Machines as Modulations of Risk: the Case of Algorithmic Decision-Making Systems

This paper frames risk and technical failure in algorithmic decision-making systems using a non-instrumentalist concept of technology derived from Niklas Luhmann. Familiar accounts of technology focus on its differences from and similarities to nature: as instrumental action upon the natural world (Aristotle), as imitation of natural causal relations (Bacon), or as a way that humans mediate their relationship to the rest of nature (Marx). Luhmann offers an alternate definition, grounded in systems theory and anthropologically agnostic: technology is a simplification and insulation of causal relations against interference. This focus on the exclusion of chance allows us to interpret the history of technical development in terms of escape from contingency and failure, especially in the era of high technologies that produce novel risks. Risk is a form of time, a way of relating to the future (as probability rather than destiny, progress, evolution, etc.). However, as greater technosocial complexity extends the duration needed to obtain adequate information for forecasting, risk evaluation itself becomes riskier. The difficulty is compensated by calculating machines that externalize cognition. In principle, such machines seem to achieve the insulation that is the ideal of technicity. However, in the case of some complex algorithmic processes, the project of anticipating future likelihoods itself generates potential crises due to the cognitive inaccessibility of their operations.

Generally, technical systems must find resilience in buffers, redundancies, and failsafes to protect themselves from environmental interference, curbing rather than loosening their openness to contingency. The environment must be increasingly controlled, that is, made technical. Such complexification generates more opportunities for decision-making within the horizon of system operation, but further constrains the technical system's adaptability and the overall scope of human power over events. One ends up with a machine surrounding a machine, as in the case of nuclear power plants wherein the reactor is only one subsystem surrounded by a branching control, safety, and security apparatus. Rather than looking at the history of technology in terms of increased efficiency, then, or in terms of increased capacity for value extraction, we should (also) look at it in terms of the management, reduction, and distribution of risk.

When it comes to ricochet effects of nuclear power, medical intervention, pollutants, and so on, one is concerned with both the scale of potential side effects and their lurking presence as "unknown unknowns". But unlike failures of materials (which in principle could usually have been discovered beforehand) or the emergence of harmful externalities (which by nature cannot be discovered until after the fact), the conditions for any particular algorithmic failure are always present beforehand, yet often inaccessible. Here, risks are due almost entirely to the gap between the speed of machine computation and that of human cognition, as well as the non-intuitive nature of many procedures developed via machine learning. Examples range from familiar problems of bias in machine-learning systems emerging from apparently neutral training data to the occasional stock market "flash crashes" exacerbated by high-frequency trading programs.