Trusting Algorithms in Society 5.0

J. N. Hooker

Abstract Society 5.0 will rest fundamentally on advanced algorithms, but will people trust them? This brief essay examines some factors that may influence future acceptance or rejection of a cybernetically-integrated society. These include algorithmic honesty, competence, transparency, and flexibility, as well as our willingness to relate appropriately to nonhuman intelligent agents.

1 Introduction

Society 5.0 is envisioned as the next stage in the evolution of human society, following hunter-gatherer society, agricultural society, industrial society, and our present information society. The concept is best known as an overarching goal put forward in 2016 by the Japanese Business Federation (Keidanren) [10]. Society 5.0 is comprised of many elements, but perhaps the two most prominent ones are a fusion of physical and cybernetic space, and an intermingling of humans with various other types of intelligent agents to form a "posthumanized" society.

A society of this kind relies on advanced algorithms of all sorts, including optimization algorithms, to power its infrastructure. This raises the issue of whether people will allow algorithms to be intimately and thoroughly integrated into their lives. Will we trust algorithms, or come to loathe them? We are already seeing efforts to resist their pervasive influence [12]. This brief essay attempts to identify some of the factors that may influence the ultimate acceptance of Society 5.0.

J. N. Hooker
Carnegie Mellon University e-mail: jh38@andrew.cmu.edu

2 Trustworthy Algorithms

To begin with, we cannot trust algorithms unless they are trustworthy. Technology that lies to us is already a familiar irritation [9]. The “close door” elevator button and the pedestrian crosswalk button often seem connected to nothing. The progress bar for file downloads sometimes appears to be humbug. Amusement park fans tell us that wait time estimates at queues are systematically inflated. On a more serious level, “privacy settings” in social networking apps create a false impression of privacy even while one’s personal information is being analyzed and distributed.

One might attempt to defend a certain amount of deception as benevolent. The fake crosswalk button could encourage pedestrians to wait for the walk signal. Yet a fundamental problem with deception, benevolent or otherwise, is that it soon destroys credibility. Pedestrians become conditioned to ignore crosswalk buttons, even when they are functional and can save lives. This, in fact, lies at the root of ethical objections to deception—its rationale is self-defeating when consistently acted upon.

One of the advantages claimed for Society 5.0 is that its systematic coordination of needs and resources, based on ubiquitous information, will serve human society more efficiently. By delivering exactly what is required when it is required, Society 5.0 can presumably provide for an aging population and meet other challenges posed by our crowded planet. None of this can occur, however, if people distrust the information generated by the system. In the area of queuing, for example, we already have optimization algorithms that link product pricing to announced delivery times, allowing for more precise equilibration of supply and demand [2]. Similar efficiencies are obtained by informing customers in advance of service delays [3] and call center response times [13]. If these estimates are biased to benefit the company at customer’s expense, they will quickly lose credibility.

Even when algorithms are designed in good faith, people may judge them to be unreliable. Recent research on this issue reports mixed results. While some studies find that people are skeptical of algorithms or prefer human judgment [1, 5, 15], others conclude that people trust algorithms more than is generally supposed [16] or will lend them credence under certain conditions [6]. One conclusion that seems to emerge from the research is that people tend to view an occasional mistake as discrediting an algorithm when it would tolerated in a human. This teaches that a task should not be turned over to algorithms until they are significantly better at it than humans.

3 Transparency and Flexibility

Society 5.0 will also rest on algorithmic transparency and fairness, an issue that has justifiably received much attention in the last few years [7, 17]. At least one conference is devoted entirely to the topic [4]. To take one example, machine learning techniques for evaluating mortgage loan applications are notoriously subject to fairness objections. Loans may be denied to deserving applicants because their ethnic background or residential address correlates with past loan defaults by others. If the applicant asks why the loan was denied, the bank can only say that denial was the output of a neural network that learned such correlations from a training dataset.

Transparency is an area in which optimization has a natural advantage over machine learning and some other AI-based techniques. Unlike a neural network, an optimization model is based on explicit and precisely stated assumptions. Questionable or controversial solutions can be traced directly to the model, which can be adjusted as necessary in a transparent fashion to obtain the desired results. In the case of mortgage loans, an optimization model can award loans in such a way as to maximize net present value for all concerned. Only factors that are directly relevant to an individual applicant's credit worthiness would be built into the constraint set. Such a model may not predict repayment as accurately as deep learning, but it would be ethical: it would honor an implicit agreement between applicant and lender, according to which the applicant divulges personal financial information in exchange for consideration based on that information and not something else.

A related advantage of optimization is the availability of postoptimality analysis. Users can explore the consequences of adjusting the model's assumptions, or examine alternate solutions that are slightly suboptimal or result from different objective functions. A simple example in today's world is the GPS system that offers alternative routes. Yet postoptimality analysis tends to be underutiltized even in the current state of affairs. A conscious effort to make it a routine and pervasive feature of Society 5.0 infrastructure can allow algorithms to meet human needs with greater flexibility. This, in turn, gives users greater control over their lives and may lead to more widespread acceptance of the algorithms.

4 Toward a Posthumanized Society

An intriguing feature of Society 5.0 is that it will be populated by nonhuman intelligent agents that interact intimately with humans. Anthropologists have made the interesting observation that this was the norm until the industrial age [8]. Intelligent beings of various kinds played an integral role in traditional socities, ranging from hunting dogs and beasts of burden to spirits and departed ancestors. It was only in the last two centuries or so that machines replaced beasts and belief in spiritual forces waned, and even then only in certain parts of the world. From this perspective, Society 5.0 returns us to the status quo, one might think that we should able to deal with it.

The difference, of course, is that our new nonhuman companions will be powered by algorithms. Perhaps we can adjust to this, but only under certain conditions. At the very least, nonhuman agents must not masquerade as humans, a practice that quickly erodes trust. The practice has already started with the rise of "chatbots," which have become so realistic and prevalent that California recently passed a law requiring them to disclose that they are not human [14].

Even when there is no impersonation, we must learn how to interact with nonhuman beings that possess some human traits but not others—a skill we seem to have forgotten. We tend to anthropomorphize pets, a habit that seems to flourish in postindustrial Western societies. The habit can extend to technology. There is anecdotal evidence, for example, that nursing home residents sometimes form emotional bonds with rather low-level robots that manage Bingo games. This issue will become more acute as androids become more fully autonomous. We may eventually owe such robots ethical obligations, and vice-versa [11], but they will not be human, and we should not treat them as such. More generally, we must learn to reognize whole new categories of beings and relate to them in a fashion that is appropriate to the algorithms that power them.

References

1. V. Alexander, C. Blinder, and P. J. Zak. Why trust an algorithm? Performance, cognition, and neurophysiology. *Computers in Human Behavior*, 89:279–288, 2018.
2. G. Allon and A. Federgruen. Competition in service industries. *Operations Research*, 55:37–55, 2007.
3. M. Armony, N. Shimkin, and W. Whitt. The impact of delay announcements in many-server queues with abandonment. *Operations Research*, 57:66–81, 2009.
4. J. Boyd and J. Morgenstern, editors. *Conference on Fairness, Accountability, and Transparency (FAT*): Program*. Association for Computing Machinery, 2019.
5. B. J. Dietvorst, J. P. Simmons, and C. Massey. Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144:114–126, 2015.
6. B. J. Dietvorst, J. P. Simmons, and C. Massey. Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science*, 64:1155–1170, 2018.
7. H. Fry. *Hello World: Being Human in the Age of Algorithms*. Norton, 2011.
8. M. E. Gladden. Who will be the members of Society 5.0? Toward an anthropology of technologically posthumanized future societies. *Social Sciences*, 8, published online May 2019.
9. K. Greene. How should we program computers to deceive? *Pacific Standard*, 14 June 2017.
10. Y. Harayama. Society 5.0: Aiming for a new human-centered society. *Hitachi Review*, 66:8–13, 2017.
11. J. N. Hooker and T. W. Kim. Truly autonomous machines are ethical. *AI Magazine*, to appear.
12. K. Hosanagar. *A Human's Guide to Machine Intelligence: How Algorithms Are Shaping Our Lives and How We Can Stay in Control*. Viking, 2011.
13. O. Jouini, Z. Aksin, F. Karaesmen, M. S. Aguir, and Y. Dallery. Call center delay announcement using a newsvendor-like performance criterion. *Production and Operations Management*, 24:587–604, 2015.
14. S. Kunthara. California law takes aim at chatbots posing as humans. *San Francisco Chronicle*, 13 October 2018.
15. M. K. Lee. Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic managment. *Big Data and Society*, pages 1–16, January–June 2018.
16. J. M. Logg, J. A. Minson, and D. A. Moore. Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, 151:90–103, 2019.
17. C. O'Neill. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown, 2016.