2-2023

Legal dispositionism and artificially-intelligent attributions

Jerrold SOH
Singapore Management University, jerroldsoh@smu.edu.sg

Legal Studies (2023), 1–20
doi:10.1017/lst.2022.52

S·L·S
The Society of Legal Scholars

RESEARCH ARTICLE

Legal dispositionism and artificially-intelligent attributions

Jerrold Soh*†

Singapore Management University, Singapore
*Author email: jerroldsoh@smu.edu.sg

(Accepted 6 December 2022)

Abstract

It is conventionally argued that because an artificially-intelligent (AI) system acts autonomously, its makers cannot easily be held liable should the system's actions harm. Since the system cannot be liable on its own account either, existing laws expose victims to accountability gaps and need to be reformed. Recent legal instruments have nonetheless established obligations against AI developers and providers. Drawing on attribution theory, this paper examines how these seemingly opposing positions are shaped by the ways in which AI systems are conceptualised. Specifically, folk dispositionism underpins conventional legal discourse on AI liability, personality, publications, and inventions and leads us towards problematic legal outcomes. Examining the technology and terminology driving contemporary AI systems, the paper contends that AI systems are better conceptualised instead as situational characters whose actions remain constrained by their programming. Properly viewing AI systems as such illuminates how existing legal doctrines could be sensibly applied to AI and reinforces emerging calls for placing greater scrutiny on the broader AI ecosystem.

Keywords: artificial intelligence; autonomous systems; attribution theory; law and technology; law and psychology

Introduction

It is conventionally argued that an artificially-intelligent (AI) system's actions and their potentially harmful consequences cannot easily be attributed to the system's developers or operators because the system acts autonomously.[1] Nor can the system, which has no legal personality, be liable on its own account. Victims are left exposed to accountability gaps,[2] and thus AI disrupts law,[3] necessitating new models of legal analysis.[4] In specific doctrinal areas, this question is typically

†I thank Professors Yip Man, Lau Kwan Ho, Alex Yang, Yuko Nakato, Gary Chan, James Penner, and the participants of the inaugural Asia-Pacific Private Law Conference 2021 for helpful comments on earlier drafts, as well as the anonymous referees for their expert, constructive comments. All errors are mine.

[1]See for example, European Commission 'Proposal for a Directive of the European Parliament and of the Council on Adapting Non-Contractual Civil Liability Rules to Artificial Intelligence' COM(2022) 496 final (AI Liability Directive) p 16; DC Vladeck 'Machines without principals: liability rules and artificial intelligence' (2014) 89 Washington Law Review 117. Variations of this argument are made or described across the sources below in nn 2–11.

[2]B-J Koops et al 'Bridging the accountability gap: rights for new entities in the information society?' (2010) 11 Minnesota Journal of Law, Science & Technology 497; T Liivak 'Liability of a manufacturer of fully autonomous and connected vehicles under the Product Liability Directive' (2018) 4 International Comparative Jurisprudence 178.

[3]R Brownsword 'Law, authority, and respect: three waves of technological disruption' (2022) 14 Law, Innovation and Technology 5.

[4]H-Y Liu et al 'Artificial intelligence and legal disruption: a new model for analysis' (2020) 12 Law, Innovation and Technology 205.

framed as a 'missing person' problem: when, instead of humans, AI systems drive,[5] contract,[6] defame,[7] make art,[8] commit crimes[9] and, more broadly speaking, cause harm,[10] how should law respond?[11]

Questions of this form are beginning to reach the courts.[12] Aiming to plug this perceived gap, in 2017 the European Parliament (EP) proposed a 'specific legal status' for AI 'so that at least the most sophisticated autonomous robots could be … electronic persons responsible for making good any damage they may cause'.[13] But this resolution was strongly criticised by legal and technological experts as premised on 'an overvaluation of the actual capabilities of even the most advanced robots, a superficial understanding of unpredictability and self-learning capacities, and a robot perception distorted by Science-Fiction'.[14] The proposal was promptly shelved, and a 2020 resolution would instead emphasise that electronic personality was unnecessary because 'all physical or virtual activities… driven by AI systems … are nearly always the result of someone building, deploying, or interfering with the systems'.[15] This position is reflected in recent EU legal instruments including the draft AI Act which imposes regulatory obligations on providers, distributors, and users of certain AI systems.[16]

Of course, AI technology did not become any less sophisticated between 2017 and today.[17] The primary difference between the 2017 and 2020 resolutions lies in how each conceptualised AI systems. In 2017, they were intelligent, autonomous beings analogised to Prague's Golem and Frankenstein's Monster.[18] In 2020, they were software units programmed by humans to act within pre-defined boundaries. This paper examines how these opposing AI conceptions animate legal debates surrounding fault and liability attributions for AI systems. Drawing upon psychological 'attribution theory', or the study of how 'ordinary people [attribute] causes and implications [to] the events they witness',[19] the paper contextualises the 'AI autonomy' frame above as one built on folk 'dispositionism', a well-documented concept in attribution theory, and demonstrates how easily dispositional AI narratives can be manipulated to promote a desired legal conclusion. It then characterises recent proposals to

[5]T Mackie 'Proving liability for highly and fully automated vehicle accidents in Australia' (2018) 34 Computer Law & Security Review 1314; J Soh 'Towards a control-centric account of tort liability for automated vehicles' (2021) 26 Torts Law Journal 221.

[6]V Ooi 'Contracts formed by software: an approach from the law of mistake' (2022) 2 Journal of Business Law 97. See also Mance IJ's dissent in *Quoine Pte Ltd v B2C2 Ltd* [2020] 2 SLR 20 at [152].

[7]See Part 2(d) below.

[8]T Aplin and G Pasqualetto 'Artificial intelligence and copyright protection' in RM Ballardini et al (eds) *Regulating Industrial Internet Through IPR, Data Protection and Competition Law* (Alphen aan den Rijn: Kluwer Law International, 2019); M Svedman 'Artificial creativity: a case against copyright for AI-created visual artwork' (2020) 9 IP Theory 1.

[9]R Abbott and A Sarch 'Punishing artificial intelligence: legal fiction or science fiction' (2019) 53 UC Davis Law Review 323.

[10]P Čerka et al 'Liability for damages caused by artificial intelligence' (2015) 31 Computer Law & Security Review 376 at 377; R Abbott 'The reasonable computer: disrupting the paradigm of tort liability' (2018) 86 George Washington Law Review 1.

[11]For more examples see Y Bathaee 'The artificial intelligence black box and the failure of intent and causation' (2018) 31 Harvard Journal of Law & Technology 889 at 891; HR Sullivan and SJ Schweikart 'Are current tort liability doctrines adequate for addressing injury caused by AI?' (2019) 21 AMA Journal of Ethics 160 at 162–163; J Goldenfein et al 'Through the Handoff lens: competing visions of autonomous futures' (2021) 35 Berkeley Technology Law Journal 835 at 844.

[12]See Part 2 below.

[13]European Parliament Resolution of 16 February 2017 with Recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)) [2018] OJ C252/239 at [59(f)].

[14]'Open Letter to the European Commission: Artificial Intelligence and Robotics', available at http://www.robotics-openletter.eu/ (last accessed 25 January 2023).

[15]European Parliament Resolution of 20 October 2020 with Recommendations to the Commission on a Civil Liability Regime for Artificial Intelligence (2020/2014(INL)) [2020] OJ C404/107 at [7].

[16]European Commission 'Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts' COM (2021) 206 final.

[17]Conversely, advances since then include 'foundation models' that represent an AI 'paradigm shift': R Bommasani et al 'On the opportunities and risks of foundation models' (Stanford Institute for Human-Centered Artificial Intelligence, 2001) p 1.

[18]European Parliament, above n 13 at [A].

[19]L Ross 'The intuitive psychologist and his shortcomings' (1977) 10 Advances in Experimental Social Psychology 173 at 174.

https://doi.org/10.1017/lst.2022.52 Published online by Cambridge University Press

focus on identifying human actors responsible for AI system behaviours as 'situationist' responses which view AI systems as what Hanson and Yosifon call 'situational characters' – entities whose behaviours are driven more by external than internal forces.[20] Reviewing the technical capabilities of contemporary AI systems, the paper argues that they are better understood through a situationist lens. Unlike human DNA, which forms part of our *natural* dispositions, today's AI systems' decisional processes are written, controlled, and continually re-written by human actors.[21]

Contextualising the legal AI discourse within attribution theory illuminates how future discourse and policy-making surrounding AI systems should proceed. Specifically, it reinforces proposals focusing less on AI systems themselves than on the eco-system of providers, distributors, and users around them. Conversely, arguments premised on framing AI systems as sentient, intelligent beings are put in doubt. More broadly, attribution theory provides a framework for identifying pivotal misconceptions underlying conventional arguments on the legal conceptualisation of AI systems. Dispositional versus situational narratives subtly shape the questions we ask, and the answers we give, on AI liability attributions. As with philosophy and computer science,[22] AI provides a backdrop against which 'normative structure[s] underlying our understanding of law' may be challenged and re-examined.[23] Thus, the paper's broader significance, especially for scholars interested in more than law and (AI) technology per se, lies in revisiting the implications of attribution theory for law.[24]

The paper first introduces attribution theory and its legal implications. Next, it identifies how far dispositionism animates conventional legal AI discourse by reference to jurisprudence surrounding AI liability, personality, publications, and inventions. Third, it examines how contemporary AI systems operate and argues that they are better understood situationally. The paper concludes with an attribution-theory informed framework for analysing AI-related attributions.

Before proceeding, it should be clarified that this paper focuses on the AI systems in use and development *today* and says nothing about the attainability of, and potential legal issues around, 'strong AI'[25] systems.[26] Nonetheless, as the technology continues to develop, this work would form an important plank for understanding how the law should conceive of and respond to increasingly sophisticated AI systems. Further, this paper is primarily concerned with fault and liability attributions; issues and materials on AI ethics and governance will only be referenced briefly where relevant.

[20]J Hanson and D Yosifon 'The situation: an introduction to the situational character, critical realism, power economics, and deep capture' (2003) 152 University of Pennsylvania Law Review 129.

[21]See also J Cobbe 'Administrative law and the machines of government: judicial review of automated public-sector decision-making' (2019) 39(4) LS 636 at 639.

[22]See AM Turing 'Computing machinery and intelligence' (1950) 59 Mind 433; JR Searle 'Minds, brains, and programs' (1980) 3 Behavioral and Brain Sciences 417; BJ Copeland 'The Turing test*' (2000) 10 Minds and Machines 519. The debate persists, for instance, in R Manzotti and A Chella 'Good old-fashioned artificial consciousness and the intermediate level fallacy' (2018) 39 Frontiers in Robotics and AI 1; EM Bender and A Koller 'Climbing towards NLU: on meaning, form, and understanding in the age of data', *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (Online: Association for Computational Linguistics, 2020).

[23]M Zalnieriute et al 'The rule of law and automation of government decision-making' (2019) 82 Modern Law Review 425 at 426. See also G Samuel 'The challenge of artificial intelligence: can Roman law help us discover whether law is a system of rules?' (1991) 11 LS 24.

[24]This retraces to Lloyd-Bostock's observation that 'little attempt ha[d] been made to relate' the 'extensive body of literature in psychology on the attribution of causes and responsibility' to law: S Lloyd-Bostock 'The ordinary man, and the psychology of attributing causes and responsibility' (1979) 42 MLR 143. This was partially a response to Hart and Honoré's discussion on common sense causality in HLA Hart and T Honoré *Causation in the Law* (Oxford: Clarendon Press, 2nd edn, 1985). Subsequent work on law and attribution theory have revolved around the latter's relevance to causality, especially in criminal law. See, for example, NJ Mullany 'Common sense causation – an Australian view' (1992) 12 Oxford Journal of Legal Studies 431; A Summers 'Common-sense causation in the law' (2018) 38 Oxford Journal of Legal Studies 793; A du Bois-Pedain 'Novus actus and beyond: attributing causal responsibility in the criminal courts' (2021) 80 Cambridge Law Journal S61.

[25]The strong/weak taxonomy originates from Searle, above n 22.

[26]On this see B Schafer et al 'A fourth law of robotics? Copyright and the law and ethics of machine co-production' (2015) 23 Artificial Intelligence and Law 217; N Bostrom *Super-Intelligence: Paths, Dangers, Strategies* (Oxford: Oxford University Press, 2014).

https://doi.org/10.1017/lst.2022.52 Published online by Cambridge University Press

1. Attribution theory, artificial intelligence, and law

First observe that the conventional 'missing person' frame oversimplifies. Law does not necessarily require specific action(s) to be taken by specific person(s). Instead, established rules of attribution are deployed to deem one's actions (or liability) as another's.[27] These rules are usually premised on familiar doctrines such as control.[28] Thus a company can be liable for employee wrongs,[29] a platform can publish user-created content,[30] a landlord can be responsible for a tenant's nuisance,[31] and an animal's keeper can be liable if it bites.[32] The difficulty with AI is better thought of as a problem with *applying* these attribution rules in light of AI's apparent autonomy. Lloyd-Bostock distinguishes between attribution as 'a relatively unreflective … process of making sense of and getting about in the world' on one hand and a deliberate 'social act' where norm-violating events are to be explained on the other.[33] AI systems challenge both kinds of attributions. For the former, the technology's complexity makes intuitive assessments of factual cause-and-effect in relation to AI systems difficult. For the latter, AI's ostensible independence from human control obfuscates assessments of whom their actions should be attributed to. As Pasquale notes, drawing clear lines of AI responsibility is difficult because 'both journalists and technologists can present Al as a technological development that exceeds the control or understanding of those developing it'.[34]

(a) The dispositionist default

Insofar as the problem is one of attribution, it follows that law can draw important lessons from attribution theory. Attribution theorists would understand the missing person frame and our resulting search for new personalities to fault as a classic dispositionist response. Dispositionism models an agent's behaviour as primarily driven by the agent's *internal* calculus – its personality, character traits, and preferences.[35] 'Good' and 'evil' are basic adjectives for moral dispositions, though law prefers more nuanced terms such as 'dishonest', 'careless', and 'reckless'.

Dispositionism offers an elegant mechanism for attributing moral blame and legal liability because assuming that internal nature drives external behaviour lets us infer the former from observing the latter. One who returns a dropped wallet does so *because* they are good and honest; one who keeps it is evil or dishonest. The wallet-keeper, having demonstrated a morally-suspect disposition, is then blameworthy. One can fairly be held liable for one's actions, and their consequences on the world, because these actions are by and large expressions, and reflections, of one's true nature.[36]

The roots of dispositionism have been traced to Western philosophy,[37] finding expression in Aristotelian conceptions of virtue,[38] Cartesian notions of individual will,[39] and Lockean social

[27]See generally R Stevens *Torts and Rights* (Oxford: Oxford University Press, 2007) ch 11.

[28]See Soh, above n 5.

[29]N Campbell and J Armour 'Demystifying the civil liability of corporate agents' (2003) 62 Cambridge Law Journal 290; J Dietrich and I Field 'Statute and theories of vicarious liability' (2019) 48 Melbourne University Law Review 515.

[30]*Fairfax Media and Others v Voller* [2021] HCA 27.

[31]*Lawrence and Another v Fen Tigers Ltd and Others (No 2)* [2014] 3 WLR 555.

[32]T Beven 'The responsibility at common law for the keeping of animals. Baker v Snell, [1908] 2 KB 352, 825' (1909) 22 Harvard Law Review 465.

[33]Lloyd-Bostock, above n 24, at 146.

[34]F Pasquale 'Data-informed duties in Al development' (2019) 119 Columbia Law Review 1917 at 1919.

[35]G Harman 'Moral philosophy meets social psychology: virtue ethics and the fundamental attribution error' (1999) 99 Proceedings of the Aristotelian Society 315; DP Field and PN Hineline 'Dispositioning and the obscured roles of time in psychological explanations' (2008) 36 Behavior and Philosophy 5 at 11–13.

[36]J Hanson and M McCann 'Situationist torts' (2008) 41 Loyola of Los Angeles Law Review 1345 at 1360. On the relationship between control, competence, and responsibility see J Raz 'Being in the world' (2010) 23 Ratio 433.

[37]I Choi et al 'Causal attribution across cultures: variation and universality' (1999) 125 Psychological Bulletin 47; K Levy 'Does situationism excuse? The implications of situationism for moral responsibility and criminal responsibility' (2015) 68 Arkansas Law Review 731 at 746.

[38]Harman, above n 35, at 317.

[39]M Ciurria 'Situationism, moral responsibility and blame' (2013) 41 Philosophia 179 at 184.

https://doi.org/10.1017/lst.2022.52 Published online by Cambridge University Press

contract theory.[40] It also features in Western legal theory,[41] for instance in the 'will theory' of contracts[42] and the 'autonomy doctrine' for attribution.[43] Thus legal fault is often premised on dispositionist notions of intention, control, and consent.[44] The more dispositional the injurer's actions, and the less dispositional the victim's, the more we are likely to fault the former, and seek remedy for the latter.[45] Dispositions are not reserved for natural persons; companies and organisations are commonly ascribed with personalities as well.[46]

(b) The situationist critique

To situationists, however, dispositionism commits a 'fundamental attribution error',[47] being 'the error of ignoring situational factors and overconfidently assuming that distinctive behaviour or patterns of behaviour are due to an agent's distinctive character traits'.[48] The situationist case, which finds support across psychology, moral philosophy, and law,[49] is premised on empirical evidence of human behaviour. One canonical example[50] is Milgram's obedience experiment,[51] where a surprising majority (65 per cent) of volunteers were willing to administer a full course of intense electric shocks (up to 450 volts) to unseen human 'learners' in another room, despite the latter's vigorous, albeit staged protests.[52] Situationists attributed Milgram's results to the power of the volunteers' situation: the gradual shift from the innocuous to the potentially fatal, the experimenter's authority, and the confusing circumstances participants were thrust into.[53] Because we are geared to 'see the actors and miss the stage',[54] these situational forces, though obvious in hindsight, were largely overlooked.

Situationists therefore argue that we assign more moral and legal weight to disposition than empirical truths about human behaviour suggest is warranted. Since 'our attributions of causation, responsibility, and blame — and our assessments of knowledge, control, intentions, and motives — are not what we suppose they are',[55] insofar as law relies on dispositionist conceptions of these doctrines, it risks *itself* committing the fundamental attribution error. If anti-social behaviour is produced more by situation, and less by disposition, than commonly thought, then law's focus on correcting faulty dispositions cannot effectively deter bad behaviour; situational causes of such behaviour must be rectified.

To be sure, Milgram's experiments have been subjected to two waves of criticism arguing that they had been misrepresented and misinterpreted.[56] Nonetheless, modern situationist work, while still

[40]J Hanson and D Yosifon 'The situational character: a critical realist perspective on the human animal' (2004) 93 Georgetown Law Journal 1 at 10–12.

[41]C Haney 'Making law modern: toward a contextual model of justice' (2002) 8 Psychology, Public Policy, and Law 3 at 5–6.

[42]See generally AS Burrows 'The will theory of contract revived – Fried's "contract as promise"' (1985) 38 Current Legal Problems 141.

[43]G Williams 'Finis for novus actus?' (1989) 48 Cambridge Law Journal 391 at 393.

[44]Hanson and McCann, above n 36, at 1368.

[45]Ibid.

[46]See also Part 3(b) below.

[47]Ross, above n 19, at 184.

[48]Harman, above n 35, at 315.

[49]See for example Ross, above n 19; Harman, above n 35; Hanson and Yosifon, above n 20; Ciurria, above n 39; Levy, above n 37; A Kaye 'Does situationist psychology have radical implications for criminal responsibility' (2008) 59 Alabama Law Review 611.

[50]Referenced in Hanson and Yosifon, above n 40, at 150–154; Hanson and McCann, above n 36, at 1362; M McKenna and B Warmke 'Does situationism threaten free will and moral responsibility?' (2017) 14 Journal of Moral Philosophy 698 at 703.

[51]S Milgram 'Behavioral study of obedience' (1963) 67 The Journal of Abnormal and Social Psychology 371.

[52]For details on the experiment see Harman, above n 35.

[53]L Ross and RE Nisbett *The Person and the Situation: Perspectives of Social Psychology* (London: Pinter and Martin, 2nd edn, 2011) pp 63–66.

[54]Hanson and Yosifon, above n 20.

[55]Hanson and McCann, above n 36, at 1369.

[56]I Nicholson '"Torture at Yale": experimental subjects, laboratory torment and the "rehabilitation" of Milgram's "obedience to authority"' (2011) 21 Theory & Psychology 737; D Kaposi 'The second wave of critical engagement with Stanley

https://doi.org/10.1017/lst.2022.52 Published online by Cambridge University Press

referencing Milgram, rests on a broader evidential base.[57] More importantly, social psychologists have shifted from 'strong situationism' towards 'interactionism' – explaining behaviour as interactions between disposition and situation (though their explanatory shares unsurprisingly remain disputed).[58] Thus, the claim is not that situation *alone* drives behaviour, nor that situation is always *completely* missed.[59] In extreme cases, such as the classic gun to the head, situation is prominent enough to be detected.[60] This is consistent with how exculpatory situations such as duress, inevitable accident, and circumstantial reasonableness are not foreign to law. The argument, more precisely, is that law *under-appreciates* situation while over-prioritising disposition. Therefore, while situationism has its own critics,[61] this paper's thesis does not require one to unconditionally accept situationism nor categorically reject all of dispositionism. Rather, the former is advanced as a *completing* rather than competing account of AI systems.

(c) Disposition versus situation in law

Given attribution theory's implications for legal fault attributions, legal scholarship on attribution theory is surprisingly scarce, particularly in the context of AI systems.[62] Situationism has primarily been applied in the context of criminal responsibility[63] and American tort law.[64] Thus, before examining how attributional frames shape the AI discourse, an illustration with a classic English case is useful.

In *Miller v Jackson*,[65] the Millers claimed in nuisance against a cricket club for cricket balls repeatedly landing in the former's property. Holding against the Millers, Denning MR's dissent predictably framed the Millers dispositionally. They were 'newcomer[s] who [were] no lover[s] of cricket' and who specifically 'asked' the court to stop the sport.[66] In this narrative, the Millers had *moved themselves* into their present position. Conversely, the cricket club had 'done their very best to be polite'[67] and did 'everything possible short of stopping playing cricket on the ground'.[68] But the Millers 'remained unmoved'.[69]

The majority's Millers were cast differently. For Lane LJ, cricket balls had been landing dangerously in their property: one had 'just missed breaking the window of a room in which their (11 or 12 year old) son was seated'.[70] To Cumming-Bruce LJ, cricket balls were 'falling like thunderbolts from the heavens'.[71] The neighbouring Milners, and their nine-month-old infant, were also subject to this danger.[72] In this narrative, the residents had merely sought to go about their daily lives, 'picking raspberries in the garden',[73] but simply could not because of the situation they had been thrust into.

Milgram's "obedience to authority" experiments: what did we learn?' (2022) 16(6) Social and Personality Psychology Compass e12667.

[57]HT Reis 'Reinvigorating the concept of situation in social psychology' (2008) 12 Personality and Social Psychology Review 311 at 312.

[58]Ibid, at 313.

[59]Situation can have a *weaker* influence than situationists hypothesise: Ross and Nisbett, above n 53, pp 5–6.

[60]Hanson and Yosifon, above n 23, p 157.

[61]See for example K Kristjánsson 'An Aristotelian critique of situationism' (2008) 83 Philosophy 55; E Mik 'The erosion of autonomy in online consumer transactions' (2016) 8 Law, Innovation and Technology 1 at 6.

[62]See above n 27.

[63]See eg Kaye, above n 49, at 670; Levy, above n 37.

[64]See eg Hanson and Yosifon, above n 20; Hanson and Yosifon, above n 40; Hanson and McCann, above n 36.

[65]*Miller v Jackson* [1977] 3 WLR 20.

[66]Ibid, at 25.

[67]Ibid.

[68]Ibid.

[69]Ibid.

[70]Ibid, at 31.

[71]Ibid, at 36.

[72]Ibid.

[73]Ibid, at 31.

https://doi.org/10.1017/lst.2022.52 Published online by Cambridge University Press

All three judges heard the same evidence, but the narrative each side told differed in the precise manner attribution theory predicts.[74]

In this way, attribution theory yields descriptive, predictive, and prescriptive insights for law. Descriptively, injurers may be cast as actors who chose certain intended actions giving rise to harmful events; victims are vulnerable persons being moved by, rather than moving, those events, and often rely on the injurer's dispositional control.[75] Predictively, the extent dispositional/situational narratives can be sustained for claimants/defendants provides an indication of how parties may argue, how judges may decide, and how those decisions may come to be justified. Prescriptively, situationism suggests that law should be cognisant of narrative manipulation. If our conclusions regarding concepts like volition and control turn on narrative framing, it is worth asking how reliable they are as tools for attributing fault. Notice that, to portray the Millers as situational characters, the majority dispositionise the cricket balls, describing them as 'thunderbolts' bearing down on the plaintiffs. Yet 'if ever there was an item that is moved more obviously by something other than its own volition, it is a ball'.[76] What then about those who struck the cricket balls to begin with?

2. Artificial intelligence as dispositional actors

If balls can be dispositionised to influence law, it is not surprising that AI systems, which appear to behave as humans do, could also be. Since lawyers are not typically trained in the technicalities of AI systems,[77] we naturally ascribe what Dennett calls 'intentionality' towards AI systems so as to explain and manage what we cannot otherwise comprehend.[78] This section demonstrates how far AI dispositionism shapes legal discourse, in the process examining popular conceptions of AI alongside debates on AI liability, personality, publications, and inventions.

(a) Popular culture

In science fiction, AI systems typically present as sentient, embodied robots who reason, act, and want.[79] Influenced by such imagery, popular culture tends to describe non-fictional AI systems as 'evil'[80] and 'biased',[81] imputing to them thoughts and emotions. In 2016, the chatbot 'Sophia' made headlines by answering, '[o]k. I will destroy humans' in response to a question from its creator David Hanson. One contemporary headline reported that a '[c]razy-eyed robot wants a family – and to destroy all humans'.[82]

Did Sophia 'want' to do so, or was it merely programmed to reproduce these words? That is, did the answer stem from 'her' internal disposition, or were they simply coded as a set piece in the chatbot's software? AI experts preferred the latter, arguing that Sophia was a mere 'puppet' with neither free will

[74]See also J Toomey 'Narrative capacity' (2022) 100 North Carolina Law Review 1073.

[75]Control, reliance, and vulnerability, are 'salient features' which Australian courts consider to test for duty of care: *Graham Barclay Oysters Pty Ltd v Ryan* [2002] 211 CLR 540. On English and Australian case law on control see Soh, above n 5, at 230–233.

[76]Hanson and McCann, above n 36, at 1357.

[77]R Calo 'Robots as legal metaphors' (2016) 30 Harvard Journal of Law & Technology 209 at 237.

[78]D Dennett 'Intentional systems' (1971) 68 The Journal of Philosophy 87; D Dennett 'Intentional systems theory' in B McLaughlin et al (eds) *The Oxford Handbook of Philosophy of Mind* (Oxford: Oxford University Press, 2009) pp 339–441.

[79]See N Sharkey and L Suchman 'Wishful mnemonics and autonomous killing machines' (2013) 136 Proceedings of the AISB 14 at 16.

[80]D Leprince-Ringue 'Evil AI: these are the 20 most dangerous crimes that artificial intelligence will create' *ZDNet* (5 August 2020), https://www.zdnet.com/article/evil-ai-these-are-the-20-most-dangerous-crimes-that-artificial-intelligence-will-create (last accessed 25 January 2023).

[81]W Knight 'AI is biased. Here's how scientists are trying to fix it' *Wired* (19 December 2019), https://www.wired.com/story/ai-biased-how-scientists-trying-fix (last accessed 25 January 2023).

[82]M Starr 'Crazy-eyed robot wants a family — and to destroy all humans' *CNET* (20 March 2016), https://www.cnet.com/news/crazy-eyed-robot-wants-a-family-and-to-destroy-all-humans/ (last accessed 25 January 2023).

https://doi.org/10.1017/lst.2022.52 Published online by Cambridge University Press

nor autonomy.[83] Its creators had deliberately cast the robot in a dispositional light as a 'publicity stunt'[84] and 'political choreography' to market the technology.[85] This notwithstanding, Sophia remains an icon for modern AI technologies frequently covered by news outlets[86] and was in 2017 granted legal citizenship in Saudi Arabia.[87]

The dispositional AI narrative is not limited to sensationalist tabloids. By selectively prioritising quotes sourced from AI companies and deliberately drawing parallels between AI systems and humans, the general media constructs expectations of a 'pseudo-artificial general intelligence' that does not exist.[88] In turn, this narrative shapes popular thinking around AI liability. In 2018, history's first pedestrian fatality linked to automated vehicles (AVs) occurred in the United States. One contemporaneous headline reported that a '[s]elf-driving Uber kill[ed] Arizona woman in first fatal crash involving pedestrian',[89] implying the primary culprit was the *vehicle* itself, not Uber the company, nor anyone else involved in the vehicle's development or use. A similar framing emerges from another headline, '[s]elf-driving Uber car that hit and killed woman did not recognise that pedestrians jaywalk'.[90]

(b) AI liability

The law is not wholly determined by lay conceptions of liability. But it may not escape its influence either. The question AVs pose to law is conventionally framed in terms of a missing person problem: when AI replaces human drivers, who – if anyone –is liable for accidents?[91] Notice how the idea of AI 'driving' begins to dispositionise the system: the main actor seems to be 'the AI' itself, but since AI systems are not legal persons, they cannot be liable despite being the perpetrator which dispositionism points towards. Thus, the European Commission has questioned the 'appropriateness' of traffic liability regimes which either 'rely on fault-based liability' or are 'conditional on the involvement of a driver'.[92]

More broadly, Chesterman calls this the 'problem of autonomy' which AI systems pose to law.[93] Since the vehicle acted 'autonomously', it appears that no person, human or legal, can be faulted for the accident. The crux lies in how far AI driving systems (ADS) can properly be said to be

[83]J Vincent 'Facebook's Head of AI really hates Sophia the Robot (and with good reason)' *The Verge* (18 January 2018), https://www.theverge.com/2018/1/18/16904742/sophia-the-robot-ai-real-fake-yann-lecun-criticism (last accessed 25 January 2023).

[84]J Urbi and M Sigalos 'The complicated truth about Sophia the robot – an almost human robot or a PR stunt' *CNBC* (5 June 2018), https://www.cnbc.com/2018/06/05/hanson-robotics-sophia-the-robot-pr-stunt-artificial-intelligence.html (last accessed 25 January 2023).

[85]J Parviainen and M Coeckelbergh 'The political choreography of the Sophia robot: beyond robot rights and citizenship to political performances for the social robotics market' (2021) 36 AI & Society 715.

[86]M Hennessy 'Makers of Sophia the robot plan mass rollout amid pandemic' *Reuters* (25 January 2021), https://www.reuters.com/lifestyle/oddly-enough/makers-sophia-robot-plan-mass-rollout-amid-pandemic-2021-01-25 (last accessed 25 January 2023).

[87]E Reynolds 'The agony of Sophia, the world's first robot citizen condemned to a lifeless career in marketing' *Wired* (1 June 2018), https://www.wired.co.uk/article/sophia-robot-citizen-womens-rights-detriot-become-human-hanson-robotics (last accessed 25 January 2023).

[88]JS Brennen et al 'What to expect when you're expecting robots: futures, expectations, and pseudo-artificial general intelligence in UK news' (2022) 23 Journalism 22.

[89]S Levin and JC Wong 'Self-driving Uber kills Arizona woman in first fatal crash involving pedestrian' *Guardian* (19 March 2018), https://www.theguardian.com/technology/2018/mar/19/uber-self-driving-car-kills-woman-arizona-tempe (last accessed 25 January 2023).

[90]P McCausland 'Self-driving Uber car that hit and killed woman did not recognize that pedestrians jaywalk' *NBC* (10 November 2019), https://www.nbcnews.com/tech/tech-news/self-driving-uber-car-hit-killed-woman-did-not-recognize-n1079281 (last accessed 25 January 2023).

[91]See for example S Shavell 'On the redesign of accident liability for the world of autonomous vehicles' (2020) 49 JLS 243 at 244. Soh, above n 5, at 226 discusses the AV framing problem more comprehensively.

[92]European Commission, Directorate-General for Justice and Consumers 'Liability for artificial intelligence and other emerging digital technologies' (2019) pp 16–17.

[93]S Chesterman 'Artificial intelligence and the problem of autonomy' (2020) 1 Notre Dame Journal on Emerging Technologies 211.

https://doi.org/10.1017/lst.2022.52 Published online by Cambridge University Press

autonomous. Chesterman notes that 'autonomy' requires the ADS to be 'capable of making decisions without input from the driver'; such a system would differ from mere 'automations' like cruise control.[94]

The line between automation and autonomy, however, is seldom clear. Most legal commentators adopt the Society of Automotive Engineers' (SAE) six levels of driving automation, found in a standards document indexed 'J3016'.[95] First published in 2014, J3016 was substantially revised in 2016, 2018, and 2021. Since 2016, the standard has only used 'automation', even to refer to vehicles at the highest levels. The SAE deliberately avoided 'autonomy', arguing that the term could 'lead to confusion, misunderstanding, and diminished credibility' because:[96]

> in jurisprudence, autonomy refers to the capacity for self-governance. In this sense, also, 'autonomous' is a misnomer as applied to automated driving technology, because even the most advanced ADSs are not 'self-governing'. Rather, ADSs operate based on algorithms and otherwise obey the commands of users.

Because the engineers' definition of 'autonomy' only requires that a system 'ha[s] the ability and authority to make decisions independently and self-sufficiently',[97] it encapsulates a range of technologies, such as thermostats,[98] to which attributing *legal* autonomy would be strange. Legal commentaries have nonetheless continued to use the term.[99] Beyond AVs, AI autonomy remains cited as a key challenge to existing liability regimes.[100]

Smith thus identifies the 'inconsistent use of several key terms [relating to autonomous systems] within and across the legal, technical, and popular domains' as a source of 'potential and ultimately unnecessary confusion'.[101] Indeed, the engineering literature itself, displays 'a profusion of concepts and terms related to autonomy'[102] and oscillates between conceptions of autonomy as self-*governance* (i.e. the primacy of internal control) and self-*directedness* (i.e. freedom from external control).[103]

Therefore, the issue here is less a problem *of* autonomy than one *with* autonomy.[104] Both the definition of autonomy and its application to identifying truly 'autonomous' systems are ambiguous[105] and subjective.[106] Since 'automation' frames the system situationally, while 'autonomy' presupposes

[94]Ibid, at 212.

[95]On-Road Automated Driving Committee *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles* (2021), https://saemobilus.sae.org/content/J3016_202104/ (accessed 7 February 2023) (J3016 (2021)).

[96]This has been in J3016 since it was first amended in 2016 and remains in the latest version. See On-Road Automated Driving Committee *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles* (2016) p 26, https://www.sae.org/content/j3016_201609 (last accessed 25 January 2023); On-Road Automated Driving Committee *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles* (2018) p 28, https://www.sae.org/content/j3016_201806 (last accessed 25 January 2023); J3016 (2021), ibid, p 34.

[97]J3016 (2021), above n 95, p 34.

[98]'The rights and wrongs of autonomous systems' (24 July 2021), https://www.saab.com/newsroom/stories/2021/july/the-rights-and-wrongs-of-autonomous-systems (last accessed 25 January 2023).

[99]See eg Chesterman, above n 93; Shavell, above n 91.

[100]See eg European Commission 'Report on the safety and liability implications of artificial intelligence, the internet of things and robotics' COM (2020) 64 final; AI Liability Directive, above n 1.

[101]BW Smith 'Lawyers and engineers should speak the same robot language' in R Calo et al *Robot Law* (Cheltenham: Edward Elgar, 2016) p 83.

[102]J Sifakis 'Autonomous systems – an architectural characterization' in M Boreale et al (eds) *Models, Languages, and Tools for Concurrent and Distributed Programming* vol 11665 (Cham: Springer International Publishing, 2019).

[103]JM Bradshaw et al 'The seven deadly myths of "autonomous systems"' (2013) 28 IEEE Intelligent Systems 54.

[104]Alluded to in S Chesterman *We, The Robots? Regulating Artificial Intelligence and the Limits of the Law* (Cambridge: Cambridge University Press, 2021) p 60.

[105]See also RK Helm et al 'Guilty plea decisions: moving beyond the autonomy myth' (2022) 85 Modern Law Review 133 at 149, offering a similar critique of "autonomy" in the English criminal system.

[106]On the politics behind AV autonomy see MI Ganesh 'The ironies of autonomy' (2020) 7(1) Palgrave Humanities and Social Sciences Communications 1.

https://doi.org/10.1017/lst.2022.52 Published online by Cambridge University Press

and implies disposition, the term one chooses, and the resultant analysis, could be driven by motivated reasoning.[107]

(c) AI personality

The longstanding debate on whether AI systems should have legal personality,[108] was brought into focus by the 2017 EP resolution which proposed limited electronic personality for 'at least the most sophisticated autonomous robots'.[109] The ensuing controversy plays out as attribution theory expects. The 2017 resolution demonstrated a classic, pop-culture informed tendency to dispositionise AI. It emphasised AI autonomy, referring to science fiction to make the point.[110] The expert critique offered a situationist response: first noting that claims of AI autonomy are overblown, and second calling out stakeholders 'in the whole value chain who maintain or control' the AI system's risks.[111] Echoing how AI personality was unnecessary, the 2020 resolution highlighted the situational forces underlying AI systems – their behaviours 'are nearly always the result of someone building, deploying, or interfering with the systems'.[112]

AI personality scholarship demonstrates similar tendencies. Proponents generally offer two types of arguments.[113] First are arguments based on the *inherent* qualities of AI, including but not limited to autonomy, intelligence, and consciousness. For instance, Hubbard argues that, given the Lockean imperative that all humans should be treated equally because we all possess 'the same faculties', any AI system which possesses these faculties should likewise have a *prima facie* right to personhood.[114] Second are instrumental arguments based on the extrinsic usefulness of AI personhood. For instance, Čerka and colleagues argue that establishing liability against AI developers is difficult under present laws because of the AI system's 'ability to make autonomous decisions, independently of the will of their developers, operators or producers'.[115] Likewise, Koops and colleagues identify challenges with determining the applicable law and enforcing it with AI becoming 'increasingly autonomous'.[116] Personality is proposed to bridge this 'accountability gap'.[117]

While dispositionism directly underpins the *inherent* arguments, instrumental arguments implicitly build on it also: legal gaps asserted critically assume that AI autonomy precludes the operation of existing laws. Unsurprisingly, the case against personality essentially contests how far AI systems are autonomous or intelligent.[118] The issue, once again, is whether AI systems are better understood dispositionally or situationally.

[107]Dixon uses 'autonowashing' to describe how AV companies abuse vehicular 'autonomy' to oversell their AVs: L Dixon 'Autonowashing: the greenwashing of vehicle automation' (2020) 5 Transportation Research Interdisciplinary Perspectives 100113.

[108]See for example MS Willick 'Artificial intelligence: some legal approaches and implications' (1983) 4 AI Magazine 5; B Solum 'Legal personhood for artificial intelligences' (1992) 70 North Carolina Law Review 1231.

[109]European Parliament, above n 13, at [59(f)].

[110]See above n 18.

[111]For critique of the 2017 Resolution see P Księżak and S Wojtczak 'AI versus robot: in search of a domain for the new European civil law' (2020) 12 Law, Innovation and Technology 297.

[112]European Parliament, above n 15, at [7].

[113]S Chesterman 'Artificial intelligence and the limits of legal personality' (2020) 69 International & Comparative Law Quarterly 819.

[114]PF Hubbard '"Do androids dream?": personhood and intelligent artifacts' (2011) 83 Temple Law Review 405 at 431. Such arguments may be traced back to Willick, above n 108.

[115]P Čerka et al 'Is it possible to grant legal personality to artificial intelligence software systems?' (2017) 33 Computer Law & Security Review 685 at 688.

[116]Koops et al, above n 2, at 508.

[117]Ibid. See also K Ziemianin 'Civil legal personality of artificial intelligence. Future or utopia?' (2021) 10 Internet Policy Review 1 at 19.

[118]See generally Solum, above n 108, at 1264–1272.

https://doi.org/10.1017/lst.2022.52 Published online by Cambridge University Press

(d) AI publications

Courts considering when an algorithm's developers 'publish' defamatory material the algorithm produces have likewise reached opposite conclusions on similar facts in a manner which attribution theory predicts. Those holding that developers are not publishers typically highlight how there is 'no human input' in the results' production.[119] 'It has all been done by the web-crawling "robots"';[120] the developer merely plays a 'passive'[121] role in facilitating the same. Conversely, courts holding that developers can be publishers stress that they *intentionally* designed, developed, and deployed the algorithm. Thus, Beach J in *Trkulja v Google (No 5)* held that 'Google Inc intended to publish the material that its automated systems produced, because that was what they were designed to do'.[122] McDonald J, in a related case, highlighted 'the human input involved in the creation of the algorithm' and how the defamation was 'a direct consequence' of the search engine operating 'in the way in which it was intended to operate'.[123]

More recently, in *Defteros v Google LLC* the Victorian Court of Appeal reiterated that Google's search engine was 'not a passive tool' but something 'designed by humans who work for Google to operate in the way it does'.[124] This was reversed by a High Court of Australia majority who did not consider Google's role in communicating the defamatory material sufficiently active.[125] The dissenting justices argued that, given how search engines operated, Google was more than a 'passive instrument' conveying information[126] and had 'intentionally' participated in communicating the material.[127]

Every case in the *Trkulja-Defteros* litigation involved the same search engine and operator, but each court's reasoning on publication was shaped by whether they understood the algorithms and its creators dispositionally or situationally. Tracing the EP resolutions, if search companies are not to be liable for defamation, we might describe the content as generated by 'sophisticated', 'autonomous', and 'intelligent' robots. But if they are to be liable, we might emphasise how search outputs are always 'the result of someone building, deploying or interfering with the [algorithm]'.[128]

To be sure, outcome differences in these cases must also be explained by reference to key factual differences that in turn shaped how the complex law and policy considerations surrounding online defamation applied.[129] For instance, in the *Trkulja* cases the search company had notice of the defamatory material; in *Metropolitan* and *Bleyer* they did not. The narrow point here is that the dispositional/situational framing of Google's search algorithms *influences*, although it may not wholly determine, judicial analysis on algorithmic publications. It is also remarkable that every court above was, regardless of how they reasoned, happy to base their framing of Google's algorithms on broad narrations, instead of specific technical details, of how those algorithms operate.[130]

[119] *Metropolitan International Schools Ltd v Designtechnica Corp & Others* [2011] 1 WLR 1743 at [50]; *Bleyer v Google Inc* [2014] NSWSC 897 at [83].

[120] *Metropolitan*, ibid at [50]–[53].

[121] *Metropolitan*, ibid at [49]; *Bunt v Tilley* [2007] 1 WLR 1243 at [21].

[122] *Trkulja v Google (No 5)* [2012] VSC 533 at [18].

[123] *Trkulja v Google Inc* [2015] VSC 635 at [45], affirmed in *Trkulja v Google LLC* [2018] 263 CLR 149 at [38].

[124] *Defteros v Google LLC* [2020] VSC 219, [40]; affirmed in *Defteros v Google LLC* [2021] VSCA 167 at [83]. See also *Google Inc v Duffy* [2017] SASCFC 130 at [151].

[125] *Google LLC v Defteros* [2022] HCA 27 at [49] (Kiefel and Gleeson JJ), [58]–[59] (Gageler J), [212]–[221] (Edelman and Steward JJ).

[126] Ibid, at [100], [144].

[127] Ibid, at [109], [136].

[128] European Parliament, above n 15, at [7].

[129] See GKY Chan 'Search engines and internet defamation: of publication and legal responsibility' (2019) 35 Computer Law & Security Review 330.

[130] For an example of how courts describe the algorithm see *Duffy*, above n 124, at [21]–[30].

https://doi.org/10.1017/lst.2022.52 Published online by Cambridge University Press

(e) AI inventions

The AI and intellectual property[131] literature was recently brought under judicial scrutiny by the 'Artificial Inventor Project' (AIP), which aims to secure 'intellectual property rights for inventions generated by an AI without a traditional human inventor'.[132] The AIP applied for patents worldwide nominating an AI system 'DABUS'[133] as sole inventor. Predictably, the AIP describes DABUS in vividly dispositional terms. The system was described as being 'sentient' and as having 'an emotional appreciation for what it conceives'.[134] To DABUS' creator Thaler, 'DABUS perceives like a person, thinks like a person, and subjectively feels like a person, abductively implicating it as a person'.[135]

Attributing sentience to an AI system – however sophisticated – remains controversial amongst AI experts.[136] Nonetheless, such assertions were submitted to patent offices and courts worldwide to justify granting DABUS the patent. The English filing describes the system as an 'autonomous machine' which 'independently conceived' of the invention.[137] The Australian filing claimed that the invention was 'autonomously generated by an artificial intelligence'.[138]

These filings elicited different conclusions from different judges. In the Court of Appeal's latest decision on DABUS, it was uncontested that the Patents Act 1977 (c 37) requires 'inventors' to be 'persons', which DABUS was not.[139] For the majority, the issue was whether Thaler could apply for the patents as a person 'entitled to the whole of the property in' DABUS' inventions under section 7(2)(b) of the Act.[140] They held otherwise because there was no rule of English property law applying accession to intangible property produced by tangible property.[141] Neither did section 7 establish that a machine's owner owns the machine's inventions.[142] Notably, such reasoning frames DABUS as a mere machine (i.e. tangible property) rather than a kind of (artificial) person. Otherwise, the AIP could arguably have relied on standard rules for attributing one *person's* intellectual product to another.[143]

Dissenting, Birss LJ thought the case could be resolved on section 13(2), which required applicants to identify the person(s) who devised the invention. For Birss LJ,this could be satisfied by stating an honest belief that the invention has no human inventor, and this Thaler had fulfilled.[144] Such reasoning implicitly frames DABUS as something beyond a mere machine. To illustrate, suppose DABUS was a fax machine which, one day, 'autonomously' printed a document detailing the invention. Thaler files the same application stating that the fax machine invented something. It would be difficult to accept this 'belief' as honestly held, whether subjectively or objectively, unless we are prepared to see something in DABUS (a capacity to invent) which we would not see in a fax machine.

[131]See above n 8. For recent scholarship on DABUS see PH Lim and P Li 'Artificial intelligence and inventorship: patently much ado in the computer program' (2022) 17 Journal of Intellectual Property Law & Practice 376; MD Lizarralde and HA Contreras 'The real role of AI in patent law debates' (2022) 30 International Journal of Law and Information Technology 23. These pre-date the Federal Court of Australia's recent Full Court judgment, referenced below n 155.

[132]See https://artificialinventor.com/ (last accessed 25 January 2023).

[133]Short for 'Device for the Autonomous Bootstrapping of Unified Sentience'.

[134]See https://web.archive.org/web/20221001083818/https://artificialinventor.com/dabus/ (accessed 7 February 2023).

[135]SL Thaler 'Vast topological learning and sentient AGI' (2021) 8 Journal of Artificial Intelligence and Consciousness 81 at 108.

[136]See N Al-Sibai 'Researchers furious over claim that AI is already conscious' *Futurism* (12 February 2022), https://futurism.com/conscious-ai-backlash (last accessed 25 January 2023); R Metz 'No, Google's AI is not sentient' *CNN* (14 June 2022), https://edition.cnn.com/2022/06/13/tech/google-ai-not-sentient/index.html (last accessed 25 January 2023).

[137]*Thaler v The Comptroller-General of Patents, Designs and Trade Marks* [2020] EWHC 2412 (Pat) at [5].

[138]*Thaler v Commissioner of Patents* [2021] FCA 879 at [8].

[139]*Thaler v Comptroller General of Patents Trade Marks and Designs* [2021] EWCA Civ 1374 at [54], [102], [116]–[123].

[140]Ibid, at [124].

[141]Ibid, at [124]–[137].

[142]Ibid, at [137].

[143]The majority judgment raises one example of such a rule: ibid, at [129].

[144]*Thaler (EWCA)*, above n 139, at [79]–[80].

https://doi.org/10.1017/lst.2022.52 Published online by Cambridge University Press

The issue under the Australian Patents Act 1990 (Cth) was whether DABUS could be an 'inventor' under section 15(1) of the Australian Act.[145] Beach J's decision, which goes furthest in the AIP's favour, is also most evidently shaped by dispositionism. The judge was expressly against 'anthropomorphising algorithms',[146] and had also rejected Thaler's 'more ambitious label' of DABUS as a fully 'autonomous' system.[147] Nonetheless, Beach J accepted that DABUS was a 'semi-autonomous'[148] system 'capable of adapting to new scenarios without additional human input',[149] and 'not just a human generated software program'.[150] Since 'machines have been autonomously or semi-autonomously generating patentable results for some time now', recognising AI systems as inventors would be 'simply recognising the reality'.[151]

Such reasoning labours under the precise problem *with* autonomy explained above. The judgment does not substantiate why DABUS (or any contemporary AI system) is properly regarded as (semi-) autonomous.[152] While Beach J delves into remarkable detail on neural networks in general and DABUS in particular, the judgment mostly echoes the AIP's dispositional narrative.[153] Autonomy is assumed, not argued. This is surprisingly clear from the judgment, which expressly 'assumes' that the system 'set[s] and define[s] its own goal', has 'free choice' of how to achieve that goal, and 'can trawl for and select its own data'.[154]

As the Full Court's decision on appeal points out, these assumptions were not substantiated by the evidence.[155] Beach J may have been giving Thaler the benefit of the doubt on matters which the patent office left unchallenged but, if so, the judgment should arguably not have purported to make any 'general point' about the autonomy of AI systems meant to 'reflect the reality'.[156]

Unsurprisingly, the Full Court overturned Beach J's decision.[157] The dispositionism which occupied much of Beach J's decision was conspicuously absent from the appellate judgment. Instead, the court observed that while the AI inventor debate was 'important and worthwhile', it had 'clouded consideration of the prosaic question before the primary judge, which concerned the proper construction of' the relevant Australian statutes.[158] The dispositional narrative DABUS was clothed in misled the lower court into conflating assumed fact with non-fiction.

Once a court sees through the ruse, however, the legal analysis and outcome takes on a different complexion. Of course, how far a court dispositionises an AI system does not solely determine whether they will rule in 'its' favour. Outcome differences in the English and Australian courts (before the Full Court's recent holding aligned the jurisdictions) should be explained by differences between the English and Australian Patent Acts and related jurisprudence.[159] That said, none of the patent offices nor courts involved questioned DABUS's dispositionist clothes. Even the Court of Appeal majority accepted without questioning the premise that 'DABUS made the inventions'.[160]

[145] *Thaler (FCA)*, above n 138, at [177].
[146] Ibid, at [17].
[147] Ibid, at [18].
[148] Ibid.
[149] Ibid, at [41].
[150] Ibid, at [41].
[151] Ibid, at [126].
[152] Lizarralde and Contreras, above n 131, at 31–33.
[153] See generally *Thaler (FCA)*, above n 138, at [16]–[56].
[154] Ibid, at [126]–[128].
[155] *Commissioner of Patents v Thaler* [2022] FCAFC 62 at [41]–[44].
[156] *Thaler (FCA)*, above n 138, at [131].
[157] *Thaler (FCAFC)*, above n 155, at [113].
[158] Ibid, at [119].
[159] Ibid, at [122]. Unlike ss 7 and 15 of the Act, s 15 of the Australian Act does not expressly define 'inventors'.
[160] *Thaler (EWCA)*, above n 139, at [137]. The High Court expressly made an assumption, not a finding: *Thaler (EWHC)* n 146 above at [6].

https://doi.org/10.1017/lst.2022.52 Published online by Cambridge University Press

3. Artificial intelligence as situational character

The intuition beneath legal AI dispositionism might be reduced to a variation on Descartes' *cogito*: AI *appears* to think, therefore it is.[161] The more we think they are, the more it ostensibly follows that fault, liability, and personality can and should be attributed to them. But *appearing* to think does not mean machines actually do so.[162] This section explains how today's AI systems operate, before discussing how their actions are determined by their training and deployment situation.

(a) Contemporary AI systems are weak AI systems

A leading textbook defines AI as a branch of computer science focused on creating machines that think or act humanly or rationally.[163] However, Turing's seminal paper argued that when a machine can be said to 'think' was 'too meaningless to deserve discussion'.[164] Instead, Turing proposed an 'imitation game': if a machine mimicked human conversation so well that a human could not tell it was a machine, for practical purposes we may say it is artificially-intelligent.[165]

The Turing test's focus was not on the machine's internal disposition but its external behaviour. This exemplifies 'behaviourist' definitions of intelligence.[166] Insofar as we are then invited to infer internal disposition from external behaviour, Turing's test demands the very inference that situationists contest. But merely *appearing* to speak like a human does not imply the machine is thinking like one.[167] Indeed, likening the Turing test to Justice Stewart's famous 'test' for obscenity,[168] Casey and Lemley argue that defining AI legally may be impossible.[169]

We may leave aside the philosophical question of whether machines *in general* can 'think' and focus on whether AI systems in practical use today 'think' in the sense AI dispositionism assumes. Today's AI systems can be broadly classified into machine learning (ML) versus rules-based systems.[170] ML is a branch of AI which programs computers by using statistical optimisation to infer patterns from data.[171] Such systems are illustratively juxtaposed against rules-based or 'symbolic' AI where decision formulae are manually specified.[172] Since explicitly coded rules pose fewer complications, ML systems are typically highlighted as the source of legal uncertainty.[173] This paper thus focuses on ML systems, though the following arguments apply to both kinds of AI.

Consider an AI system meant to predict recidivism.[174] A rules-based approach may involve the programmer manually specifying the formula below:

$$\begin{aligned} \textit{recidivism score} = 2 \times \textit{no of antecedents} + 3 \times \textit{no of violent antecedents} \\ + 1 \times \textit{first conviction age} \end{aligned}$$

[161] Alluded to in R Abbott 'I think, therefore I invent: creative computers and the future of patent law' (2016) 57 Boston College Law Review 1079; Chesterman, above n 113, at 830.

[162] This tracks classic debates between behaviourist and non-behaviourist 'intelligence': Searle, above n 22; N Block 'Psychologism and behaviorism' (1981) 90 Philosophical Review 5; H Ben-Yami 'Behaviorism and psychologism: why Block's argument against behaviorism is unsound' (2005) 18 Philosophical Psychology 179.

[163] S Russell and P Norvig *Artificial Intelligence: A Modern Approach* (New Jersey: Pearson, 4th edn, 2022) pp 1–4.

[164] Turing, above n 22, at 442.

[165] Ibid, at 433–434.

[166] Searle, above n 22, at 423.

[167] The best-known rejoinder is Searle's Chinese Room argument: see generally ibid.

[168] *Jacobellis v Ohio* 378 US 184, 197 (1964).

[169] MA Lemley and B Casey 'You might be a robot' (2020) 105 Cornell Law Review 287 at 343.

[170] H Surden 'Artificial intelligence and law: an overview' (2019) 35 Georgia State University Law Review 1305 at 1310.

[171] A Zuckerman 'Artificial intelligence – implications for the legal profession adversarial process and rule of law' (2020) 136 LQR 427 at 428–429.

[172] Zalnieriute et al, above n 23, at 432–434.

[173] Ibid, at 428; Zuckerman, above n 171, at 428–432.

[174] This being a well-known, controversial use of AI: J Angwin et al 'Machine bias' *ProPublica* (23 May 2016) at https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (last accessed 25 January 2023).

https://doi.org/10.1017/lst.2022.52 Published online by Cambridge University Press

The only factors this system considers are the offender's (violent) antecedents and the age of first conviction of any crime.[175] With weight three, the number of violent antecedents impacts the overall score the most. Of course, this stylised formula will be wholly unsuited for the task. In practice, these factors, their weights, and the formula for mathematically aggregating them, will be more sophisticated. Deep Blue, the AI chess master, was a rules-based system.

However, specifying the right factors, weights, and formulae can be difficult, particularly if a model is meant to approximate legal principles.[176] ML, by contrast, attempts to uncover the same using statistical computations. Data on offenders – whether they re-offended, antecedent counts, and other relevant factors – would be fed through a 'learning algorithm' which computes correlations between said factors and recidivism. Often, though not always, the algorithm essentially identifies a best fit curve for the data then used for out-of-dataset predictions. To illustrate, the ML process may yield the following prediction formula:

$$score = 1.0 \times antecedents + 1.5 \times violent\ antecedents$$

$$- \ 0.1 \times first\ conviction\ age$$

The key difference between the rules- and ML-based models is that the latter's decision formulae and weights are statistically computed, not manually specified. One widely used learning algorithm, 'gradient descent', begins by initialising all weights at an arbitrary value, often zero. The putative prediction formula is then used to predict outcomes for the dataset. Since setting all weights to zero results simply in predicting zeros (i.e. no re-offending) for all subjects, the initial formula will predict most outcomes wrongly. The weights are then adjusted in a manner informed by the aggregate prediction error.[177]

Inferring weights from data gives machine 'learning' its name. Thus, ML always involves two algorithms.[178] First is the *prediction* algorithm, such as the formula above, which generates the predictions. Second is the *learning* algorithm which produces the prediction algorithm to begin with.

To be sure, this brief treatment does not exhaust the depth and sophistication of rules-based or ML-based AI. There exist a vast library of learning algorithms that approach the optimisation problem differently.[179] Different learning algorithms run on the same dataset may yield different prediction algorithms.[180] Nonetheless, the principles above generalise to even the most sophisticated AI in use today. This includes the 'neural networks' (NNs) which have driven much of the strong AI narrative. NNs are *one* class of ML algorithms typically trained using a generalised version of gradient descent known as backpropagation. Despite the name, NNs have no physical form. They too are statistical algorithms for computing weights from data. NNs are 'deep learning' algorithms because they comprise multiple layers of standalone algorithms ('neurons') whose outputs become inputs to yet more algorithms. This allows NNs to approximate a large class of arbitrary formulae.[181] There is no theoretical limit to an NN's architecture; myriad neuron types may be linked together in myriad ways. Nonetheless, the computer scientists who invented backpropagation noted that this 'learning procedure … is not a probable model of learning in brains'.[182]

[175]These are standard variables for predicting recidivism: N Tollenaar and PGM van der Heijden 'Which method predicts recidivism best?: A comparison of statistical, machine learning and data mining predictive models' (2013) 176 Journal of the Royal Statistical Society: Series A (Statistics in Society) 565 at 570.

[176]Zuckerman, above n 171, at 430.

[177]For a worked example see J Mahanta 'Keep it simple! How to understand Gradient Descent algorithm' *KDNuggets* (2017), https://www.kdnuggets.com/2017/04/simple-understand-gradient-descent-algorithm.html (last accessed 25 January 2023).

[178]J Kleinberg et al 'Discrimination in the age of algorithms' (2018) 10 Journal of Legal Analysis 113 at 132.

[179]For a technical overview see A Mackenzie *Machine Learners: Archaeology of a Data Practice* (Cambridge, MA: MIT Press, 2017). For a legally-oriented overview see Zuckerman, above n 171.

[180]Algorithms for recidivism prediction are reviewed in Tollenaar and van der Heijden, above n 175.

[181]K Hornik 'Approximation capabilities of multilayer feedforward networks' (1991) 4 Neural Networks 251.

[182]DE Rumelhart et al 'Learning representations by backpropagating errors' (1986) 323 Nature 533 at 536.

https://doi.org/10.1017/lst.2022.52 Published online by Cambridge University Press

Our technical detour clarifies two critical attributes of contemporary AI systems.

(b) The problem with dispositionising mathematics

First, both rules-based and ML-based AI systems are *mathematical* systems (of equations). ML's focus is not on any physical 'machine' or hardware, but the numerical weights which algorithms *figuratively* 'learn' from data. While AI systems are often *embodied* within hardware systems such as cars or humanoid robots, putting form to mathematics does not change its inherent nature any more than painting a face on a volleyball should.[183]

Dispositionising maths is, to be clear, not a problem per se. We routinely dispositionise everything from cricket balls to companies and the legal system itself.[184] Ascribing intentionality to systems whose inner workings are opaque to us may be the most practical way to manage them.[185] With these systems, however, dispositionism has limits. Corporate personality only arises when formal requirements are met and never argued solely on the basis of a company's 'autonomy'. Moreover, corporate decisions are ultimately made by people whose minds and wills are, following standard corporate attribution rules,[186] taken to represent the company's.[187] In speaking of corporate 'wants', we are ultimately personifying *human* dispositions, not mathematical formulae. Thus the corporate form is often acknowledged as fiction.[188]

By contrast, lawyers framing AI dispositionally seldom seem to realise they may be personifying maths. Reinforced by science fiction, our dispositionist tendencies lead us to conceive of AI systems as autonomous beings, seeing disposition when we should be seeing situation. This tendency to personify AI has been identified by AI researchers as an 'anthropomorphic bias'[189] and by legal scholars as an 'android fallacy'.[190]

That dispositionism misleads lawyers is unsurprising, for even computer scientists do not escape its grasp. ML parlance routinely describes algorithms anthropomorphically: they have 'neurons' that are 'trained' to pay 'attention' and hold 'memory'.[191] McDermott famously called these 'wishful mnemonics': terms used to reflect what programmers *hope* the algorithm does, not what it actually does.[192] More recently, Bender and Koller argue that 'claims in both academic and popular publications, that [AI] models "understand" or "comprehend" natural language … are overclaims' and that 'imprudent use of terminology in our academic discourse … feeds AI hype in the popular press'.[193]

Legal narratives which dispositionise AI must therefore be scrutinised. Notwithstanding the imagery that wishful AI mnemonics conjure, they are inexact metaphors for inevitably *statistical* computations.[194] To recall, 'neurons' are standalone statistical algorithms which compute numerical weights from data. 'Training' is the process of passing data through algebra to compute these weights. 'Attention' means increasing the numerical weights accorded to outputs from certain parts of the

[183]This reference to *Cast Away* originates from Hanson and McCann, above n 36, at 1355.

[184]ED Elliott 'Holmes and evolution: legal process as artificial intelligence' (1984) 13 Journal of Legal Studies 113.

[185]See Dennett (2009), above n 78.

[186]*Stone & Rolls Ltd v Moore Stephens* [2009] 3 WLR 455; *Crown Prosecution Service v Aquila Advisory Ltd* [2021] 1 WLR 5666.

[187]J Chen and P Burgess 'The boundaries of legal personhood: how spontaneous intelligence can problematise differences between humans, artificial intelligence, companies and animals' (2019) 27 Artificial Intelligence and Law 73 at 84–85.

[188]Sanford A Schane 'Corporation is a person: the language of a legal fiction' (1986–87) 61 Tulane Law Review 563.

[189]D Proudfoot 'Anthropomorphism and AI: Turing's much misunderstood imitation game' (2011) 175 Artificial Intelligence 950; D Watson 'The rhetoric and reality of anthropomorphism in artificial intelligence' (2019) 29 Minds and Machines 417.

[190]NM Richards and WD Smart 'How should the law think about robots?' in Calo et al, above n 101.

[191]For an overview of these terms see Watson, above n 189.

[192]D McDermott 'Artificial intelligence meets natural stupidity' [1976] ACM SIGART Bulletin 4.

[193]Bender and Koller, above n 22, pp 5185–5186.

[194]ER MacCormac 'Scientific metaphors as necessary conceptual limitations of science' in N Rescher (ed) *The Limits of Lawfulness* (Washington DC: University Press of America, 1983) pp 185–203; D West and L Travis 'The computational metaphor and artificial intelligence: a reflective examination of a theoretical falsework' (1991) 12 AI Magazine 64.

https://doi.org/10.1017/lst.2022.52 Published online by Cambridge University Press

network.[195] 'Memory' is particular type of neuron (i.e. computation) which feeds into itself such that previous computations influence subsequent ones more directly.[196] These metaphors make the maths *appear* as if it has its own mind but neither entail nor imply that it does. As Cardozo CJ famously held, '[m]etaphors in law are to be narrowly watched, for starting as devices to liberate thought, they end often by enslaving it'.[197] Likewise, Calo notes that judges' 'selection of a metaphor or analogy for a new technology can determine legal outcomes' surrounding AI.[198]

(c) Mathematical dispositions are not human dispositions

Secondly, even if we wanted to dispositionise maths, maths does not think or act as we do. Whether an AI system's internal formulae are manually specified or statistically learned, its 'disposition' is entirely encapsulated in those formulae. Since these dispositions are mathematically expressed, they can also be mathematically explained. To illustrate, we might say that our recidivism predictor above 'prefers' those with no violent antecedents the most, since its formula weights that factor most. Moreover, these formulae are fixed after training, and only updated if the *learning* algorithm is run on new data. Thus the predictor's 'disposition' is stable and deterministic: the same inputs always produce the same outputs. By contrast, we cannot ascribe numbers to how the human mind weighs factors; these weights can and do change over time.

To be sure, much depends on the specific algorithm(s) used. For large NNs that compute billions of weights across millions of factors, unravelling how the system weights each factor can be prohibitively difficult. Even assuming an AI system's prediction algorithm is stable, inputs received in real-time deployment may be ephemeral, prompting split-second changes in the system's outputs. Such opacity indeed challenges fault and liability attributions where victims often need to prove specific software defects and identify person(s) at fault for those failures.[199] While AI researchers have dedicated an entire sub-field towards AI explainability,[200] explanations created from those techniques are often not the kind law requires.[201]

Opacity must, however, be distinguished from autonomy. An NN may perform ten billion computations and tweak its output ten times per microsecond, but maths writ large is still maths. If one linear regression is neither sentient nor (truly) autonomous, what changes, if anything, when one links together a (hundred) thousand regressions? Opacity does not imply autonomy, even assuming the converse holds. Our legal system is opaque to most laypersons, and the best lawyers often cannot predict how it will behave, but we do not say that it therefore acts autonomously and in a way which justifies legal personality, rights, and obligations. Crucially, unlike humans, today's AI systems cannot act beyond what they are programmed to do, even to fulfil their 'wants'.[202] Our recidivism predictor may 'prefer' offenders with fewer violent antecedents, but it cannot, say, propose laws for reducing violent crime. Likewise, Sophia can only produce textual responses to textual prompts. 'She' cannot take steps towards starting a family or destroying humans. This is not to say that AI systems have no 'autonomy' at all, only that the label attaches primarily in an *engineering* sense.[203]

[195]A Vaswani et al 'Attention is all you need' *Proceedings of the 31st International Conference on Neural Information Processing Systems* (California, 2017).

[196]S Hochreiter and J Schmidhuber 'Long short-term memory' (1997) 9 Neural Computation 1735.

[197]*Berkey v Third Ave Ry Co* 155 NE 58, 61 (NY 1926).

[198]Calo, above n 77, at 237.

[199]European Commission, above n 100, at 15–16.

[200]See generally J Burrell 'How the machine "thinks": understanding opacity in machine learning algorithms' (2016) 3 Big Data & Society 1; B Mittelstadt et al 'Explaining explanations in AI' *Proceedings of the Conference on Fairness, Accountability, and Transparency* (Association for Computing Machinery, 2019).

[201]C Reed et al 'Non-Asimov explanations: regulating AI through transparency' in L Colonna and S Greenstein (eds) *Nordic Yearbook of Law and Informatics* (Swedish Law and Informatics Research Institute, 2022).

[202]Kleinberg et al, above n 178, at 132.

[203]See Part 2(b) above.

https://doi.org/10.1017/lst.2022.52 Published online by Cambridge University Press

Thus, today's AI systems remain instances of Searle's 'weak' AI.[204] The disposition of weak AI systems, insofar as they exist, remain dictated by situation: weights produced from the training data and learning algorithm used, how the tasks they trained to do are defined, and the inputs received in their deployed environments. These invariably involve choices made and actions taken by the AI's developers, operators, and users – batters of the AI cricket ball.

4. Situating AI in law

Diagnosing myths afflicting dispositional AI discourse lends itself to two natural prescriptions. First, legal scholars, regulators, and judges must consciously question and take issue with anthropomorphic AI narratives presented before them. Assertions that an AI system is 'autonomous' cannot simply taken as given as they shape the legal conclusion. Particular care is required because anthropomorphisms can be embedded within seemingly descriptive words.[205] For instance, stating that an AI system 'drove itself' and 'caused' an accident may be grammatically correct, but implies a factual disposition which attracts legal responsibility.[206] Moreover, AI developers and operators have incentives to dispositionise their technology to drum up attention and funding while diverting legal consequences away from themselves.

Secondly, situational AI risks must be deliberately highlighted. Conventional dispositionism centralises the legal inquiry around individual 'bad' actors like drivers, resulting in what Elish calls 'moral crumple zones': human actors who, despite having limited control over a complex system, bear 'the brunt of the moral and legal responsibilities when the overall system malfunctions'.[207] Once the AI autonomy myth is avoided, however, it is obvious how eco-systemic stakeholders collectively determine the risks that AI systems pose to society.[208] Crawford notes that the very idea of AI is inextricably intertwined with the socio-economic forces which build and sustain the technology, calling for 'a topological approach' which 'understand[s] AI in a wider context by walking through the many different landscapes of computation'.[209] Likewise, Edwards argues that AI is 'a system delivered dynamically through multiple hands', involving a 'complex web of actors, data, models, and services' who could be held accountable.[210] Edwards thus critiques the original draft Act for centralising 'primary responsibility…on an initial provider' and 'fail[ing] to take on the work … of determining what the distribution of sole and joint responsibility should be contextually throughout the AI lifecycle'.[211] In other words, the Act rightly foregrounds situational actors but is not yet situational enough. Notably, Edwards' critique also applies to the AI Liability Directive because 'fault' as defined is closely tied to breaches of AI Act obligations.[212]

Arguments for refocusing attention onto organisational stakeholders in the AI risk creation process are thus reinforced by attribution theory. Once we see how extensively an AI system's behaviour is determined by programming, it is eminently foreseeable that errors in building and/or deploying AI systems could harm.[213] Therefore, a situationist framing of the legal AI discourse would shift our focus from individual, human dispositions to collective, sociotechnical *systems*.[214]

[204]R Fjelland 'Why general artificial intelligence will not be realized' (2020) 7 Nature: Humanities and Social Sciences Communications 2.

[205]Lloyd-Bostock, above n 24, at 152–153.

[206]But see the Automated and Electric Vehicles Act 2018 (c 18), ss 3 and 4.

[207]MC Elish 'Moral crumple zones: cautionary tales in human-robot interaction' (2019) 5 Engaging Science, Technology, and Society 40 at 41.

[208]Soh, above n 5.

[209]K Crawford *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence* (New Haven: Yale University Press, 2021) pp 10–11.

[210]L Edwards *Regulating AI in Europe: Four Problems and Four Solutions* (Ada Lovelace Institute, March 2022) p 6.

[211]Ibid, p 7.

[212]AI Liability Directive, above n 1, Arts 4(2) and 4(3).

[213]Mackie, above n 5, at 1320; HY Lim *Autonomous Vehicles and the Law: Technology, Algorithms, and Ethics* (Cheltenham: Edward Elgar, 2018) pp 21–23.

[214]Goldenfein et al, above n 11.

https://doi.org/10.1017/lst.2022.52 Published online by Cambridge University Press

Notably, this does not necessarily mean abandoning existing (tort) law entirely. Negligence standards, given their focus on circumstantial reasonableness, are compatible with situationist models of responsibility.[215] Moreover, once any misconception of contemporary AI systems as science-fictitious, autonomous thinking machines is avoided, existing (dispositional) doctrines generally have fewer problems encompassing AI systems. Recalling the *Trkulja* litigation, once we acknowledge that search algorithms merely produce results their programmers designed and built them for, search companies can be said to have intentionally published those results. It has also been argued that the doctrine of control, clarified for AVs, could be meaningfully applied towards determining AV liability.[216] What needs to change is not existing laws per se, as conventionally asserted, but how the law conceptualises AI systems.

To illustrate, suppose a developer D creates an AI system S that Qs with legal consequence L. Assuming L is a harmful consequence, D would like to avoid being fixed with L and argues that S Q-ed autonomously, independent from D's control, intention, and design. The first step must be to ascertain S's technical nature, stripped of any dispositionist baggage Q presents in. While courts may not have the expertise to delve into technical complexities, those who claim their AI to be autonomous may fairly be expected to prove it.

Next, regardless of step one's outcome, deliberate attention should be paid to situational player(s) who shaped S's behaviour. This points first to D, but might also identify other stakeholders, for instance, if D sold S to operator O. Consistent with standard product liability principles, had O deployed S in an environment which D expressly warned against, O's risk contribution cannot be ignored. This step might therefore identify multiple attribution targets.

Selecting the 'right' target(s) from this list turns on specific laws and facts at play, but the relative contribution each target makes towards determining S's behaviour is a key consideration. If L is a legally divisible consequence like liability, L might be apportioned proportionately to harm/risk contribution. Indivisible obligations like contracts may be best attributed to the party who contributed the most.

To be clear, situationism's insights would be wasted if it were merely used to identify targets for conventional dispositionist analysis. Each stakeholder's contributions should ideally be assessed situationally as well. We should consider, for instance, actions taken by other stakeholders, the scientific state-of-the-art, and inputs received by AI systems from their deployed environments. This explains why commentaries adopting more technically accurate views of AI systems favour apportioning safety and compensatory obligations across multiple stakeholders.[217] Such inquiries may, of course, be more complex and expensive than we are used to. Thus, situationism may support moving more radically towards no-fault systems financed by eco-systemic actors,[218] as well as policies targeting systemic change[219] (eg building AI literacy[220]). However, these proposals fall beyond this paper's scope and are best explored in future work.

Conclusion

This paper situates legal debates on AI within the context of attribution theory and uses situationism in particular as a foil to highlight law's traditionally dispositionist tendencies and critique

[215]Ciurria, above n 39.

[216]Soh, above n 5.

[217]See ibid, at 242; Edwards, above n 210, p 7.

[218]Proposed in KS Abraham and RL Rabin 'Automated vehicles and manufacturer responsibility for accidents: a new legal regime for a new era' (2019) 105 Virginia Law Review 127; Shavell, above n 91.

[219]See also Hanson's 'Systemic justice project', available at https://systemicjustice.org/about-us/ (last accessed 25 January 2023).

[220]European Commission Committee on Legal Affairs 'Opinion on the Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts' 2021/0106(COD), proposed recitals 14a and 14b.

https://doi.org/10.1017/lst.2022.52 Published online by Cambridge University Press

unquestioned AI dispositionism. Folk conceptions of AI permeate the conventional legal, regulatory, and judicial AI discourse, leading to the exact attributional errors that situationists have long criticised. This not only threatens the credibility of legal AI analyses; because dispositional AI narratives are easily manipulable, allowing them to shape legal outcomes is problematic. Overcoming AI dispositionism does not necessarily require total reform; recognising AI systems as situational characters, as recent legal instruments are beginning to do, is sufficient. Implementing this paradigm shift may be challenging, but the more we are interested in an account of AI based on fact rather than fiction, the more we should be willing to abandon fallacious AI anthropomorphisms and re-direct attention to the situational forces driving how today's AI systems 'think', 'act', and harm.

Cite this article: Soh J (2023). Legal dispositionism and artificially-intelligent attributions. *Legal Studies* 1–20. https://doi.org/10.1017/lst.2022.52