Zauberlehrling"' (1919) 39 Journal of the American Oriental Society 295. I have also seen suggestions that Goethe was inspired by the story of the *golem* of Prague.

[5] See Bronislaw Malinowski, *Magic, Science, and Religion, and Other Essays* (The Free Press 1948); Stanley J Tambiah, 'The Magical Power of Words' (1968) 3(2) Man 175, 184.

[6] David A Westbrook, 'Magical Contracts, Numinous Capitalism' in Brian Moeran and Timothy de Waal Malefyt (eds), *Magical Capitalism: Enchantment, Spells, and Occult Practices in Contemporary Economies* (Springer International Publishing 2018) 62. Empahsis added.

[7] Joshua Trachtenberg, *Jewish Magic and Superstition: A Study in Folk Religion* (University of Pennsylvania Press 2012) 80,84.

One possible origin for the word is an Aramaic phrase: *avra kedabra*. This sentence means 'I create as I speak'... The verb *avra*, 'I create', in Aramaic is cognate with the verb *evra* in Hebrew. But Hebrew makes a distinction between different types of creation: if you make something from nothing, you say 'evra', but if you make something from something else, you say 'etzor'. So abracadabra is saying 'I create something from nothing when I speak, just as my words come from nothing'.[8]

Accepting this story for the sake of argument, at least, there is an implicit nominalist ontological (and possibly metaphysical[9]) claim here: by *naming* something we *create* it, calling it into being from the primordial chaos of potentiality.[10] If we know the right words (semantics) and grammar (syntax), we can organise (and arguably, create) matter itself.

Of course, words do not really act on physical objects directly in this way. I can shout 'Open Sesame!' at a door all I want – it will not budge. But it should not be controversial to say that words do have immense constitutive power in domains of ritual and convention and it should not be controversial to identify the law as such a domain. Legal words have invisible effects that are not unlike magic,[11] causing changes in institutional legal reality. Saying the words 'I promise' creates contractual 'rights' and 'duties' out of thin air.[12] Writing words on a piece of paper might not animate a clay *golem*, but the correct words will transform it from an inert scrap into a legally potent 'warrant'.[13] As J.H. Sommer observes in the law of financial instruments, *negotiability* 'is almost purely a question of magic words, not external facts.'[14] If we get the spell right, we create an artefact that moves between transacting parties in the world of law, leaving a wake of invisible effects (namely: newly minted rights and obligations) behind it. Contracts themselves are magical, argues Westbrook, or at least 'nearly so'. Using a 'contract spell', he argues, '[l]egal persons (not just anybody, indeed often not a person) can... invoke an invisible power (the law) in order to change local facts.'[15] Our institutions project visions of legal reality and the coercive power of the community is brought to bear to bring the world into conformity with it – to ensure isomorphism between the *de facto* state of affairs and *de jure* state of affairs that we have, collectively, spun from words.

---

[8] Patrick Dunn, *Magic, Power, Language, Symbol: A Magician's Exploration of Linguistics* (1st ed, Llewellyn Publications 2008) , xi. The creation of something from nothing in the Kabbalistic tradition has to do with the 'contraction' of the Infinite Creator to make room for finite 'formed' creation and the paradox of the Infinite remaining unchanged by the act. Pompeu Casanovas has drawn my attention to the fascinating parallel between my own discussion of 'vases' in the context of D Alan Cruse, 'Some thoughts on agentivity' (1973) 9(1) Journal of Linguistics 11 and the 'shattering of the vessels' in Kabbalistic thought (whereby the 10 manifestations of the Infinite in creation shattered and cracked the vessels of matter), however I am unable to develop that parallel here. See Mario Macías, Pompeu Casanovas, and John Zeleznikow, 'Shevirat Ha-Kelim: Jewish Mysticism and the Catalan Matrix for Dialogue and Conflict' [2020] Journal of Catalan Intellectual History 64.

[9] The relationship between 'ontology' and 'metaphysics' is sometimes drawn quite differently, and the matter has a complex intellectual history. For present purposes, I adopt a certain – not uncontroversial – approach to 'ontology' and 'social ontology' in particular which embraces many of the things typically associated with metaphysics, namely, the study of the *mode of existence* of what exists, not just the classification of what exists. This is consistent with my use of John Searle's brand of 'social ontology' in particular, which I find most accessible and operationalized for the purposes of legal theory and law and technology studies; this does not amount to a wholesale endorsement of Searle's scheme, or a wholesale rejection of other approaches to (social) ontology and metaphysics.

[10] God creates light *ex nihilo* by 'saying' *fiat lux*: See Jonathan Yovel, 'The Creation of Language and Language without Time: Metaphysics and Metapragmatics in Genesis' (2012) 20(3) Biblical Interpretation 205. *Genesis uses* אמר *' while abracadabra putatively derives from* רכר *' the difference reflects that between 'to say' and 'to speak'.*

[11] See Alois von Brinz, *Lehrbuch der Pandekten* (Second Edition, Deichert 1857); Lars Lindahl, *Position and Change: A Study in Law and Logic* (Springer Netherlands 1977) 211.

[12] See Neil MacCormick and Ota Weinberger, *An Institutional Theory of Law: New Approaches to Legal Positivism* (vol 3, Springer Netherlands 1986) 37; Dick WP Ruiter, *Legal Institutions* (Kluwer Academic Publishers 2010) 8.

[13] See H Klinghoffer, 'Jurisdiction and Illegality. A Study in Public Law. By Amnon Rubinstein. [Oxford: Clarendon Press, 1965. xii + 250 pp. 45s net.]' (1966) 1(3) Israel Law Review 510, 8 citing *The Duchess of Kingston's Case* (1776) 20 St. Tr. 355, *Holdsworth v Lane* (1841) 1 QB 18, and *Corrat v Morley* [1962] 2 AC 147.

[14] J.H. Sommer explains the effect of the words 'to order' and 'to bearer' under the Uniform Commercial Code Art. 3 to argue his nominalist (versus realist) ontological approach to electronic payment systems: Joseph Sommer, 'Where is a Bank Account?' (1998) 57(1) Maryland Law Review 1, 62.

[15] Westbrook (n 6) 47.

The Magic of Modern Computer Systems

Here, I wish to argue that ideas from the anthropology of magic can help us to understand certain aspects of (and problems raised by) advancing digital technologies. We should not push this parallel too far, but the sections to follow will argue that a legal system is a projection of an essentially fictional state of affairs onto the world in which we can 'act' through linguistic means (i.e., through speech acts and the creation of documents memorialising them).[16] As Laurence Diver observes:

> Code text is 'performative', creating a new state of affairs in the world when it is executed, but also is also documentary, both describing what will happen and telling us something meaningful about the conditions of its production… code could be said to be a kind of speech act: it is a written text with latent performativity that creates a new state of affairs in the world when it is executed.[17]

As we use sophisticated computer systems to represent this system of 'augmented legal reality' and as those systems interact with other domains of social reality and indeed with physical reality in more complex and direct ways, we are witnessing a new and powerful form of techno-legal 'magic'. As this section elaborates, my view rests on a certain view of law that stresses the role of speech acts in the creation of legal (and other social) institutions. As the 'world of law' is spun from digital words and documents instead of just spoken words and documents written on paper (or some comparably static medium of inscription), interesting things start to happen. First and foremost, the law need not always address human agents to have effect on the physical world. The typical mode of action of the law is to stipulate a norm, which I take to be a proposition that x (rather than y or z) 'ought' to be the case, such that some agent *makes* x *the case* through intentional action pursuant to the norm. By responding to legal 'spells', in a sense we are conscripted to carry out the legislator's purposes. However, we do not thereby become *golems* – automata who are, ourselves, speechless. One of the main differences between law and magic is that the law generally presumes human agents, whereas magic is claimed to work directly on the physical world.[18] Even a 'functionary', such as a commissioner, police officer, or tax official, is not a *golem* because, although they carry out the directives of the legal system, they do so through their natural capacities for thought and action and these are part of what keeps the legal system 'human'.

As machines can be programmed to respond automatically and intelligently – in increments, more *autonomously* – the law (and those that lay it down), need not address human subjects as its middlemen anymore. This highlights the importance of algorithms that *seem* to perform legally relevant acts without (further) human input: they take the law one step towards the magical by replacing human agency with agency *ex machina*. In certain limited contexts, such as financial trading, it is already arguably the case that algorithms are performing (apparent) acts-in-the-law such as forming (as well as performing) contracts with each other and with humans, who may or may not be aware that the 'agent' on the other end of the bargain is a bot initiated by their ultimate contractual partner.[19]

As Diver elaborates, the consensual threshold for execution of code is much lower, and in many cases non-existent; as mentioned above, once execution has been initiated, there is little in the way of scope for mitigatory interpretation or collective agreement not to recognize or act upon the 'performative' – there is no official to refuse enforcement of an encoded rule. This, in turn, 'up-ends law's scheme of performativity, which is built upon institutive rules that

---

[16] I am grateful to Jaap-Henk Hoepman for our discussion on this point.

[17] Laurence Diver, 'Interpreting the Rule(s) of Code: Performance, Performativity, and Production' [2021] MIT Computational Law Report; Brian Moeran and Timothy De Waal Malefyt (eds), *Magical Capitalism: Enchantment, Spells, and Occult Practices in Contemporary Economies* (Springer International Publishing 2018) 3,12.

[18] Westbrook (n 6) 47.

[19] See, e.g., the discussion in Simon Gleeson, 'The Legal Consequences of Automated Mistake' in Jason G Allen and Peter Hunn (eds), *Smart Legal Contracts: Computable Law in Theory and Practice* (Oxford University Press 2022) Ch 19.

[20] Diver (n 17) 12.

require the recognition of a community in order to have any practical effect in the world.'[20]

Within the compass of a human lifetime, the notion of a rich-state Turing machine has moved from a heretical theory to an everyday, even throw-away, commodity.[21] These machines have been networked extensively over the past three decades, in ways that have affected every aspect of human interaction and justify speaking of a 'networked society'. Network protocols such as Ethereum purport to create 'stateful' Turing machines across the globe and to use these machines to animate novel forms of associational behaviour. What we are now witnessing is the entwinement of law and computer systems to form a complex socio-technical 'stack' that appears even more magical than before. Worlds governed by axiomatic (human-stipulated) rules, not the laws of physics, are brought to life daily, in richer ways in modern computer hardware systems and these intermesh with the 'world(s) of law' inextricably in the context of everyday economic and financial transactions, as well as in a growing sphere of social and political life. These changes to the very fabric of social reality demand new theory and this article could be understood as a programmatic sketch of the kind of theory that is required. It stresses the interaction between different layers – human and non-human, conventional and novel, social and technical - in the 'stack'.

The Epistemic and Ontic Functions of Computer Systems

The study of human-computer interactions has been described (perhaps not accurately)[22] as falling into three 'waves', each informed by a different 'metaphor of inter-action' that operationalises different epistemological and ontological commitments.[23] According to DJ Gunkel, the first wave was characterised by human control of computational mechanisms and its chief concern was the best way to design input and output affordances to facilitate effective human/machine couplings. The second wave shifted this focus from questions of efficient control and ergonomics to computational capability and information processing and transmission. Its chief concern was the flow of information into and out of the device, and the transformations in data that occur in this process. It was, therefore, concerned with 'communication' in the sense implied in the information theory of Shannon and Weaver.[24] The third wave represents a *phenomenological turn*: the chief concern is 'interaction as phenomenologically situated'. Human-computer interaction supports 'situated action in the world' and questions concern 'how to complement formalized, computational representations and actions with the rich, complex, and messy situations at hand around them.'[25]

Gunkel thus characterizes the third wave as a kind of inversion of the others; it is concerned not with the ontological capabilities of the two interacting components (human user and the computational artefact) as 'subject' and 'object', but with the relationship that is situated between them.[26] The turn is 'phenomenological' because its aim is to determine less what something *is* than what it *appears to be*.[27] In the domain of *social* reality, if agents define a situation as real, it is real in its consequences.[28] This is a crucial point. Assemblages of human users and computational artefacts are now representing states of affairs as existing in virtue of being represented in a digital sys-

---

[21] See Alan Turing, 'Intelligent Machinery, A Heretical Theory' (1996) 4(3) Philosophia Mathematica 256; Alan M Turing and Jack Copeland, *The Essential Turing: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life plus The Secrets of Enigma* (Oxford University Press 2004); Marvin Minsky, *Computation: finite and infinite machines* (Prentice-Hall 1967) Chapter 8.

[22] In particular, Gunkel's 'third wave' may have begun much earlier, even in the 1990s: see, e.g., Philip Robbins and Murat Aydede (eds), *The Cambridge Handbook of Situated Cognition* (Cambridge University Press 2008) and William J Clancey and others, 'Brahms: simulating practice for work systems design' (1998) 49(6) International Journal of Human-Computer Studies 831.

[23] David J Gunkel, 'The Relational Turn: Third Wave HCI and Phenomenology' in Michael Filimowicz and Veronika Tzankova (eds), *New Directions in Third Wave Human-Computer Interaction: Volume 1 - Technologies* (Springer International Publishing 2018) 12.

[24] Claude E Shannon and Warren Weaver, *The Mathematical Theory of Communication: Volume 1 - Technologies* (University of Illinois Press 1963).

[25] Gunkel (n 23) 12.

[26] ibid 13.

[27] ibid 15.

[28] Mireille Hildebrandt, 'Law as Information in the Era of Data-Driven Agency' (2016) 79(1) The Modern Law Review 1, 26. This notion has been variously expressed over the centuries. See e.g. Thomas Hobbes, *Man and Citizen* (Bernard Gert and Thomas Hobbes eds, Anchor 1972) XV:13.

tem. Digital assets, for example, exist because some digital ledger represents them as existing and some human users treat those quasi-mathematical objects as 'assets' with value.[29] In their autonomous or automatic nature,[30] so-called smart contracts now perform agreements within such domains of social reality and therefore assume a kind of agency themselves. When aggregated into a 'nexus' for joint action, they create a kind of group agency that deserves theoretical attention.[31]

As Philip Brey explains, computer systems have ceased to provide only 'epistemic' functions (i.e. the representation, storage, and manipulation of information as an augmentation of our central nervous system). The epistemic relation between humans and computers has been supplemented by an 'ontic' relation in that computer systems are able to simulate virtual and social environments that extend the interactive possibilities found in humans' physical environment. They are 'portals to the worlds we inhabit' that provide visual and sonic outputs to the sensory perceptions of a human user.[32] The most obvious example of this (from the user perspective at least[33]) is in the rich game-worlds that have become commonplace in recent decades,[34] but it now goes much further. Drawing on John Searle's social ontology, Brey explains:

> Since the 1990s, many computers have become part of large-scale computer networks and specifically the Internet. *When computers become networked, and they are used as means for social interaction by users, social structure emerges.* Internet, specifically, has given rise to an online social reality that includes social roles and statuses, groups and organizations, institutions and social events, where most or all of this social structure is *realized digitally...* Internet users can make it true that certain displacements of bits on the Internet count as buying, gambling, marrying, signing a contract or winning a game by *assigning the appropriate statuses*

to digital structures and events. Since the 1990s, this has occurred on a large scale, with the result that cyberspace now contains evolved social structures that collectively define new social realities for their users. These social structures need not be represented graphically and are often represented by means of text and simple buttons or windows. But like graphical representations, they contribute to the ontic function of contemporary computers: *they represent interactive objects with which users interact and that are part of their everyday ontology.*[35]

This points to the possibility of developing deeper theoretical parallels between the disciplines of legal theory and computer science. In object-oriented programming languages, for example, *classes* of entity are stipulated to exist in the relevant *universe* and are then organised into a *hierarchical class structure*. Entities within a class, *objects*, are ascribed *methods* (behaviours or capabilities) consistent with their being an instance of a class (in the relative class structure). In law too, we effectively stipulate that a class of entity (e.g., 'persons') exists and then we assign objects (many of them artificial, in the sense of being the product of human artifice and not entities of brute creation) to that class in a hierarchical schema. Entities (e.g., different types of persons such as 'natural persons' and 'juristic persons') have different abilities within the 'universe of law' which we map as legal 'positions' or 'bundles' of jural relations such as rights and correlative duties. Traditionally, we have used words, physical symbols (such as regalia and uniforms) and paper documents to represent social reality. As we use (for example) object-oriented programming languages that can execute functions, as well as represent information, in some runtime environment, new dimensions open up for the scale, speed, and complexity of our social world.

---

[29] See, e.g., Jason G Allen, 'Property in Digital Coins' (2019) 8(1) European Property Law Journal 64.

[30] Kevin Werbach, 'Trust, but verify: Why the blockchain needs the law' (2018) 33(2) Berkeley Technology Law Journal 487, 507.

[31] See Jason G Allen, 'Bodies without Organs: Law, Economics, and Decentralised Governance' (2020) 4(1) Stanford Journal of Blockchain Law & Policy 53, 53.

[32] Philip Brey, 'The Epistemology and Ontology of Human-Computer Interaction' (2005) 15(3) Minds and Machines 383, 383.

[33] See the discussion in ibid 393-394 which I have omitted here for the sake of brevity.

[34] See Jason G Allen, 'Law's Virtual Empires: Game Analogies and the Concept of Law' in Jorge Luis Fabra-Zamora and Gonzalo Villa Rosas (eds), *Conceptual Jurisprudence: Methodological Issues, Classical Questions and New Approaches* (Springer International Publishing 2021) Chapter 15.

[35] Brey (n 32) 396.

Doing Things With Algorithms

The idea that law *does* things is critical to understanding law: legal norms attach legal effect to specific acts, behaviours, states or occurrences We generate 'real' legal consequences by defining legal conditions. This in turn implies that law is not merely *information about* such consequences, it also transforms our reality; for example, '[i]f specific legal conditions are fulfilled, a couple is married or a person becomes the owner of a house.'[36] This identifies our present question: how does the law operate in the rich virtual extensions of reality afforded by modern computer networks? And are humans still the only entities that could be described as 'subjects' there?

This ability of words to *do* things as well as to *say* things is of interest to philosophers of language and many of their examples are drawn from law, ritual, religion, and magic. JL Austin observes that many of the 'acts' which concern the jurist are or include the utterance of 'performatives' – utterances intended not to assert the truth or falsity of a proposition but to *do something in virtue of being spoken.*[37] Before Austin's lectures, the Munich phenomenologist of law Adolf Reinach observed that when we perform certain types of act, provided that certain conditions are satisfied, the *world itself changes.*[38] The paradigm example of such an utterance is, of course, the contractual promise – saying that I promise to do something is a 'performative' speech act to the extent that it serves an outward-looking, social function, namely creating a contractual relationship. Such

an utterance may be 'unhappy'[39] (because mistaken, under duress, etc.), but the function of it is less to tell the world that I (truly) intend to buy the thing as to create a new juridical fact (i.e., a binding contract) on the basis of my performative utterance.

Paul Amselek has observed that the theory of speech acts (or 'linguistic phenomenology', as Austin called it) opened up new possibilities for legal philosophy from its very outset.[40] Austin was, of course, a major influence on the doyen of 20th century English legal positivism, HLA Hart.[41] One of the leading successors to Austin, in turn, is John Searle, who has expanded on the concepts of speech act theory to create a rich school within the branch of analytic metaphysics known as social ontology.[42] Perhaps in virtue of drinking from the same intellectual fountain, in my view Searle's investigations of what he calls 'institutional facts' are particularly ripe (and oddly overdue) for integration into the didactically familiar school of Oxford legal philosophy. Searle's work has, in turn, found fruitful application in analytical jurisprudence, especially in the 'institutional legal theory' of scholars working in Hart's tradition, including most notably Neil MacCormick.[43] This is perhaps due to, and enhanced by, MacCormick's proximity to another school peculiarly concerned with the apparent mysticism of law and the need to reconcile the outlandish 'facts' of the legal universe with a basically materialist metaphysics: Scandinavian Realism.[44] This basket of resources – Oxford linguistic philosophy, Hartian neo-positivism, Searleian social ontology, Scandinavian Realism and institutional legal theory – provides my preferred toolkit for making sense

---

[36] Hildebrandt, 'Law as Information in the Era of Data-Driven Agency' (n 28) 9.

[37] John L Austin, *How to do things with words* (Second edition, Harvard University Press 1975) 19.

[38] Barry Smith, 'How to Do Things with Documents' [2012] (50) Rivista di estetica 179, 181. See Adolf Reinach and John Crosby (eds), *The Apriori Foundations of the Civil Law* (De Gruyter 2012) 9.

[39] In Austin's theory of performative speech acts, an utterance is 'unhappy' (also 'infelicitous') where something is wrong with them; because performative, as opposed to constative, utterances have no truth claim, they cannot be true or false, but they can be unhappy.

[40] Paul Amselek, 'Philosophy of Law and the Theory of Speech Acts' (1988) 1(3) Ratio Juris 187, 187.

[41] See, e.g., Herbert Lionel Adolphus Hart, *The Concept of Law* (Oxford University Press 1961).

[42] See, e.g., John R Searle, *Making the Social World: The Structure of Human Civilization* (Oxford University Press 2010).

[43] See Neil MacCormick, *H.L.A. Hart* (vol 93, Stanford University Press 1981). For examples of institutional legal theory, see MacCormick and Weinberger (n 12); Carlos Bernal, 'Collective Intentionality and the Ontological Structure of Law' (2014) 45(3) Rechtstheorie 335; Carlos Bernal, 'In search of the ontological structure of law' [2012] Analisi e Diritto 59; Jaap Hage, 'The (Onto)logical Structure of Law: A Conceptual Toolkit for Legislators' in Michał Araszkiewicz and Krzysztof Płeszka (eds), *Logic in the Theory and Practice of Lawmaking* (Springer International Publishing 2015); Ruiter (n 12).

[44] See, e.g., Alf Ross, 'Tû-Tû' (1957) 70(5) Harvard Law Review 812; see also Jakob v HHoltermann, 'A Straw Man Revisited: Resettling the Scope Between HLA Hart and Scandinavian Legal Realism' (2017) 57 Santa Clara L. Rev. 1. Other important scholars in this school are Karl Olivecrona and Axel Hägerström. See Gregory S Alexander, 'Comparing the Two Legal Realisms-American and Scandinavian' (2002) 50(1) The American Journal of Comparative Law 131.

of new technological developments as they challenge conventional doctrinal categories. From them, we can begin to build a coherent view of current developments that explains how machines are intervening in conventionally 'human' social phenomena and processes including and especially, the law.

In the passage cited above, Brey applied some concepts from Searleian ontology to explain how 'objects' represented in computer systems come to be objects of social relations and therefore come to be part our social reality.[45] In Searle's account, institutional facts are *epistemically objective* but *ontologically subjective.* That is, we can say that it is 'true' as a matter of 'fact' that I owe $10 or that you are 'married', but these are only 'facts' because people think they are true. Such facts have a very different ontology to 'facts' of the brute universe that refer, for their truth, on mind-independent phenomena. According to Searle, 'institutional facts' are created through the imposition of a *status function* on a brute fact, such as a brute object. The imposition of a status *function* transforms the brute input into a *legal output.* This occurs when we treat a woman as a 'wife', with all that entails,[46] a person who is 18 years + one day old as an 'adult' or a group of human beings as a 'corporation'.

A status function, for Searle, is constituted by a bundle of *deontic powers*, which includes rights, duties, etc. Institutional facts are created by what Searle calls *status function declarations*, i.e. speech acts that stipulate that X (e.g., a group of co-adventurers) are to 'count as' Y (e.g., a corporation) in the relevant context C. This scheme, which Searle argues can be applied in a general fashion to explain the whole structure of social reality, draws on his previous work in the philosophy of mind and in the philosophy of

language. It is in virtue of the representation of X as Y in C *by a reasoning, intentional agent* that creates the deontic powers that constitute the institutional fact. These powers are invisible, ontologically subjective, they exist as psychological dispositions only. However, this does not make them less than 'real', it merely means that they bear the mode of existence of *institutional facts*, facts by institution as it were.

Declarations, in turn, have what Searle calls a *double direction of fit.* Direction of fit is a concept from the philosophy of mind, to do for example with the nature of mental states such as belief and intention.[47] Austin probably coined the phrase – the concept is much older – when he distinguished between fitting a name to an item and fitting an item to a name.[48] Searle and Vandervaken later applied the notion to a taxonomy of speech acts, distinguishing between speech acts with a *word-to-world* direction of fit, a *world-to-word* direction of fit and a *double direction of fit.*[49] In the first case, an utterance is successful if its propositional content fits an independently existing state of affairs in the world (e.g. 'you are married'). In the second case, an utterance is successful if the world changes to fit its propositional content (e.g. 'you need to get married'). And in the third case, the utterance is successful if the uttering of it changes the world to fit its propositional content (e.g. 'I hereby declare you married'). By asserting a belief about the state of the world, declarations purport to alter the state of the world to match the propositional content of the declaration.[50]

---

[45] Searle's account is set out in John R Searle, *The Construction of Social Reality* (Free Press 1995); Searle, *Making the Social World: The Structure of Human Civilization* (n 42) The latter provides a more concise explanation.

[46] For example, in 19th century English law, the doctrine of couverture entailed that the woman lost the ability to perform certain acts-in-the-law, such as making a 'deed', independently of her husband: see *Whitmarsh v Robertson* (1845) 1 Coll 571, 575.

[47] See, e.g., John R Searle, *Intentionality: An Essay in the Philosophy of Mind* (Cambridge University Press 1983) 7. See also Lloyd Humberstone, 'Direction of Fit' (1992) 101(401) Mind 59 for discussion.

[48] John L Austin, 'How to Talk: Some Simple Ways' (1953) 53(1) Proceedings of the Aristotelian Society 227.

[49] John R Searle and Daniel Vanderveken, *Foundations of Illocutionary Logic* (Cambridge University Press 1985) 52.

[50] The validity of a declaration could be postulated to be governed by the accepted rules of the context. We do not (always) write on a blank slate; in the present context, status function declarations about corporate personality (for example) will derive 'validity' by reference to the aggregate of existing declarations (and other forms of utterance such as directives made by constituted institutions), but that aggregate body of rules will not be completely static, and is subject to change according to the same logical structure.

From Speech Acts to Document Acts to Dynamic Documents

Lawyers attempting to come to terms with technological developments need to adopt (or develop) a more sophisticated understanding of documents, building on the theory of speech acts. The result of our efforts, I think, will be a better theory of what I call 'dynamic documentality'. A concept of dynamic documentality would incorporate the learning sketched out in the preceding sections on documentary speech acts with the learning on the ontic function of digital communications systems.

The leading theorist of documentality *per se* is Maurizio Ferraris. To start from the beginning, Ferraris asks what makes social objects 'objects' at all. Some philosophers, he argues, would insist that social, mind-dependent phenomena (like corporations and other legal artefacts) are not 'objects' at all, for want of physicality; others would insist that even physical objects are socially constructed because they are the result of our theories about the world.[51] Taking a *via media*, Ferraris stresses the importance of the 'inscription' of performative acts for the creation of social objects. These are distinct from brute objects – mountains and lakes, trees and atoms exist even if we don't think about them, and they would still exist even if no human had ever inhabited earth – and from apparently Platonic objects ('ideal objects existing out of space and time, but still independently of subjects', such as numbers and theorems). Social objects rely for their existence not only on human beings but also on 'inscriptions'.[52]

Ferraris offers a series of examples, ranging from the nation state to the automobile conglomerate Fiat, to a conventional telephone company (with landline infrastructure), to a mobile phone company (with satellite and tower infrastructure), to the debt of a multinational dairy concern embroiled in a fraud conspiracy. The question 'where is the being of a debt', he observes, is the hardest of all, as there is no obvious physical element to a debt, which appears as a 'negative entity'.[53] but entities such as debts do exist insofar as and because, they are tethered to physical reality through inscription:

> There is also in this case, to speak the truth, a physical object, although it is rather a peculiar one: it is something written on a paper or a computer, or simply in the mind of someone. This is the secret of social objects. [W]e do not need massive physical objects (a territory, a human body, a piece of paper or a coin) to obtain a social object: in many cases a few ink molecules, or some neurons, are enough… [S]ocial objects depend on social acts concerning at least two persons – two persons that should be able to remember what they do – and an inscription is the only physical thing it takes. The inscription, *viz.* the trace, is a concept on which Derrida insisted a lot: without inscriptions of some kind, in the head at least, there would not be society, although, obviously, there may be inscriptions (for instance the notches on Robinson [Crusoe]'s calendar) without there being a society… [T]his condition holds specifically for social objects: physical objects, like mountains, or ideal objects, like theorems, exist without inscriptions, but social objects don't (a society without memory is, strictly speaking, unconceivable). It is in this sense that I propose to transform Derrida's thesis (untenable as it is) 'There is nothing outside of the text' into 'There is nothing social outside of the text'.[54]

In Ferraris' account, inscription covers everything broadly: from human memory ('traces' in the neuronal structure of a human brain) to paper writing to digital recording. And as more advanced societies have greater needs for documentation, they generally create more resources – today in the form of digital technologies – to meet that need.

Although I would not endorse Ferraris' theory wholesale, it provides an essential and often-ignored inroad into un-

---

[51] Maurizio Ferraris, 'Cultures. Conflict - Analysis - Dialogue' in Christian Kanzian and Edmund Runggaldier (eds), *Proceedings of the 29th International Ludwig Wittgenstein-Symposium in Kirchberg, Austria* (De Gruyter 2007) 386.

[52] ibid 387.

[53] (ibid 396). As an anonymous Editor of the Yale Law Journal observed in 1925, the idea that a thing ('*res*') is needed to enliven jurisdiction led to the 'fallacy of hypostatisation', to the 'tendency to reify legal relations wherever the court [decided] to exercise jurisdiction.' Herman Thomas Stichman, 'Jurisdiction and the "Situs" of Debts' (1925) 34(6) The Yale Law Journal (Leon Arthur Tulin ed 652. See also Sommer (n 14); Pippa J Rogerson, 'The Situs of Debts in the Conflict of Laws—Illogical, Unnecessary and Misleading' (1990) 49(3) The Cambridge Law Journal 441.

[54] Ferraris (n 51) 397, 399.

derstanding digital record-keeping systems and the social artefacts to which they give rise. Although the concepts of speech act theory (set out above) have become commonly accepted, qualitative changes occur when a society moves from oral speech acts to documentary acts. Documents are self-contained, complete and they can endure self-identically over a period of time in isolation from the individuals who participated in their creation. Crucially, documents can be combined into more stable and sophisticated 'document-complexes' that enable extended forms of social organisation such as joint stock companies and global insurance schemes.[55]

Ferraris' account of documentality runs into problems when applied to digital documents that are dynamic rather than static.[56] Paper documents are an inscription that is made at a point in time (T_0) and that remains the same unless it is altered by a human being, which effectively means its destruction and the creation of a new document at T_1. First-generation digital documents, such as Portable Document Files ('PDFs') are essentially digitised versions of these static documents. But documents have become dynamic too, as David Koepsell explains in the context of blockchain technology:

> Blockchains like Bitcoin... are 'dumb' in the sense that they do nothing but record verified transactions. They cannot 'act', they can only be acted upon. Some blockchains allow not only the recording of static entries, but they also allow instructions about how to deal with certain types of new entries, entities, or other inputs such that they can execute actions based upon logic embedded in those entries. These blockchains are the foundation for what some call 'Web 3.0', where the full potential of the 'semantic web' means that automated business and social processes can execute logic across the web's distributed architecture.[57]

As Koepsell elaborates, dynamic blockchains such as Ethereum make an entire layer of electronic expression computable; phenomena previously requiring intentionality and some 'meeting of minds' may now, it seems, be programmed and unleashed as automated processes. One task facing us, as theorists, is what all of this means for legal theory and practice in the coming decades.

We can draw parallels between modern developments and the impact of advances in transportation, advances in record-keeping and the advent of the limited liability joint-stock company in the 18th century.[58] An oft-cited example of documentation acting like leaven in the early modern economy is double-entry book-keeping. Like so many technological innovations, this is something that came from the East but was instrumentalised with remarkable effects in the West. A ledger book utilising this new methodology made new categories of human association possible, as AW Crosby explains:

> Matthäus Schwartz, accountant to the Fuggers in the sixteenth century, called [double-entry book-keeping] a magic mirror in which the adept sees both himself and others. Before we examine the mirror directly (in which I think we will see ourselves), we must go back several centuries before the Fuggers became great bankers. There were no accounts receivable or payable, very little loaning of money and no accountants. There were no companies, no firms, no economic entities apart from the actual person or persons involved. One could not be a cog in a machine of a purely economic nature, because there were no such machines. The manor was economic, yes, but also familial, social, religious, and political. The monastery was economically efficient, with platoons and labourers in its field and tending its mills, but it was first and foremost religious.[59]

---

[55] Smith (n 38) 183.

[56] See David Koepsell and Barry Smith, 'Beyond paper' (2014) 97(2) The Monist 222, 222.

[57] David Koepsell, 'Beyond Human: Smart-contracts, Smart-Machines, and Documentality' in Jason Allen and Peter Hunn (eds), *Smart Legal Contracts: Computable Law in Theory and Practice* (Oxford University Press 2022) Chapter 14.

[58] Lee Bacon and George Bazinas, '"Smart Contracts": The Next Big Battleground' (2017) 18 Jusletter IT. See also Basil Selig Yamey, 'Scientific Bookkeeping and the Rise of Capitalism' (1949) 1(2) The Economic History Review 99.

[59] See Alfred W Crosby, *The Measure of Reality: Quantification in Western Europe, 1250–1600* (Cambridge University Press 1996) 204 citing Michael Baxandall, *The Limewood Sculptors of Renaissance Germany* (Yale University Press 1980) 231. Emphasis mine.

Given this, it is unsurprising that the radical connectivity provided by public networks such as the Internet has given rise to new forms of collective enterprise and that innovations such as DLT are extending the horizons even further. A tenet of the Coaseian approach to the firm is that advances in ICT, which reduce the cost of organising spatially, 'tend to increase the size of the firm.'[60] Before the creation of the Internet, as Primavera de Filippi and Aaron Wright observe, networks like Facebook would have been inconceivable. As the Internet grew, however, people became accustomed to using it for communicating, connecting and engaging with others. At first bilaterally and then in a more networked fashion, until 'over time these relationships solidified into larger networks, linking people together and enabling a greater flow of information across geographic boundaries.'[61] If we relate these ideas back to the notion of computer systems having a dual (epistemic and ontic) function, documents such as company ledgers are augmentations of the human mind that allow successful, purpose-oriented social organisation on a scale unimaginable without stable documents. Our task, then, is to extend the theory of documentality to reflect the ontic functions of computer systems – and more than that, to reflect the agentive nature of algorithms within them.

Many records are declarative, i.e. have a double direction of fit. For example, the entry of my name in a company's register of members both asserts that I own *n* shares and also makes it the case that I own *n* shares. Until recently, documents could only be maintained by constant human intervention. A human agent was always required to bring the state of the world into line with the reality projected in the records. The promise of novel (though not very 'smart') 'smart contracts' is that a mechanism can be put in place, which then functions technologically on its own – in theory, indefinitely. Transactional algorithms have existed since before blockchain technology, but the latter promises to increase their independence and potentiality. To concretise, De Filippi and Wright explain that blockchain-based 'tokens' can be used to replace conven-

tional company shares to represent a person's interest in a collective undertaking. Each token can encode that person's rights (e.g., to participate in profit or loss, to access, manage or transfer the resources it controls, etc.). Smart contracts can be used to automate these actions, such that corporate managers in the traditional sense are displaced to a greater or lesser degree; when the required number of token-holders vote for a course of action, for example, smart contracts might perform the action automatically without further human action.[62] This, in our view, adds something new to the ontic function of computer systems in modern society. They have come to life, as it were, as digital automata that structure human relations through their independent 'action'. PG Hunn explains:

> The capabilities of smart legal contracts are predicated upon the introduction of two attributes hitherto largely absent from contracts expressed in other forms: *executability and statefulness.* The former enables a contract document to be processed, in whole or in part, as a series of machine-readable instructions. In doing so, *the contract document is able to operate as a dynamic artefact,* responding to data inputs and generating data outputs. The latter enables the condition of the agreement to be stored, based upon the resulting execution events.[63]

Smart contracts would appear to offer a new technological foundation for more sophisticated documentary complexes – complexes which *interact* with each other and with the conventional subjects of law ('natural persons' and 'juristic persons'). This will change the range of possibilities open to individual actors, it will change the role of established public and private institutions and will open up new horizons in collective action.

Artificial Agents: From Smart Contracts to AI

Previously, I argued that there are two important features of smart contracts for legal theory. One is the *formalisation of legal language* implicit in the exercise of reducing

---

[60] Ronald Harry Coase, *The firm, the market, and the law* (University of Chicago press 2012) 46.

[61] Primavera De Filippi and Aaron Wright, *Blockchain and the Law: The Rule of Code* (Harvard University Press 2018) 265 (text to fn40).

[62] ibid 136-137.

[63] Peter G Hunn, 'Smart Contracts as Techno-Legal Regulation' (2019) 7(3) Journal of ICT Standardization 269, 273. Emphasis mine.

contractual provisions to machine-intelligible code.[64] The other is the *automatic execution of actions* in fulfilment of the contractual obligations so specified. This, in turn, is interesting to legal theory for two reasons. The first and most obvious, is the 'immutable' or unstoppable nature of such an algorithm. This poses obvious questions in terms of legality, that have been explored in the literature to date.[65] The second, on which I shall now focus, is the notion that an algorithm is performing legally relevant actions and possibly even performing 'acts-in-the-law' – for example, buying and selling securities. In Hildebrandt's seminal analysis, modern law presupposes the affordances of the printing press, 'a cybernetic system that informs self, mind, and society', but this is changing and we cannot assume that the architecture of novel ICT has the same affordances as that of the printing press.[66] This is not only because the medium by which law articulates its regulation is being transformed, but also because the playing field itself is reconfigured and we find ourselves at play with previously unknown 'actors' capable of autonomous interventions that we may not be aware of, including data-driven agents.[67]

The law is ancient and has digested sweeping social and technological transformations before, but Hildebrandt warns against lawyers' natural tendency to claim that things are 'not really that different now'. The 'deep structure' of the modern law, she argues, has been built on the assumptions of static documentation – linearity and the sequential processing demands of written text evoke the need for (human) interpretation, reflection, and contestation. The study and practice of law have thus been focussed on establishing the meaning of legal norms and their applicability to relevant human interactions, while establishing the meaning of human action in the light of the applicable legal norms. While many of these processes

might be applied to code-based 'agents' such as smart contracts (allowing for new complexities), what Hildebrandt calls 'data-driven agency' builds on a different grammar; its building blocks are *information* and *behaviour*, not *meaning* and *action*.[68] We cannot take for granted that the law, which we understand to be a discursive, and therefore social, iterative normative practice, will interact with an artificially intelligent, agentive information communications infrastructure in the same way that it interacted with written and printed text and human agents.[69]

Both code-driven and data-driven automated systems increasingly both (i) construct the reality humans inhabit and (ii) effect transactions within that reality. Our task, then, is no mean one: it is to establish the legal consequences of algorithmic 'action' (or 'behaviour') in a virtual domain to which the law attaches 'real world' meaning. There remains an important, even categorical distinction between human and non-human agents, but the lines are blurring. Not only because algorithms are becoming more intelligent, but because we are giving 'dumb' (but nonetheless powerful) algorithms ever more important jobs to do, and because we are ever less certain whether a human or an algorithm is 'acting' on the other end of the transaction. As De Filippi and Wright explain, 'weak' AI systems with sufficient 'intelligence' to perform highly constrained tasks are already active and increasingly ubiquitous. They power algorithmic trading systems; sort, track, and categorise information, images, and other forms of data on the Internet; predict weather, provide personal services, and run 'chatbots' that are often our first point of contact with both private companies and public authorities.[70] Blockchain-based systems are increasingly able to provide an interoperable layer for algorithmic systems (including weak AI) to interact and perhaps sooner rather than later,

---

[64] See editor's Introduction in Jason G Allen and Peter Hunn, *Smart Legal Contracts: Computable Law in Theory and Practice* (Oxford University Press April 2022).

[65] See, e.g., Max Raskin, 'The Law and Legality of Smart Contracts' (2016) 1 Georgia Law and Technology Review 305.

[66] Hildebrandt, 'Law as Information in the Era of Data-Driven Agency' (n 28) 23 referring to Mireille Hildebrandt, *Smart Technologies and the End(s) of Law. Novel entanglements of Law and Technology* (Edward Elgar 2015).

[67] Hildebrandt, 'Law as Information in the Era of Data-Driven Agency' (n 28) 8.

[68] This would seem to echo the view expressed by John Searle that computers are *a priori* incapable of ascribing meaning: David Cole, 'The Chinese Room Argument' in Edward N Zalta and Uri Nodelman (eds), *The Stanford Encyclopedia of Philosophy* (Winter 2024, Metaphysics Research Lab, Stanford University 2024) This, however, draws us into a deeper debate on the nature of (artificial) intelligence and agency.

[69] Hildebrandt, 'Law as Information in the Era of Data-Driven Agency' (n 28) 2.

[70] De Filippi and Wright (n 61) 147.

coordinate themselves to form a truly decentralised autonomous organisation ('DAO').[71]

In this discussion, I have cast a broad net, embracing both symbolic logic and data-driven AI, accepting some loss of granularity and nuance in order to trace broader and deeper themes. This is risky, but I think the point is general enough. Automated systems (of whatever type) are still obviously tools or instruments used by (human) agents. But they have surely acquired *some* of the properties of agents, at least within certain contexts, and they must be recognised as distinct from simple tools (e.g., shovels and typewriters). To the extent that they can manipulate the documentary fabric of social reality without direct human control, they are potentially very powerful agents and their advent in the 'digital real' must be taken seriously. If we recognise the state of an algorithmic system as projecting the state of (affairs in) institutional legal reality, we effectively add a new category of agentive entities to the ontology of that legal system.[72]

Automata and Agentivity

At this juncture, I hope that an exercise in good, old-fashioned Oxonian language analysis is not seen as trite. It seems that there are two ways an entity can 'act'. First, technological automata act, in response to our coded instructions, on the physical world. Our words command computers in robots and drones and their mechanical hands and feet perform those commands, often with terrible efficiency. In most cases, whether a DVD is delivered by hand or by drone is no more important than whether one ploughs a field by horse or tractor. But automata may

'perform' actions, such as torts, for which the law will need to develop appropriate rules of attribution.[73] The concrete problems that will arise in the context of packet-delivering drones and self-driving cars might be more analogous to cases concerning animals than inanimate tools in the conventional sense. Already this gives a hint at what we mean (and do not mean) by 'action' and 'agency'.

Secondly, and more relevant for our present purposes, the world we inhabit is increasingly virtual. An increasing proportion of our assets, for example, exist in a purely digital realm.[74] This is most obviously true of online worlds such as *World of Warcraft* or *Second Life*. But the border between these game-worlds and the 'real' world of institutional legal reality is increasingly porous. Our financial system now exists in an almost purely digital realm. The only categorical difference between money in a *Second Life* bank and money in a 'real world' bank is that a national legal system takes cognisance of demand deposits recorded in the digital ledger of a licensed commercial bank, but not those recorded in a game bank as increasing the supply of 'broad money'. A digital repository of bank balances is ontologically similar to a repository of digital 'treasure' in a game – it is only rules in the legal system that give the money any greater reality.[75] When an algorithm can manipulate digital assets with 'real world' consequences, it is surely appropriate to describe its behaviour as 'agentive' in some sense.

Legal theorists and computer scientists tend to use 'agency' differently: the former connote self-consciousness, rational self-interest, freedom of choice and legal capacity with the term; the latter mean only the ability to 'act' in a technical system. In a universe constructed using an object-oriented programming lan-

---

[71] De Filippi and Wright (n 61) 149. See also Allen, 'Bodies without Organs: Law, Economics, and Decentralised Governance' (n 31).

[72] Ontology itself has two related meanings: (i) the branch of metaphysics dealing with the nature of being, with which I have mostly been concerned thus far, and (ii) a set of concepts and categories in a domain that shows their properties and the relations between them. While I generally use it more broadly, here I use 'ontology of law' in the second meaning, to say that 'agentive algorithms' are now an object in the law's recognised *taxonomy* of all that exists and is the case.

[73] See Jason G Allen, 'Agency and Liability' in Charles Kerrigan (ed), *Artificial Intelligence: Law and Regulation* (Edward Elgar Publishing 2022). This section expands on some of the discussion in that chapter.

[74] You are likely reading this paper as a PDF, which you accessed from behind a paywall by entering login credentials. It was probably paid for by your institution via electronic funds transfer, meaning that digitally-represented debt relationships between your institution, its bank, the content provider, and the content provider's bank.

[75] Indeed, I could start an Excel spreadsheet in which 'digital dollars' were represented, and this could become an international reserve system if only the people of the world would accept it as such. See generally Allen, 'Law's Virtual Empires: Game Analogies and the Concept of Law' (n 34).

guage, for example, objects 'message' each other 'requests' according to their 'methods' and it is the 'behaviour' of these objects within the computer's hardware that creates the epistemic and ontic functions of the computer system described above. Hildebrandt's distinction seems to imply that 'action' is available to 'agents' in the legal sense and that agents in the computer science sense are confined to 'behaviour'. This may be a valid distinction to draw, but it would seem that we must recognize that 'behaviour' as *agentive* at least.

In a brief paper on agentivity from the early 1970s, DA Cruse set out some thoughts that I think are quite helpful in the present context. First, Cruse observes that the notion that verbs ('doing words') can be 'agentive' is intuitive, but that it was difficult to determine what exactly agentivity actually means. It would seem that agentivity implies a *relation between a verb and a noun*. In effect, we are asking what kind of noun can be the subject of a relation with a verb.[76] Cruse thus approaches the definition of 'agentivity' as a feature of the meaning of the verb 'to do'. For example, consider the relationship between the noun and the verb in: (i) *The wind overturned the dustbin*; (ii) *John accidentally overturned the dustbin*; and (iii) *John overturned the dustbin*. Is it appropriate to describe the wind as 'doing' something ('overturning') here? What about John in (ii)?

Cruse suggests that we classify clauses according to the 'preferred' form of the corresponding 'identifying clause'. *John punched Bill* is a 'do-clause', because *What John did was punch Bill* is preferred to *What happened to John was that he punched Bill*. The latter seems to miss something essential about John. On the other hand, *The vase broke* is a 'happen-clause', since *What happened to the vase was that it broke* is preferred to *What the vase did was break*.[77] This approach, argues Cruse, can be expressed using the logic of necessary implication or entailment: *John broke the vase* entails that *John did something*; *The vase broke*

does not entail *The vase did something*, but rather *Something happened to the vase*. So, in this case, it would seem to be happier to speak of the vase as a non-agentive noun, an object unfitting for the ascription of action in a sentence.

Yet, if we take this approach, we are stuck with cases such as *The wind blew the tree down*. We might not want to ascribe the wind agency, but it would be unhappy to say *What happened to the wind was that it blew the tree down*. Cruse suggests that 'inanimate objects can, as it were, acquire a temporary "agentivity by virtue of their kinetic (or other) energy" and that it is appropriate to ascribe this on the basis of the "causation of some external effect"'.[78] But there are cases which cast even this restriction into doubt. 'Apparently', he concludes, 'in certain circumstances the difference between *do* and *happen* is neutralised, *do* being the normal form.'[79]

In order to prevent our conception of 'agentive algorithms' becoming too broad, a concept from legal theory might provide a suitable limitation: the basic distinction between *acts-in-fact* and *acts-in-the-law*. An act-in-the-law is an act which, by reference to a set of rules, purports to change the state of the world (that is, of institutional legal reality).[80] In this investigation, we are chiefly concerned with the apparent ability of smart contracts to effect legal transactions, which might be internal to the organisation (i.e., governing the relations between token-holders *inter se*) or between the organisation and third parties (i.e., forming and effecting a financial trade). Categorically, we would seem to be concerned with *automating the performance of acts-in-the-law.*

In this context, Cruse's distinction between four aspects of 'agentivity' would seem to be relevant. Cruse distinguishes between relations between a noun and a verb that are (i) *volitive*, (ii) *effective*, (iii) *initiative* and (iv) *agentive*.

---

[76] Cruse (n 8) 14.

[77] ibid 13, citing Michael Alexander Kirkwood Halliday, 'Notes on transitivity and theme in English Part I' (1967) 3(1) Journal of Linguistics 37.

[78] Cruse (n 8) 16.

[79] ibid 17.

[80] See Hester DS Van Der Kaaij, *The Juridical Act: A Study of the Theoretical Concept of an Act that aims to create new Legal Facts* (vol 129, Springer International Publishing 2019) for a contemporary discussion of an old concept that draws on speech act theory and the institutional theory of law within which my investigation is, broadly, couched.

The *volitive* feature is present when an act of will is stated or implied in the verb. 'Willing is a kind of doing, whether what is willed is a state, process, or action.'[81] Sometimes, the 'do-feature' of an agentive noun is due entirely to its volitive properties. In other cases, actions are inherently non-volitional (e.g., sneezing, tripping, mistaking). *Effectivity*, in turn, refers to a feature in a sentence which refers to something which exerts a force of some kind, literal or metaphorical, not in virtue of an internal engine but because of its position, momentum, etc, for example *The stone broke the window* or *The pillar supports the roof*. Here, the stone seems more truly 'agentive' than the pillar, which could equally be described as 'instrumental' only.[82] The third closely related feature is *initiative*, when the noun in question initiates an action, such as *The warden marched the prisoners across the yard*. It is obviously the prisoners doing the 'marching' and the action of 'marching' on the part of the warden rather means that she caused the prisoners to act. Finally, Cruse finds (true) *agentivity* in any sentence referring to an action 'performed by an object which is regarded as using its own energy in carrying out the action,' which might include 'living things, certain types of machine and natural agents.'[83] Agentivity in this sense is found in sentences such as *John moved to avoid the car, The machine switches off at 6:00, The fire spread through the building*, which can all be 'tested' by the insertion of a reflexive element (e.g. *John moved (himself). . .*, etc.).

Turing machines are the kind of machines to which it is at least intuitively right to attach agentivity. We speak of computers 'doing things' all the time. To be 'agentive', it would seem that an algorithm need only be capable of effecting a change within a digital information system without human intervention. The question is whether there is a class of verbs which we desire, for some reason of logic or policy, to restrict to conventional animals or to humans in particular.

The feature of *volatility* is of particular importance in the concept of an act-in-the-law, for legal systems generally attach different legal consequences to identical actions based on the state of mind of the actor. Although agentivity and volitivity co-occur frequently, they are distinct.[84] It would seem to me that, while it is coherent to speak of *agentivity* in algorithms, we are not yet at the point where we could speak meaningfully of volatility. This may have implications for the legal consequences attached to the actions of algorithms transacting with each other in a manner largely independent of human oversight and control.[85] Perhaps the legal system should ascribe no legal consequences to such 'action' or 'behaviour' at all or make the legal consequences conditional on robust attribution to a human agent.

As we grapple with the idea of automated acts-in-the-law, it will become necessary to set out with more rigour a logic of action that can be applied to conventional agents and agentive automata alike. While it is beyond the scope of the present contribution, at the least this will include a robust distinction between the capacitative and deontic aspects of acts-in-the-law – that is, what an agent (or quasi-agent) *may* do and what they (or it) *can* do. Elsewhere, I have described this by reference to Lars Lindahl's concept of *Spielraum* comprising an 'ability-space' and a 'liberty-space' set by rules of competence and rules of conduct, respectively.[86]

Conclusion: *Die ich rief, die Geister!*

The parallels between automatic law and magic also sound a warning. In Goethe's poem, the apprentice's automaton fills the bathtub as instructed and then keeps filling it. 'Stop! Stop!' he cries, but he has forgotten the necessary

---

[81] Cruse (n 8) 18.

[82] ibid 20.

[83] ibid 21.

[84] ibid 19.

[85] See Gleeson (n 19).

[86] See Jason G Allen, *Non-Statutory Executive Powers and Judicial Review* (Cambridge University Press 2022) Ch 7 citing Lars Lindahl, 'Hohfeld relations and spielraum for action' (2006) 26(2) Análisis Filosófico 325, 325; Lars Lindahl, 'Hohfeld relations and spielraum for action' in Christian Dahlman (ed), *Studier i rättsekonomi: festskrift till Ingemar Ståhl* (Studentlitteratur AB 2005).

command. 'Those I called! The spirits! Now I can't get rid of them!' He is saved when the old witch master returns and orders the spirits down.

The apprentice's error – instructing the spirits to 'Fetch water!', instead of instructing them to 'Fill the bath-tub (and stop when it's full)!' – suggests that those giving instructions to automata should also exercise caution. De Filippi and Wright identify this as perhaps the most critical legal concern surrounding DAOs, for example:

> [The operations] of a DAO ultimately depend on the operations of the underlying blockchain-based network. *As long as the DAO collects enough funds to operate, it will keep working toward furthering its own mission, without paying attention to the implications this might have on society. Once deployed, illicit or defective code will operate as planned, even if it contravenes the law or the fundamental purpose of the organisation.* The code will be automatically enforced by the underlying blockchain infrastructure, with little room for a court or other third party to enforce an amendment. Because no-one has control over the smart contracts underpinning a DAO, no single party can intervene to amend the code or bring it to a halt, even if the code does not work as expected.[87]

Data-drive automata are even more problematic. The problem is even more serious when humans are reintroduced into the equation, because it is always possible for an opportunistic human to exploit an automated system. This is, in effect, what happened in the notorious incident of TheDAO.[88] The response of the Ethereum community was to collude in order to fork the blockchain before the affected funds could be accessed. A certain Elijah of Chelm is reputed to have created a *golem* in the mid-16th century of monstrous size and strength; upon reckoning its destructive potential, Elijah is said to have torn the life-giving Name from the *golem's* brow, upon which it crumbled into dust.[89] Perhaps it is necessary to ensure that humans retain

a 'kill-switch' to ensure that digital spirits do not threaten the values which they were animated to promote.

Like denizens of the spirit world, computers are not very good at deriving meaning from the context in which a human utters his spell; they are either ignorant of human purposes, or they do not care; they have an immanent teleology to follow. This problem plays out differently in the context of code-driven and data-driven automata. In the former case, the problem is one of formal expression of human intentions, generally in an environment shorn of context.

In the DAO example, this underscores the currently anaemic state of 'blockchain corporate governance.' But it points to a deeper problem: even an excellent translation implies a bit of 'treason'.[90] Information (in the mathematical sense) can be compressed and transferred, decoded and recoded without loss of integrity, but its *meaning* remains contingent upon the 'reading' of its (ultimately human) recipient. As Hildebrandt explains, '[w]hereas from the perspective of a mathematical theory of information this may be irritating, because it turns straightforward information into something ephemeral such as meaning, for the law this is crucial.'[91] This problem could be dismissed as a teething problem in an emerging field. But we ignore the cautionary tale of Goethe's apprentice at our peril.

In the case of data-driven automata, the problem is one of the inherent opacity of technologies such as language models, which are very good at approximating speech and appear to be good at divining intention from natural language, but which leave a categorical gap. This is a problem, for example, in the context of AI use in public administration, where public law principles of good administration require that an outcome be explicable by and justifiable by reference to standards of good decision-making that include things like absence of bias and proper purpose orientation. Though a system might function flawlessly

---

[87] De Filippi and Wright (n 61) 154-155. Emphasis added.

[88] For an explanation of the (now-familiar) exploit of TheDAO, see Muhammad Izhar Mehar and others, 'Understanding a Revolutionary and Flawed Grand Experiment in Blockchain: The DAO Attack' (2019) 21(1) Journal of Cases on Information Technology (JCIT) 19.

[89] Trachtenberg (n 7) 85.

[90] Hildebrandt, 'Law as Information in the Era of Data-Driven Agency' (n 28) 17 citing G Rabassa, *If this be Treason: Translation and Its Dyscontents : a Memoir* (New Directions Book 2005).

[91] Hildebrandt, 'Law as Information in the Era of Data-Driven Agency' (n 28) 27.

(i.e., as intended) in the majority of cases, any gap between their actions and our ability to understand the outcome is cause for concern.

It is increasingly difficult to maintain oversight over the capacity of machines to simulate true agentivity, and their apparent ability to understand our natural language speech and document acts blurs the line further. At some point we may reach the stage where the 'utterances' of data-driven agents start to change the state of the (legal) world, too.[92]

Wherever acts-in-the-law (e.g., contracts, administrative determinations) and acts with legal consequences (e.g., torts) are performed by automata, conventional understandings of the rule of law will need to be fundamentally reconsidered and underlying 'social purposes' will need to be maintained.[93] Because there is a range of circumstances in which algorithmic actions should be recognised by the legal system to have legal consequences, I think that we should expect a DAO, say, or an AI system used in public administration to embody something *at least as good* as the current body of corporate governance norms or the principles of good administration. And this entails that we must leave conceptual room for a *human* layer in the socio-technical 'stack' that is deployed. My concern is that the ethos motivating the development of many technology applications discourages innovators from adequately addressing this problem, based on a naïve belief that the 'immutability of code' or the quest for ever more capable AI will render the problem itself a thing of the past.

References

Alexander GS, 'Comparing the Two Legal Realisms-American and Scandinavian' (2002) 50(1) The American Journal of Comparative Law 131.

Allen JG, 'Property in Digital Coins' (2019) 8(1) European Property Law Journal 64.

— 'Bodies without Organs: Law, Economics, and Decentralised Governance' (2020) 4(1) Stanford Journal of Blockchain Law & Policy 53.

— 'Law's Virtual Empires: Game Analogies and the Concept of Law', in JL Fabra-Zamora and G Villa Rosas (eds), *Conceptual Jurisprudence: Methodological Issues, Classical Questions and New Approaches* (Springer International Publishing 2021).

— 'Agency and Liability', in C Kerrigan (ed), *Artificial Intelligence: Law and Regulation* (Edward Elgar Publishing 2022).

— *Non-Statutory Executive Powers and Judicial Review* (Cambridge University Press 2022).

Allen JG and Hunn P, *Smart Legal Contracts: Computable Law in Theory and Practice* (Oxford University Press April 2022).

Amselek P, 'Philosophy of Law and the Theory of Speech Acts' (1988) 1(3) Ratio Juris 187.

Austin JL, 'How to Talk: Some Simple Ways' (1953) 53(1) Proceedings of the Aristotelian Society 227.

Austin JL, *How to do things with words* (Second edition, Harvard University Press 1975).

Bacon L and Bazinas G, '"Smart Contracts": The Next Big Battleground' (2017) 18 Jusletter IT.

Baxandall M, *The Limewood Sculptors of Renaissance Germany* (Yale University Press 1980).

Bernal C, 'In search of the ontological structure of law' [2012] Analisi e Diritto 59.

— 'Collective Intentionality and the Ontological Structure of Law' (2014) 45(3) Rechtstheorie 335.

Brey P, 'The Epistemology and Ontology of Human-Computer Interaction' (2005) 15(3) Minds and Machines 383.

Brinz A von, *Lehrbuch der Pandekten* (Second Edition, Deichert 1857).

Clancey WJ and others, 'Brahms: simulating practice for work systems design' (1998) 49(6) International Journal of Human-Computer Studies 831.

Clarke AC, *Profiles of the Future: An Inquiry Into the Limits of the Possible* (Harper & Row 1973).

---

92 '[T]he practice of law is identifiable as such by the specific ensemble of constraints that determines what counts as a legally relevant fact, whether a legal act (concluding a contract), a legally relevant act (committing a tort), a legally relevant event (birth) or specific status (being an animal, a corporation or a notary public), coupled with the constraints that determine the legal effect (an obligation to deliver a good, to pay compensation, the status of a legal subject, or, simply the competence to create legal effect).' Hildebrandt, 'Law as Information in the Era of Data-Driven Agency' (n 28) 14.

93 ibid 26.

Coase RH, *The firm, the market, and the law* (University of Chicago press 2012).

Cole D, 'The Chinese Room Argument' in EN Zalta and U Nodelman (eds), *The Stanford Encyclopedia of Philosophy* (Winter 2024, Metaphysics Research Lab, Stanford University 2024).

Crosby AW, *The Measure of Reality: Quantification in Western Europe, 1250–1600* (Cambridge University Press 1996).

Cruse D, 'Some thoughts on agentivity' (1973) 9(1) Journal of Linguistics 11.

De Filippi P and Wright A, *Blockchain and the Law: The Rule of Code* (Harvard University Press 2018).

Diver L, 'Interpreting the Rule(s) of Code: Performance, Performativity, and Production' [2021] MIT Computational Law Report.

Dunn P, *Magic, Power, Language, Symbol: A Magician's Exploration of Linguistics* (1ˢᵗ ed, Llewellyn Publications 2008).

Ferraris M, 'Cultures. Conflict - Analysis - Dialogue' in C Kanzian and E Runggaldier (eds), *Proceedings of the 29th International Ludwig Wittgenstein-Symposium in Kirchberg, Austria* (De Gruyter 2007).

Gleeson S, 'The Legal Consequences of Automated Mistake' in JG Allen and P Hunn (eds), *Smart Legal Contracts: Computable Law in Theory and Practice* (Oxford University Press 2022).

Gunkel DJ, 'The Relational Turn: Third Wave HCI and Phenomenology' in M Filimowicz and V Tzankova (eds), *New Directions in Third Wave Human-Computer Interaction: Volume 1 - Technologies* (Springer International Publishing 2018).

HHoltermann J v, 'A Straw Man Revisited: Resettling the Scope Between HLA Hart and Scandinavian Legal Realism' (2017) 57 Santa Clara L. Rev. 1.

Hage J, 'The (Onto)logical Structure of Law: A Conceptual Toolkit for Legislators' in M Araszkiewicz and K Płeszka (eds), *Logic in the Theory and Practice of Lawmaking* (Springer International Publishing 2015).

Halliday MAK, 'Notes on transitivity and theme in English Part I' (1967) 3(1) Journal of Linguistics 37.

Hart HLA, *The Concept of Law* (Oxford University Press 1961).

Hildebrandt M, *Smart Technologies and the End(s) of Law. Novel entanglements of Law and Technology* (Edward Elgar 2015).

— 'Law as Information in the Era of Data-Driven Agency' (2016) 79(1) The Modern Law Review 1.

Hobbes T, *Man and Citizen* (Gert B and Hobbes T eds, Anchor 1972).

Humberstone L, 'Direction of Fit' (1992) 101(401) Mind 59.

Hunn PG, 'Smart Contracts as Techno-Legal Regulation' (2019) 7(3) Journal of ICT Standardization 269.

Klinghoffer H, 'Jurisdiction and Illegality. A Study in Public Law. By Amnon Rubinstein. [Oxford: Clarendon Press, 1965. xii + 250 pp. 45s net.]' (1966) 1(3) Israel Law Review 510.

Koepsell D, 'Beyond Human: Smart-contracts, Smart-Machines, and Documentality' in J Allen and P Hunn (eds), *Smart Legal Contracts: Computable Law in Theory and Practice* (Oxford University Press 2022).

Koepsell D and Smith B, 'Beyond paper' (2014) 97(2) The Monist 222.

Lindahl L, *Position and Change: A Study in Law and Logic* (Springer Netherlands 1977).

— 'Hohfeld relations and spielraum for action', in C Dahlman (ed), *Studier i rättsekonomi: festskrift till Ingemar Ståhl* (Studentlitteratur AB 2005).

— 'Hohfeld relations and spielraum for action' (2006) 26(2) Análisis Filosófico 325.

MacCormick N, *H.L.A. Hart* (vol 93, Stanford University Press 1981).

MacCormick N and Weinberger O, *An Institutional Theory of Law: New Approaches to Legal Positivism* (vol 3, Springer Netherlands 1986).

Macías M, Casanovas P, and Zeleznikow J, 'Shevirat Ha-Kelim: Jewish Mysticism and the Catalan Matrix for Dialogue and Conflict' [2020] Journal of Catalan Intellectual History 64.

Malinowski B, *Magic, Science, and Religion, and Other Essays* (The Free Press 1948).

Mehar MI and others, 'Understanding a Revolutionary and Flawed Grand Experiment in Blockchain: The DAO Attack' (2019) 21(1) Journal of Cases on Information Technology (JCIT) 19.

Minsky M, *Computation: finite and infinite machines* (Prentice-Hall 1967).

B Moeran and T De Waal Malefyt (eds), *Magical Capitalism: Enchantment, Spells, and Occult Practices in Contemporary Economies* (Springer International Publishing 2018).

Rabassa G, *If this be Treason: Translation and Its Dyscontents : a Memoir* (New Directions Book 2005).

Raskin M, 'The Law and Legality of Smart Contracts' (2016) 1 Georgia Law and Technology Review 305.

Reichling GA, 'A Demotic Folk-Tale the Basis of Goethe's "Der Zauberlehrling"' (1919) 39 Journal of the American Oriental Society 295.

A Reinach and J Crosby (eds), *The Apriori Foundations of the Civil Law* (De Gruyter 2012).

P Robbins and M Aydede (eds), *The Cambridge Handbook of Situated Cognition* (Cambridge University Press 2008).

Rogerson PJ, 'The Situs of Debts in the Conflict of Laws—Illogical, Unnecessary and Misleading' (1990) 49(3) The Cambridge Law Journal 441.

Ross A, 'Tû-Tû' (1957) 70(5) Harvard Law Review 812.

Ruiter DWP, *Legal Institutions* (Kluwer Academic Publishers 2010).

Searle JR, *Intentionality: An Essay in the Philosophy of Mind* (Cambridge University Press 1983).

—— *The Construction of Social Reality* (Free Press 1995).

—— *Making the Social World: The Structure of Human Civilization* (Oxford University Press 2010).

Searle JR and Vanderveken D, *Foundations of Illocutionary Logic* (Cambridge University Press 1985).

Shannon CE and Weaver W, *The Mathematical Theory of Communication: Volume 1 - Technologies* (University of Illinois Press 1963).

Smith B, 'How to Do Things with Documents' [2012] (50) Rivista di estetica 179.

Sommer J, 'Where is a Bank Account?' (1998) 57(1) Maryland Law Review 1.

Stichman HT, 'Jurisdiction and the "Situs" of Debts' (1925) 34(6) The Yale Law Journal (Tulin LA ed 652.

Tambiah SJ, 'The Magical Power of Words' (1968) 3(2) Man 175.

Trachtenberg J, *Jewish Magic and Superstition: A Study in Folk Religion* (University of Pennsylvania Press 2012).

Turing A, 'Intelligent Machinery, A Heretical Theory' (1996) 4(3) Philosophia Mathematica 256.

Turing AM and Copeland J, *The Essential Turing: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life plus The Secrets of Enigma* (Oxford University Press 2004).

Van Der Kaaij HDS, *The Juridical Act: A Study of the Theoretical Concept of an Act that aims to create new Legal Facts* (vol 129, Springer International Publishing 2019).

Werbach K, 'Trust, but verify: Why the blockchain needs the law' (2018) 33(2) Berkeley Technology Law Journal 487.

Westbrook DA, 'Magical Contracts, Numinous Capitalism' in B Moeran and T de Waal Malefyt (eds), *Magical Capitalism: Enchantment, Spells, and Occult Practices in Contemporary Economies* (Springer International Publishing 2018).

Wiener N, 'Some Moral and Technical Consequences of Automation' (1960) 131(3410) Science 1355.

Yamey BS, 'Scientific Bookkeeping and the Rise of Capitalism' (1949) 1(2) The Economic History Review 99.

Yovel J, 'The Creation of Language and Language without Time: Metaphysics and Metapragmatics in Genesis' (2012) 20(3) Biblical Interpretation 205.

A reply: On the Importance of Grounding and Causality for Agency in Humans and LLMs

Lea Frermann • Senior Lecturer, School of Computing and Information Systems, The University of

Melbourne, lea.frermann@unimelb.edu.au

Abracadabra! Text, Data, and Agency in the Digital Society explores the anticipated impacts of advancing AI and language models on the institution and functioning of the law. At its core, it conceptualizes law as a shared *social* reality, which lacks grounding in the *physical* world. The institution of law relies on several layers of shared belief within a society. These shared beliefs and conventions allow legal artifacts, such as contracts, oaths, or promises, to be 'spoken into existence' – not unlike a magic spell! To date, most legal 'spells' are carried out by humans, who have earned the right to do so through their status (e.g., election as a judge or the award of an appropriate university degree); and they possess the qualities of agency, free will, and consequently, liability for their actions. With the increasing power of AI and large language models (LLMs), Allen asks: What happens once we allow AI to perform 'the magic of the law'?

My expertise lies in natural language processing (NLP), where I combine methods from cognitive science, computational linguistics, and machine learning to better understand how humans conceptualize complex problems, how machines (still) fall short of this goal and how they might be improved. Based on this background, I will explore the commonalities and differences in how humans and AI models function, and, as a result, how they can engage in social reality – in law or otherwise. When I talk about AI or large language models (LLMs), I refer to recent transformer-based models such as GPT-4o, Llama or Claude.

Human existence is embedded in a shared reality. Most fundamentally, this includes the physical environment, and the way we perceive it through vision, smell, pleasure, or pain. Additionally, we share a social reality, driven by desires and beliefs. We pursue our goals within the context of the goals of our colleagues, friends, and families, creating a complex web of social interaction. This enables us to create intricate, shared social constructs such as law, society, or nationality. All of this operates based on constant streams of multi-modal data that we process: physical data from our surroundings, and information we consume digitally or otherwise.

The reality of LLMs looks starkly different because it is restricted only to incoming digital data. LLMs are superhuman data processors with a wondrous ability to analyze and store vast amounts of information. By processing massive amounts of human-generated text, LLMs distill and store complex patterns underlying our language in their enormous parameter space, learning to reproduce text that follows the same patterns – and looks magically human-like. But all they do is compute the highest probability continuation of an input prompt. LLMs, as of today, have no notion of broader context: how information might look, smell, or feel, or how it might impact others. Nor can they refuse to complete a task for any reason other than deeming that refusal is the most likely response given the input text.

Allen writes:

> As machines can be programmed to respond automatically – in increments, more autonomously – legal spells need not address human subjects as their middlemen anymore. This highlights the importance of algorithms that seem to perform legally relevant acts without (further) human input: they take the law one step towards the magical by replacing human agency with 'agency ex machina'.

It is ultimately a philosophical question of where the line is drawn between a 'predictive response' and an 'agentic

act'. However, without goals grounded in the shared social reality of the law, and without a model of those affected by the law, it is questionable whether (a) agency can be ascribed, and (b) there is any hope that the behavior of an 'LLM agent' will align with the goals and values of its creator.

LLMs, like most machine learning models, are *predictive* models. They learn useful correlations in the data that allow them to solve specific tasks. Unlike humans, they do not have a *causal* model of the underlying principles that would allow them to make broad generalizations. This is not only a theoretical point but also of great practical relevance. By relying only on correlations in historical data, a model is vulnerable to manipulation: the training data could be strategically selected by adversarial actors to support some desired associations. More fundamentally, a lack of generalization ability based on underlying causal principles means that models generalize poorly from high-resource domains of law to niche scenarios. Therefore, AI models will be significantly more beneficial and reliable

in (the few) well-resourced areas of law – such as English-written case law, family law, or criminal law. Many niche areas and smaller language communities will be left behind. While human agency and liability support social constructs around the world, the performance and utility of 'LLM agents' will vary systematically. How the legal systems of the world will respond to this is another aspect to explore.

Allen draws thought-provoking (and entertaining!) parallels between the fields of computer science, magic, and law, and fully succeeds in opening new horizons by playful connection of concepts. My points above are offered with a view towards grounding Allen's framework more firmly in the reality of today's machine learning, as well as the fundamental gaps between human learning and cognition that seem to underlie meaningful agency in social reality. I am excited to see the cross-disciplinary conversations unfold that Allen's article, and the platform of this journal, will spark!

Author's reponse

Jason Grant Allen

First, thanks to Dr. Lea Frermann for her careful reading of and comment on my paper. This response sets out (i) some points of clarification and (ii) some reflections on my argument, its limitations, and the philosophical programme it seeks to address.

First, two necessary clarifications:

- My concept of law is not one 'which lacks grounding in the *physical* world', at least not in such terms: everything social *is* grounded in the physical world, but the connection between the social and the physical needs ontological explication. The challenge is to connect the realms of molecules and mothers-in-law, and the tradition I follow uses the philosophy of language and mind to do that (*pace* Searle) and documents are integral to the story (*pace* Ferraris).
- The question is not (just) what happens once we allow machines to 'cast spells', but whether we should *ever* allow them to do so, *categorically* – now, and in some future state of higher sophistication.

The philosophical question Frermann articulates ('where the line is to be drawn between a "predictive response" and an "agentic act"') somewhat flattens the problem-space. The first question is whether machines (now, or later) can plausibly be ascribed 'agency'. Even if we answer 'no', the question is whether there are *nonetheless* changes to the way the law operates *qua* human institution – for example, because the type and extent of automation makes attribution of machine behaviour to humans a fiction.

This paper has gestated, and mutated, since 2018. The argument has been (re)directed towards good old-fashioned software, symbolic AI, blockchain 'smart contracts', and generative AI. It risks chasing multiple rabbits (and catching none). I try to embrace more 'agentive' machines *and* mundane automata, and I try to embrace data-driven machines that produce legal text *and* code that embodies legal text. Perhaps I should have re-written this paper (again) to focus exclusively on one or the other – but I felt that something would be lost with that approach.

A central question is the relationship between *tools* and *agency*. What changes when we use tool X, Y, or Z to do something? Maybe nothing – we could call this the 'mere tools' approach. Maybe these things are not really tools (objects) at all, but are more like subjects – we could call this the 'digital agents' approach. The truth is (now, and possibly categorically) in the middle, but we do not have a conceptual or lexical apparatus for dealing with 'agentive' tools (whether data-driven or code-driven). That is what we are currently trying to build. The law of animals, or of slaves in a system like Roman law, may be helpful. So might the law of artificial group agents (corporations). In future work, I would address this question (more) directly. My approach in this paper was perhaps too discursive. But my aim was less to hit a moving target, and more to illustrate, in a reasonably systematic way, the intellectual resources we need if we hope to hit it. I hope the paper will remain useful in that regard.