

A Legal Theory for Autonomous Artificial Agents

Samir Chopra and Laurence F White

(University of Michigan Press, 2011)

The blurb on the back cover of this book claims that it is about artificial intelligence and the law. This fails to do justice to it. This is a book about artificial *agency* and the law, not artificial *intelligence* and the law. The distinction is important. As the authors say, ‘we deprecate the terms *intelligent agent* or *artificial intelligence* as we wish to emphasize the embedded, social, real-world nature of artificial agents, rather than merely their disembodied intelligence’.¹ Although this is *in part* a work of speculative legal theory about how the law can and should respond to anticipated technological developments, it is also about the here and now. It makes the case, very persuasively in my opinion, that (at least partially) autonomous artificial agents are already with us, and that the law needs to catch up with this fact. What is more, it makes a wide variety of always interesting, and often compelling, suggestions about how that might happen, at the core of which is the argument of Chapter Two that artificial agents should (at least in certain circumstances) be treated as legal agents.

Although, this is a book about the law, profound philosophical issues are never far away. Chapter Three addresses questions about knowledge and artificial agents. In particular, it addresses two questions: ‘In what circumstances should we attribute knowledge to artificial agents?’, and ‘In what circumstances should we attribute the knowledge of an artificial agent to its principal (that is the person, human or corporate, on behalf of whom the artificial agent is acting)?’. The authors offer an analysis of knowledge for artificial agents which draws on ancient philosophical debates about the nature of knowledge, as well as contemporary debates about the practice of law. This requires ‘a delicate balancing act in trying to devise an analysis of knowledge for artificial agents that meshes with intuitions, while not introducing undue complications to the law’.²

Not content with a purely theoretical analysis of knowledge for artificial agents, Chopra and White go on to apply their analysis to a variety of practical legal issues. One of the most interesting of these applications is their discussion of whether email filters can literally be said to read email,

¹ Samir Chopra & Laurence F White, *A Legal Theory for Autonomous Artificial Agents* (2011) 28.

² *Ibid* 75.

and whether companies like Google can violate their customer's privacy by acquiring and using information when no humans have access to that information.

Chapter Four is about Tort Law. Here the authors draw on philosophical debates about the nature of causation, going back to Hume, to discuss the circumstances in which artificial agents should be held legally responsible for harms, and the circumstances in which holding them responsible would 'break the chain of causation' so as to alleviate responsibility from the designer, operator, or owner of the artificial agent.

The fifth and final chapter is about whether, and in what circumstances, the law should treat artificial agents as people. It draws on longstanding philosophical debates about the distinction (or alleged distinction) between the concept of a *person* and that of a *human being*. It also draws on the fascinating history of the evolution of the legal concept of a *person* and the distinction between dependent and independent personhood. The authors make it clear that the personhood of artificial agents need not be an all or nothing matter. Artificial agents may be treated (as corporations and a host of other entities have been treated) as legal persons for some purposes, but not for others. This chapter is important, not so much for the conclusions it reaches, as for the questions it raises. The authors do not offer (or attempt to offer) any precise criteria artificial agents would have to meet in order to be treated as legal persons, nonetheless they do make it clear that there can be no good *a priori* reason to rule out the possibility or desirability of ever treating artificial agents as people.

Although it's both an academic book and a law book, *A Legal Theory for Artificial Autonomous Agents* is almost entirely free of both academesse and legalese. It does, of course, contain some technical vocabulary, and even a bit of Latin, but these things are always explained clearly in plain English. What's more, the writing style is lucid and engaging. The only serious flaws are flaws of omission. For example, there is no chapter on military law, despite the growing importance of artificial agents (such as drones) to military strategy. There is also no chapter on criminal law, although the discussion of responsibility in tort law in Chapter Three has some clear implications for our understanding of criminal responsibility. Such omissions are inevitable in any work, such as this, which opens up a genuinely new field of research, and I look forward to seeing them remedied in future editions.

Dr David Coady
Lecturer
School of Philosophy
University of Tasmania