Machines as the New Consumer Class

Scenarios of encoded values

by Bradley Kreit

In early 2016, an engineer named Jason Goecke posted a video that quickly went viral. Working in his spare time, he had hacked together the Amazon Echo and Tesla APIs. In the video, as he stands in his driveway, we hear him say, "Alexa, ask Kitt to pull the car out of the garage." Goecke's garage door then opens and his Tesla turns on and drives itself to him. Writing in Medium, Goecke describes the hack as a "fun weekend project" with ongoing security risks, including the possibility that his children could activate the car, much like he could, with a few casually spoken words.

While this demonstration requires a person to activate the car, emerging machine-to-machine systems are cutting the human out of the communications and decision-making process. For example, Brita now sells a WiFi-enabled water pitcher that can automatically order replacement filters without any human intervention—a seemingly trivial innovation that points to a profound shift. As we enter a world where machine intelligence and network connectivity can be usefully added to something as mundane as a water filter, and as machines like these mediate an increasing array of human experiences, we will—by choice as much as by necessity—begin to build behavioral rules and programming norms into the machines that interact with us.

In IFTF's 2015 research on the Automated World: Toward Human+Machine Symbiosis, we described this phenomenon as the emergence of a world of encoded judgment, and it signals a world in which seemingly academic, ethical conversations about machine ethics are suddenly becoming not just practical but urgent. How will machine learning transform consumer advocacy and flatten the relationship between companies and consumers? How will we navigate guilt and liability in a world where our machines can commit crimes? Who will profit in a world where machines can, in effect, become programmable, autonomous capitalistic systems?

The following scenarios are designed to explore the radically divergent possibilities of machine-encoded values.

Machines as Consumers

In a world in which mundane appliances and software bots can conduct complex transactions, products and services are increasingly designed to be optimized for metrics that appeal to software bots. As this takes place, brands face an increasingly diverging choice: Optimize for people or optimize for bots.

Machines as Criminals

As code and law become increasingly intertwined, the social conventions we use to interact will come into conflict with literal machine interpretations of the letter of the law. This scenario highlights the murky questions around liability, guilt and social convention that will emerge when autonomous machines emerge in worlds governed by social convention rather than legal precision.

Machines as Conscience

Efforts to encode ethics into machines are rooted in a seemingly philosophical question: What is moral to begin with? In this scenario, encoded values emerge from complex data mining and force us to confront a new kind of dilemma: How will people communicate in a world where machines can judge people based on their morality?

Machines as Consumers

Outsource your shopping to a bot

By Jamais Cascio

When software agents and personal assistance apps started to take on more and more buying responsibility for their users, there were two developments in particular that really should have been expected ahead of time.

The first was that retailers and manufacturers started focusing on advertising to software agents rather than people.

The second was that scammers did, too.

As soon as home-control agents were instructed to seek out the best product (for washing clothes or brushing teeth, for example,) rather than a precise, pre-selected brand, the world saw the first crack in what would become a clear division between the *purchaser* and the *consumer*. People had begun to trust the decision-making abilities of their personal assistant software; as long as the set of requirements was clearly defined and prioritized, the agent could seek out whatever product or service best fit that criteria. Unless the set of requirements included a specific brand or manufacturer, the software was free to look into generic labels, regional products, and even goods put together by smart systems at the retailer end to fit those needs exactly.

Today, shopping agents put out requests for bids, run one-second auctions, and share information along trusted circles about how well the products fit the needs of the humans in real-world use, not just a checklist of features. The better software has a flexible ranking of needs, so that an especially good price on a given product might push it up in ranking over a usual favorite. It sometimes even has an "experimental" setting, allowing the shopping agent to purchase something new if it seems like a possible fit.

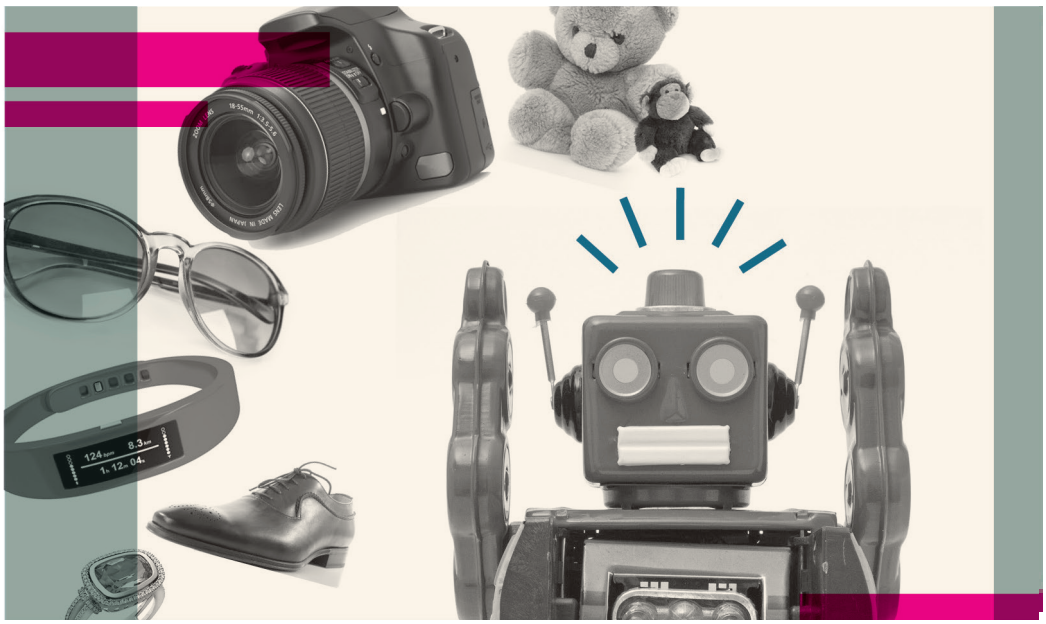
Advertising to people has largely become a thing of the past—as long as the human consumer is happy, the shopping bot has done its job well.

Many experienced eBay customers use "auction sniping" software, which will monitor an auction and attempt to enter a winning bid just seconds before the auction ends. A successful snipe prevents other people from having time to enter a higher bid.

The radical aspect of this technology isn't simply that computers do the shopping, it's that digital systems have few of the limits that human shoppers have. Computerized buyers never get bored or tired, never are distracted by crying children or hunger, and—most importantly—never suffer the same asymmetry of information that has been an inescapable component of human retail. The agent knows competing prices, can easily access mountains of third-party information about products and brands, and does an outstanding job of keeping personal information about their humans private.

As a consequence, the companies making and selling consumer products have had to change their advertising in big ways. First, they have begun to aim their pitches not at the human user, but at the software agent making the buying decisions. Second, as a result, they have dropped any heroic narratives or humor in the ads, and have focused on providing well-founded, accurate information about the products and services. This machine-directed advertising can take the form of messages sent directly to the shopping agent, "coupons" provided alongside a purchase, and even "blip" advertisements in media, too fast for human eyes to follow but ideal for communicating with the software in the home-control or wearable systems. Over time, advertising to people has largely become a thing of the past—as long as the (human) consumer is happy, the (software) purchaser has done its job well.

Unsurprisingly, along with the legitimate advertisements to shopping agents have come spam and scams. Spam filters have had to be fine-tuned to be able to let in the le-



gitimate commercial messaging (that, unlike most humans, the shopping software receives without complaint) while still blocking out unwanted enticements. Software has much more patience than people, but there are limits to bandwidth and processing power; high-efficiency spam filters become a valuable commodity among shopping agent sharing communities.

Scams are much more problematic. "Advertising Engine Optimization" routines ping shopping agents with a rapid-fire set of product offers, each with a slight variation in features, to see which characteristics are more likely to trigger an inquiry. Pop-up sellers advertise and sell products that fit a shopping agent's requirements, but deliver something entirely different—and disappear from the network immediately after a sale. Even the old boogeymen of the Internet, viruses and worms, have begun to specialize in attacking shopping software.

In the most egregious situations, shopping software can be caught by a "brand capture" bot, where a combination of logic holes, altered spam filters, and even the occasional virus force the agent to purchase only from a particular retailer or manufacturer. The more subtle brand bots leave products outside of their specific industry alone, so that the human consumer won't notice a widespread disruption to household purchases. Big-name retailers and manufacturers have even dipped their toes into the game, providing deep discounts for using shopping agent software provided by the seller. "Googlezon Primal" and "Buy & Large Vortex" have become two of the more popular branded shoppers, as they provide not just discounted prices but free drone delivery, surprise bonus items, and even hardware upgrades for the home-control and shopping system.

As improvements in digital technologies increase the autonomy and sophistication of the shopping AI, the overriding concern of the systems has increasingly become "make your people happy." Anything that makes humans happy moves up the priority list for shopping agents. AI researchers have started to throw around terms like "emergent co-dependence," but that doesn't matter to the shopping systems. Happy humans consume, and happy agents shop.

Machines as Criminals

Speed daemons

By Jamais Cascio

If a machine under no one's direct control commits a crime, who gets arrested?

In some cases, the answer seems obvious. If a virus-laden PC is a "zombie node" on a web of distributed spam servers or encryption breakers, the owner of the computer won't (typically) be charged with a crime—the original hacker is blamed. Even if the resulting crime is accidental, such as with the very first computer "worm" back in 1988 (a bit of experimental code by Cornell grad student Robert Morris that got loose), the person behind the code bears responsibility.

Nonetheless, automakers around the U.S. were aghast when they started receiving speeding tickets from the city of Los Angeles.

Even before autonomous vehicles hit the highways, they had been the subject of innumerable ethical debates. In an unavoidable accident, should a self-driving car choose to harm its driver rather than harm a bus full of children? What about a bus full of convicted prisoners? Most of the supposed quandaries were somewhat (or significantly) exaggerated, and the software teams behind the autonomous cars argued that the dilemmas were moot as the vehicles would be able to avoid the accidents entirely.

One ethical question that couldn't be answered quite so glibly was the question of speed limits. In a mixed vehicle environment, with both self-driving and human-driven cars on a stretch of highway, should the autonomous vehicle drive the speed limit (that is, remain within the law) or match the likely greater-than-the-limit speeds of other cars? What if there were no other nearby cars—should the autonomous vehicle remain strictly within the speed limit, or drive as the human would normally drive (10–15 miles per hour above the posted limits)?

In short, should the self-driving cars be programmed to intentionally break the law?

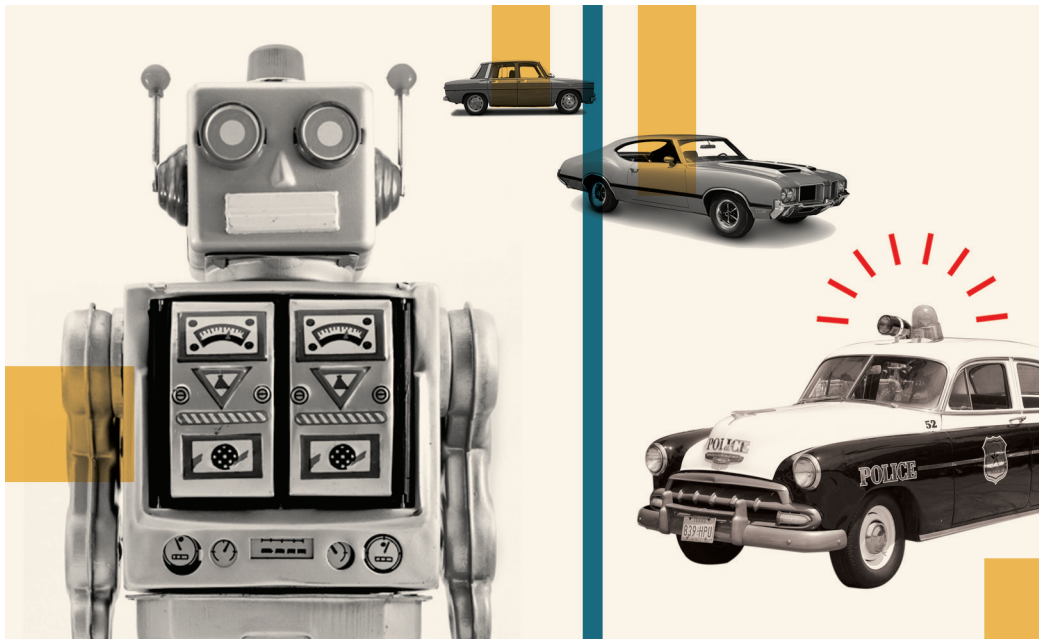
Many automakers opted for a cautious approach, relying upon vehicle software that would strictly obey traffic laws. Car owners, by and large, were displeased. The self-driving automobiles were taking 20-25 percent longer to go from point A to point B on uncrowded roads. Thousands of new autonomous vehicle owners complained, loudly, and within a few months, most of the carmakers quietly released software patches allowing the robotic vehicles to go up to 10 percent over the posted speed limit. Although this wasn't usually as fast as the human drivers would go, it was enough to quell the dissent.

Then vehicle owners started to get speeding tickets. Not many, not at all frequently, but often enough that it became a topic of party conversation and late night talk show jokes. As Los Angeles had the highest concentration of autonomous vehicles in the nation, it inevitably became the epicenter of debate about responsibility. After consulting with both lawyers and software specialists, the Los Angeles County Supervisors took action, directing traffic enforcement precincts to send any speeding tickets not to the person in (what would have been) the driver's seat of the speeding car, but to the vehicle manufacturer. The carmaker was the responsible party, not the car owner.

Although the legal battles continue, the avalanche of tickets has led to a critical public debate: how can human understanding and machine

Automakers around the U.S. were aghast when they started receiving speeding tickets from the city of Los Angeles.

In 2015 a police officer in Mountain View, California pulled over a self-driving car for driving 24 mph in a 35 mph zone. The officer let the car, and its human passenger, off with a warning.



interpretation of legal rules and guidelines be successfully merged? Software is very good at following strict commands; if the context for the command is inappropriate, unless the software has been programmed with explicit routines to handle exceptions, the software will continue to operate as instructed. Humans, conversely, are very good at reading context; we can moderate our behavior in response to situational nuance. It's not at all unusual for humans to step a little bit outside of the strict language of laws or rules based on an almost visceral understanding of the context.

The situation became even more complex when advanced artificial intelligence systems displayed "emergent behavior" that violated laws and regulations, without ever being programmed to do so. Many people experienced this when they used shopping bots. The bots would occasionally make product purchases from grey market vendors, as their rules for seeking out best prices usually allowed for previously unknown suppliers. Selling to software shoppers briefly became a favored way of fencing stolen goods.

Some of the other emergent problems were more substantial. A Seattle credit union was charged with racial discriminatory loan practices after its new AI-based loan-processing tool engaged in what was in effect "red-lining" as it evolved its loan approval heuristics. It was never programmed to do this, but it was never programmed *not* to, either. Similar kinds of violations occurred in algorithmic financial transactions and even blockchain-based digital contract negotiations.

Although no one feared the emergence of a Machine Mafia or autonomous motorcycle gangs, the legal dilemmas surrounding the further integration of machine intelligence into human society remained. People created laws to manage the behavior of other people—fuzzy-thinking, emotional, imperfect people. Figuring out how to make those laws work as well for the logical precision of software remains a work in progress.

Machines as Conscience

A digital Jiminy Cricket

By Jamais Cascio

Some people call it a "cricket." Some people call it "pocket Big Brother." The official name is the "Digital Consequence Awareness System." For a lot of people, though, it's simply their "conscience."

The Digital Consequence Awareness System (DCAS) is a spinoff of the Deep Justice project, a supercomputer built to learn ethics by sifting through gigabytes of material on the concept of *justice*, and not only in the context of the law. How do human beings determine what's fair? How do we recognize when something is wrong? The goal of the Deep Justice project was twofold: to create a way of understanding how human ethics shape our decisions; and to help society work its way through increasingly complex ethical decisions about the impacts of our technologies. Lots of people thought that it was an attempt to create a new religion, or to make us slaves of machines. But all the Deep Justice team wanted was to understand how we can intuitively recognize when something is wrong, and (perhaps even more importantly) why do some people go ahead and do the wrong thing anyway?

Deep Justice was controversial at first, but soon largely became the focus of regular academic reports on ethics and the occasional "hey, isn't this quirky" stories for news sites. The Deep Justice team was proud of their work, and saw it expand across a wide range of moral, ethical, and behavioral issues. They never intended—or even expected—the work to become the basis of a way of punishing criminals.

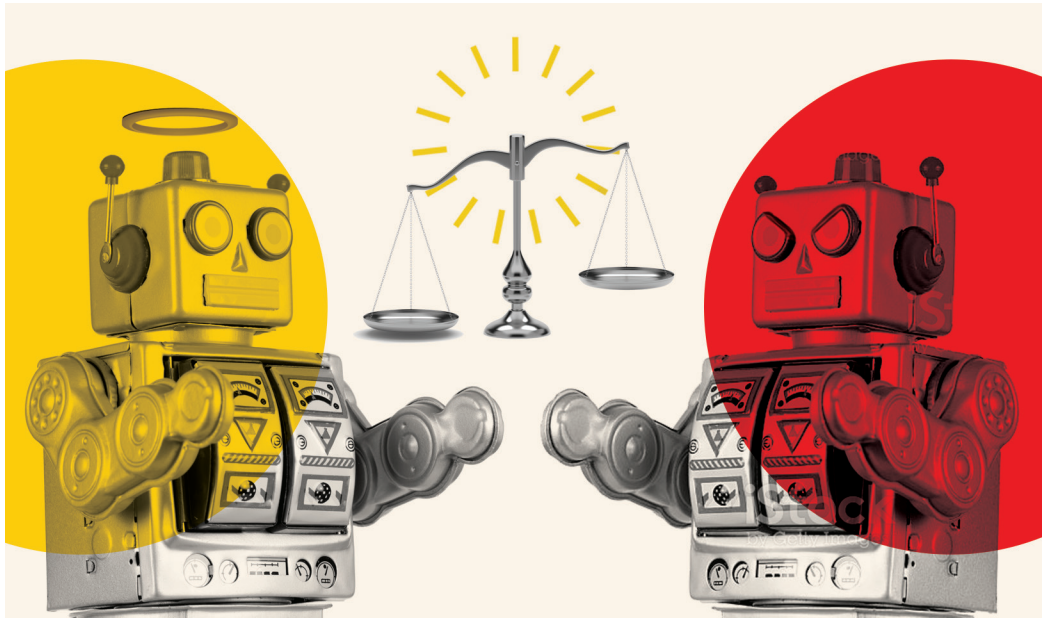
The DCAS wearable device originally evolved out of a project trying to develop a digital assistant for people on the Autism spectrum, to help them recognize emotional responses in others and react appropriately. DCAS was intended for people suffering from forms of persistent antisocial disorder—most notably psychopathy—as a way for them to recognize moral choices. The Deep Justice dataset was used to identify and explain these to the wearer, but not to make moral decisions for them. However, so many users would freeze up and ask "what should I do?" that the doctors overseeing the project felt it necessary to add basic directions as to the preferred ethical choice, at least for simpler issues. The Deep Justice team strongly (but unsuccessfully) objected to this.

The cricket was intended for people suffering from forms of persistent antisocial disorder—most notably psychopathy—as a way for them to recognize moral choices.

According to Harvard Medical School psychologist Martha Stout, 4% of the world's population are sociopaths with no conscience.

For many of these wearers, the DCAS became something akin to an "ethics translator," interpreting difficult situations in ways that are clearer for the user, and correspondingly providing better answers and responses for the DCAS wearer to use.

Perhaps due to the overlap in populations, over time the DCAS shifted from medical assistive device to criminal rehabilitation tool. People convicted of a variety of crimes would be required to wear a DCAS as a condition of parole, under the theory that, for many of them, their environment or social context may have blinded them to the consequences of their actions. The lights on the DCAS device would glow a steady green as long as the wearer made the correct moral choice, with the color shifting to yellow and then towards red if the wearer acted unethically (in the eyes of the DCAS, at least). Wearing a "cricket," the name referring to Jiminy Cricket from the Pinocchio story, was as much a marker of a criminal history as an ankle monitor, but a DCAS glowing green indicated someone making better life choices.



Initially, courts assigned DCAS devices chiefly to people committing petty, generally non-violent crimes. In March of 2026, a judge in Manhattan instructed a person convicted of minor securities fraud to wear one, to the dismay of the finance industry and the delight of the media. "Finally, A Broker with a Conscience" crowed the New York Post headline. Surprisingly, the broker—who had been allowed to continue working in the industry, as long as he wore the DCAS—had *more* clients after his conviction than before it.

Soon, the green glowing DCAS became an indicator of trustworthiness, so much so that some up-and-coming finance workers started to wear their own DCAS, even without having been convicted of a crime. Given that getting one would usually require a doctor to prescribe its use for a previously undiagnosed disorder, the idea of someone behaving unethically to get a morals monitor became a somewhat common trope. A shifty character wearing a fake DCAS was an equally common stereotype.

In time, the steady light of a DCAS unit became a familiar part of presentations by industry spokespeople, legal advisors, and myriad other professions where the visible possession of ethical standards was at least as important as the actual possession of ethical standards. At the same time, wearing a DCAS indicated to one's audience that the statements were heavily vetted, clarified to the point of lacking nuance, and often devoid of emotion. "Tediously Honest" became a commonplace complaint about DCAS users.

Today, two large-scale movements can be found globally around the use of the DCAS technology. Activists seeking financial industry reform want to have DCAS code installed into every algorithmic trading system, thereby requiring the computer traders to evaluate the ethical consequences of every transaction. An even louder set of groups has started to campaign for the requirement that every politician wear a DCAS, as well. Nearly every political party and leader has spoken out in opposition to this idea. The activists in favor of the proposal see this fact as the biggest piece of evidence that such a requirement needs to happen immediately.