Sven Nyholm, *This Is Technology Ethics: An Introduction*, (Wiley-Blackwell, 2023), 288 pages. ISBN: 9781119755579 (pbk). Paperback: $34.95.

Tobias Flattery
Wake Forest University

Forthcoming at the Journal of Moral Philosophy
(penultimate draft – please cite the final published version when available)

In recent decades, a number of monographs have been published on specific topics in or subfields of technology ethics. But, so far as I can tell, there are only a couple of monographs on technology ethics *in general* that have been written by academic philosophers and published by academic presses. One of those is Shannon Vallor's *Technology and the Virtues: a Philosophical Guide to a Future Worth Wanting* (Oxford University Press, 2016), which articulates and defends a specifically virtue ethical foundation for thinking about technology and human flourishing. Vallor's book is impressive but not ideal for some purposes, e.g., for teaching a course on technology ethics from an ethical theory-neutral point of view. The other book is Nyholm's *This Is Technology Ethics*, which is an opinionated—but not overly opinionated—introduction to the field of technology ethics, geared toward undergraduate (or even graduate) students or other interested nonspecialists. Thus Nyholm's book is already an important and welcome addition. Indeed, while preparing my own upper level undergraduate course on technology ethics last year, I struggled to find a suitable core text until I found Nyholm's just-published book, which I quickly adopted and used to good effect. Since the book is aimed mostly (though not exclusively) at university students, this review is written largely from a teaching perspective. And from that perspective, the book is, overall, a winner: it provides a gentle but solid philosophical introduction to the problems and questions discussed in contemporary technology ethics (e.g., the nature of technology, the alignment and control problems, responsibility gaps, machine agency, robots/AI and moral status, and more), delivered in a coherently organized package, and written in clear and eminently readable prose.

Although Nyholm doesn't explicitly frame the book this way, it can be thought of as comprising four main parts. The first part provides a basic philosophical background for thinking about topics in technology ethics. It does so by offering a pluralistic discussion of each of the two words in the term "technology ethics" in turn, and then a discussion of the two when combined together. Chapter 1 tackles the questions of what technology *is* and whether technologies are inherently value-neutral, and also previews later chapters. Next, Nyholm provides quick but generally useful primers on several historically influential ethical theories, notably including a couple of non-western views, viz., Confucian ethics and Ubuntu ethics (ch. 2). The first part of the book closes (ch. 3) with an examination of several competing methodologies for technology ethics (e.g., ethics by committee, ethics by applying traditional ethical theories), using the topic of self-driving cars as a case study to explain and compare these methodologies.

The second part of the book covers issues broadly related to control. It begins with a treatment of the problems of aligning technologies with our values and controlling technologies

(ch. 4), before turning to issues relating to how technologies can be used to change and control our behavior and thinking (ch. 5). Chapter 6 addresses how we should understand, and what we ought to do about, responsibility gaps that occur when we haven't managed to control our technologies (e.g., autonomous weapons), sometimes resulting in harms.

The next trio of chapters—the third part of the book—investigate moral questions arising when robots and AIs become increasingly integrated into human society. Nyholm first considers the question of whether, and in what sense, machines can be moral agents, and overviews the main positions in the field of machine ethics (ch. 7). He then turns to the question of whether, in what sense, and to what degree artificial beings can have moral status, whether we can wrong them, and whether they ever ought to have rights (ch. 8). Chapter 9 addresses the questions of whether robots or AIs can be, or ought to be, our friends, lovers, or coworkers. The final chapter considers whether we can and/or ought to merge with our technologies, as is to one degree or another predicted and advocated by transhumanists and posthumanists.

At the end of my course on technology ethics last year, I asked my students what they thought about Nyholm's book. The response was overall very positive. They reported that its chapters served well as general introductions to specific topics in technology ethics, which can then be followed up with reading primary sources on those specific topics—and indeed the book engages with much of the best known recent work in the field. Students also noted especially that the book was easy to read and understand. Nyholm's prose is smooth and engaging, and each chapter is well-organized. He also employs plenty of examples, ensuring that abstract philosophical ideas are tethered to concrete reality. Nearly every chapter opens with a real-life case as a "hook", which I found to enhance teaching . Two examples: first, the chapter (ch. 5) on behavioral change technologies begins by describing a 2018 commercial for Apple's smartwatch, in which a man, after strapping on the Apple Watch, sees successive versions of himself outdoing the other in exercising (e.g., running). The commercial then displays the message, "There's a better you in you." This example framed Nyholm's chapter well, but it also became a central case study animating a long and vigorous classroom discussion, not least because many students use smartwatches to track and change their own behavior. Second, Nyholm opened his discussion of robots and AIs as friends and lovers (ch. 9) with the case of the recently launched AI chatbot called "Replika", which is increasingly popular among people seeking a friend, lover, or even a counselor. Again, this example (and others like it in the chapter) sparked long and lively class discussion, but it also dispelled some illusions that the topic was merely the stuff of science fiction movies.

I applaud Nyholm's efforts to highlight the fundamental philosophical issues ultimately at play in particular applied topics, and to combat potential confusion with careful conceptual distinctions. (Nyholm's analytic philosophical training is evident.) However, I found myself thinking that his efforts, at times, forged a few double-edged swords in parts of the chapters dedicated to control (chs. 4-6). Take as an illustrative example Nyholm's discussion of alignment and control problems (ch. 4). Early in the chapter, he lays out what is, on the one edge, a genuinely insightful discussion of the alignment problem structured around a combination of distinctions between intrinsic/instrumental value, positive/negative value, and value alignment/misalignment. This results in a 13-page discussion of value alignment guided by the following series of section headings: "Instrumentally Positive Value-Alignment of Technologies" (p. 86), "Instrumentally Negative Misalignment of Technologies" (p. 87), "Positive Non-

Instrumental Value Alignment of Technologies" (p. 90), and "Negative Non-Instrumental Value Alignment of Technologies". These distinctions are philosophically astute, and they are delivered about as clearly as can be asked for. Even so—and here I flip to the other edge—I fear that many or most of his non-expert readers' eyes will threaten to glaze over while reading these sections.

Moreover, in an effort to show that value alignment is a concern not only with AI but with technologies in general, Nyholm discusses many other examples in this context. These range from emerging technologies such as driverless cars (p. 86), robot/AI companions (pp. 91, 95), and autonomous weapons (p. 92), to relatively old hat technologies like nuclear missiles (p. 88), data collection systems (p. 88), and even car seats and soap dispensers (p. 89). But in contemporary technology ethics discussions, the term "alignment problem" forcefully calls to mind artificial general intelligence (AGI)—and for good reason, since true AGI and especially superintelligent AGI, if ever developed, will most likely have historically unparalleled implications for humanity. Nyholm's chapter, then, might disappoint some readers at least somewhat, since the above-mentioned series of distinctions and the generality of the discussion result in a relatively small treatment of alignment and control problems related to AGI. But, of course, Nyholm *is* correct that value alignment is a concern applicable to most or all technologies, and it is good for students to know this. And, of course, Nyholm's chapter can be combined with additional readings on the alignment problem for AGI. So the concern presented here is not ultimately severe. Note that similar comments, both mildly critical and ameliorative, apply to parts of Nyholm's treatments of control (ch. 5) and responsibility gaps (ch. 6).

Another notable feature of the book is its approach to using footnotes: virtually all footnotes direct readers to website URLs for podcasts, YouTube (or other) videos, popular articles, and even a number of Wikipedia pages. Academic references are found only in annotated bibliographies at the end of each chapter. I have mixed feelings about this editorial decision. I see the value in pointing readers to more lively and approachable sources. But here are two quibbles, both concerning the stability of these sources. First, while many Wikipedia pages are reliable enough for general purposes, Wikipedia is also well-known for having less than perfect accuracy and imperfectly academically rigorous editing policies. Second, surely some of the footnoted URLs—perhaps especially for YouTube videos—will, over time, become outdated, resulting in readers attempting to access broken links.

The above concerns are relatively minor. Overall, the book is excellent. Readers who want to introduce themselves or their students to the field of technology ethics, and who want to enjoy the reading experience, would do well to pick up Nyholm's book. I myself will likely do it again soon.

Tobias Flattery
Department of Philosophy
Wake Forest University
Winston-Salem, North Carolina, USA
flattet@wfu.edu