From Dead Hand to Flash Collapse: a note on risky machine to machine chain reactions.

Erwan Le Merrer (elemerrer@acm.org), June 2020

*** HEAD: Machine learning based algorithms are now in the wild, and read their environment to react to it. The past has already exhibited relatively benign forms of cascades or chain reactions between these. With the increasing interconnectivity of networks and complexity of algorithms, unpredictable chain reactions might lead to severe flash collapses. This possibility stays largely under the shadow of the awaited strong artificial intelligence) and its associated menace. ***

Due to the sensational progress of machine learning --and in particular neural networks-- for solving cognitive tasks, was resurrected a fear of technological annihilation by an "autonomous artificial intelligence". The figure of the Terminator cyborg now constitutes an equivalent to the Golwin point in some domains of science. Techno-prophets await the day where programs become self-aware, as the indication of an imminent catastrophy. Endless discussions about its near occurrence or impossibility now populate the media. As there are still no evidence that a strong AI can ever arise, one can ask whether the techno-driven apocalypse might not find another vector than killer robots[1].

AUTONOMOUS. The complexity of computer systems and programs continues to increase sharply. This is made possible by the continuous growth of the available processing power, both with specialized types of processors and with the gathering of many of them in datacenter premises. Complex inference algorithms can now directly measure tens of thousands of their environmental variables, and then drive decision-making processes from their analysis. The unprecedented accuracy of their predictions finds applications in self driving cars, financial markets, and video synthesis (with eg, deep fakes).

An historical milestone for complex systems dates back to the cold war era. And this one specifically had retaliation in its code. Perimetr --or Dead Hand-- was built to send a deadly nuclear strike to the USA, in an automated way. Fearing the decimation of the head of the Soviet after such a nuclear strike, engineers foresee this system to operate automatically, when environmental readings seem to indicate that the West has stroke first. This included monitoring for ballistic launches, ground vibrations and radioactivity (indicating that a strike had indeed happened). The system, or algorithm, was then supposed to act as a consequence of those readings to generate a counter strike, without human decisions in the loop. Yet, This fully automatic system was not deployed, and only a semi-automatic version of it[2] was.

HIGH FREQUENCY. Once set-up and activated, this kind of systems lives its program life. And with its own reaction times and temporality, that are far ahead those of humans. More recently, High frequency trading (or HFT) was introduced for that very purpose of operating at super-human timings. The financial markets are now accessed by programs that are designed to read and react, that is to say that they assess their environmental metrics (stock values, number of waiting buy/sell orders, news feeds) and react to these so fast that they overcome "slow" (or human) trading operations. Algorithms are then optimized at the utmost regarding compute time (code optimization) and
action latency (using ultra low latency communication networks). This is key to their prevalence. It turns out that this superiority w.r.t. human action scales like orders of magnitudes faster
COMPLEX INTERACTIONS AND FLASH CRASHES. Interactions of humans and algorithms, as well as from the algorithms between themselves now lead to a tremendous complexity. In fact, all the interacting pieces of software form an emergent system, that is itself out of the reach of simple

1 Why general artificial intelligence will not be realized. Ragnar Fjelland. *Humanities and Social Sciences Communications* **volume 7**, Article number: 10 (2020)

2 The dead hand: Reagan, Gorbachev andthe untold story of the Cold War armsraceTimothy McDonnell aa Woodrow Wilson International Center for ScholarsPublished online: 23 May 2012

institutions or groups of experts. And in consequence no one can test or bound this system. This emerging impact on the markets for instance is far too complex to predict. Salient illustrations of these uncomprehended interactions are called flash crashes. There were massive actions on the markets by those algorithms, leading to the collapse of stock prices as in May 2010[3]. Experts say "the (market) instability can be further exacerbated by HFT algorithms herding one another and thereby transforming instability into widespread crashes"[4]. And with the critical impact of stocks on real life, this is far from being a good news.

ALGORITHMS PRONE TO ATTACKS. Such complex interactions, and their unpredictable effects, question potential feedback loops. A second and major concern is that algorithms reading variables from large environments are prone to attacks. Hostile high frequency trading algorithms are for instance suspected to have as objective to destabilize other algorithms, in order to get some rewards (e.g., lower stock prices after having triggered potentially unjustified sell actions). For instance a tweet from a hacked account is suspected to have triggered billions on the Standard & Poor's 500 Index's value[5]. This is highlighting the constant race in the search for good "signals" predicting future moves on markets, in order to take advantage before the competition.

Even more concerning, many of those automated algorithms are embedding neural networks, that are known to be sensitive to so called adversarial attacks. These attacks leverage the poor understanding humans have about the decision-making process of neural-networks, to trigger arguably incoherent decisions from them. This is equivalent to finding bugs in more classic algorithms, at runtime. Researchers proposed illustrative attacks in critical decision-making systems such as self-driving cars, by simply modifying their environmental inputs.

FUTURE FLASH COLLAPSE? Algorithms were up to now relatively isolated in their domain: HFT algorithms read stocks and information on the Internet, but only operate on markets. Self driving cars read environmental inputs and make decisions for maneuvering the car. With both i) the promises of the 5G to interconnect more and more computing devices and their networks, and ii) the capability of neural network models to handle larger and larger sets of data as input, we are entering a new setup, of unprecedented complexity. And if algorithms can read from larger pools of data, on different networks and on different domains, their creators will probably not refrain to do so, following the moto "information is power".

This opens up for the creation of autonomous agents that will be inclined in reacting as fast as possible to a maximal amount of information, leading to risky interactions. Algorithms can then indirectly influence other algorithms by writing the data some others are going to read, and so on and so forth. In that light, millions of algorithms may very soon depend on each other, in a non controlled way, to make decisions. And these dependencies will create cascade effects that are unlikely to be be interrupted by humans, due to their flash occurrence. The naïve will to create "automatic stops" is bound to encounter the problem that those might be impossible to model and, as their implementation precisely depends on the overwhelming complexity of the data and algorithms themselves.

We are then possibly already in a era where unstable clusters of algorithms are getting more and more interleaved, forming a emergent system reaching all domains of online interactions and decision making processes. The question now becomes: "how can a major chain reaction not occur in the coming years?".

---

3 The Flash Crash: High-Frequency Tradingin an Electronic Market. ANDREI KIRILENKO, ALBERT S. KYLE, MEHRDAD SAMADI,and TUGKAN TUZUN. THE JOURNAL OF FINANCE•VOL. LXXII, NO. 3•JUNE 2017

4 High-frequency Trading, Algorithmic Finance, and the Flash Crash: Reflections on Eventalization. Christian Borch. Economy and Society, Vol. 45, No. 3-4, 2016

5 Social Media, Financial Algorithms and the Hack Crash. Tero Karppi, Kate Crawford. Theory, Culture & Society, Vol 33, Issue 1, 2016.