**ARTIFICIAL INTELLIGENCE AND ETHICS**

Digital information technology has made information readily accessible to practically anyone, anytime and anywhere. This has had a profound effect in shaping all aspects of our society from industrial manufacturing, to distribution, to the consumption of goods and services. Inevitably, like in preceding technological revolutions, digital information technology’s impact has been so pervasive that we are no longer simply adopting it—doing what we have done before—but adapting to it by changing how we behave.

Today, digital information technology has redefined how people interact with each other socially, and even how some find their partners. Redefined relationships between consumers, producers and suppliers, industrialists and laborers, service providers and clients, friends and partners are already creating an upheaval in society that is altering the postindustrial account of moral reasoning.

We are standing at the cusp of the next wave of the technological revolution: AI, or artificial intelligence. The digital revolution of the late 20th century brought us information at our fingertips, allowing us to make quick decisions, while the agency to make decisions, fundamentally, rested with us. AI is changing that by automating the decision-making process, promising better qualitative results and improved efficiency. Successes by AI gaming systems in defeating world chess champion Gary Kasparov, and world go champion Ke Jie underscore the qualitative aspect of AI that proved to be superior than human experts in computing the impact of current decisions on potential future moves.

Unfortunately, in that decision-making process, AI also took away the transparency, explainability, predictability, teachability and auditability of the human move, replacing it with opacity. The logic for the move is not only unknown to the players, but also unknown to the creators of the program. As AI makes decisions for us, transparency and predictability of decision-making may become a thing of the past.

Imagine a situation in which your child comes home to you and asks for an allowance to go see a movie with her friends. You oblige. A week later, your other child comes to you with the same request, but this time, you decline. This will immediately raise the issue of unfairness and favoritism. To avoid any accusation of favoritism, you explain to your child that she must finish her homework before qualifying for any pocket money.

Without any explanation, there is bound to be tension in the family. Now imagine replacing your role with an AI system that has gathered data from thousands of families in similar situations. By studying the consequence of allowance decisions on other families, it comes to the conclusion that one sibling should get the pocket money while the other sibling should not.

But the AI system cannot really explain the reasoning—other than to say that it weighed your child’s hair color, height, weight and all other attributes that it has access to in arriving at a decision that seems to work best for other families. How is that going to work?

In court, past decisions bind the judges to follow precedent even when the situations are not identical but approximately similar. Consistency is important in justice, government, relations and ethics. AI has no legal requirement of stare decisis. AI decisions may truly be artificial to humans as humans tend to have limited sets of direct or indirect experiences while machines may have access to vast troves of data.

Humans cannot sift through their experiences at a long-range timescale, while machines can easily do so. Humans will rule out factors that are perceived to be irrelevant and inconsequential for a decision, while a machine will not rule anything out. This may result in decisions that do not respect precedent at a scale comprehensible to humans. As businesses and societies turn rapidly towards AI, which *may* in fact make better decisions with a far longer time horizon than humans, humans with their shorter-range context will be baffled and frustrated, eroding the only currency for a functioning society, namely trust.

To understand, how *artificial* the artificial intelligence–based decisions may be, it is important to examine how humans make decisions. Human decisions may be guided by a set of explicit rules, or by associations simply based on consequentialism, or by a combination. Humans are also selective about which information is relevant for making a decision. Lacking selectivity, machines may consider factors that humans deem impertinent in making a decision.

There are innumerable examples of this, from Microsoft [shutting down](https://urldefense.proofpoint.com/v2/url?u=https-3A__arstechnica.com_information-2Dtechnology_2016_03_microsoft-2Dterminates-2Dits-2Dtay-2Dai-2Dchatbot-2Dafter-2Dshe-2Dturns-2Dinto-2Da-2Dnazi_&d=DwMFaQ&c=vh6FgFnduejNhPPD0fl_yRaSfZy8CWbWnIf4XJhSqx8&r=1fzmffI58KtZ2pDBBDolw6rafQBw7LqTj9gz0pSFkEU&m=gThPXBOkYP0K31YBTrV0fIV4KsJgLXpSmyT2s7vwrKI&s=EvnXkPQX7L9vGnHyb_d3j4WaHYWhUXoWkGETeHYYThk&e=) its chatbot Tay after it started spewing incendiary anti-Semitic rhetoric on Twitter, to a Boston University [study](https://urldefense.proofpoint.com/v2/url?u=https-3A__arxiv.org_abs_1607.06520&d=DwMFaQ&c=vh6FgFnduejNhPPD0fl_yRaSfZy8CWbWnIf4XJhSqx8&r=1fzmffI58KtZ2pDBBDolw6rafQBw7LqTj9gz0pSFkEU&m=gThPXBOkYP0K31YBTrV0fIV4KsJgLXpSmyT2s7vwrKI&s=axBmc2pkICyfzHNssmLDXsANlmznPlkr1FQqyWJ8Wec&e=) that found gender associations between words like “boss,” “architect” and “financier” with men, and words like “nurse” and “receptionist” with women. This may be borne-out by *data*, but it stands in contrast with our explicit values. If data-driven processes rely on output generated by these AI algorithms, they will produce biased decisions, often against our ethical values.

ProPublica provided [glaring](https://urldefense.proofpoint.com/v2/url?u=https-3A__www.propublica.org_article_how-2Dwe-2Danalyzed-2Dthe-2Dcompas-2Drecidivism-2Dalgorithm&d=DwMFaQ&c=vh6FgFnduejNhPPD0fl_yRaSfZy8CWbWnIf4XJhSqx8&r=1fzmffI58KtZ2pDBBDolw6rafQBw7LqTj9gz0pSFkEU&m=gThPXBOkYP0K31YBTrV0fIV4KsJgLXpSmyT2s7vwrKI&s=6xaWAWoYPz6mMfriS2eN--1JzATUH81rPC1NZa-2Rp4&e=) evidence of this in 2016. A computer program used by U.S. courts wrongly flagged black defendants who did not recidivate over a two-year period as likely to become repeat offenders at nearly twice the rate as white defendants—45 percent as compared to 23 percent. If a human did the same, it would be decried as racist. AI exposes the schism between our explicit values and collective experiences. Our collective experiences are not static. They are shaped by important societal decisions, which in turn are guided by our ethical values. Do we really want to leave the decision-making process to machines that learn solely from the past and therefore are beholden to it, rather than shaping the future?

Given the scale of AI applications in fields like medical diagnosis, financial services and employment screening, the consequences of any single blip are immense. As algorithms rely on more and more features for improving predictability, the logic governing such decisions becomes increasingly inscrutable. Accordingly, we lose the holistic aspect of decision-making, throwing out all principles in favor of past observations. In some instances, this may be unethical, in some, illegal and in some, myopic. The recidivism algorithm blatantly flouted principles like presumption of innocence and equality of opportunity.

Consistency is indispensable to ethics and integrity. Our decisions must adhere to a standard higher than statistical accuracy; for centuries, the shared virtues of mutual trust, harm reduction, fairness and equitability have proved to be essential cornerstones for the survival of any system of reasoning. Without internal logical consistency, AI systems lack robustness and accountability—two critical measures for engendering trust in a society. By creating a rift between moral sentiment and logical reasoning, the inscrutability of data-driven decisions forecloses the ability to engage critically with decision-making processes.

This is the new world we live in, where complex decisions are whittled down to reflexive choices and reinforced by observed outcomes; where complexity is reduced to simplicity and morality is reduced to utility. Today, our sense of ethics provides a framework for making decisions. It may not be long before our decisions cast doubt on our ethics altogether.