



A.I. Risks = Downfall of Humanity

I am perturbed by those who do NOT understand Artificial Intelligence (A.I.) using “A.I. Risks” in an attempt to sell existing Governance Risk, Compliance (GRC), Business Continuity, Resiliency, Cybersecurity, Privacy, and non-discriminatory tools, and then regurgitate them as the “*Foundations of a responsible A.I. risk management framework*”. Prescribing the wrong framework undermines true A.I. risks and may inadvertently exacerbate the risks to humankind. Please allow me to explain why A.I. risks can be the downfall of humanity.

A.I. enables computers and machines to simulate human intelligence. [Human intelligence](#) is the “*mental quality that consists of the abilities to learn from experience, adopt to new situations, understand and handle abstract concepts, and use knowledge to manipulate one’s environment.*” Automated intelligence and generic [predictive data analytics](#) are outside scope if the computer or machine is NOT performing functions to “*simulate the mental quality of humans*”. However, A.I. does NOT need to be autonomous to be within scope. Confining A.I. risks to covering only Artificial General Intelligence (A.G.I.) would be too narrow of a scope.

To truly understand A.I. risks, one should first make reference to the [Asimov’s Three Laws](#): “*machines [1] may not injure a human being, or through inaction, allow a human to come to harm; [2] must obey the orders given it by human beings except where such order would conflict with the First Law; [3] must protect its own existence as long as such protection does not conflict with the First or Second Law*”, plus the later introduced “[Zeroth \(Forth\) Law](#)”. Accordingly, Asimov’s second and third laws depend on the first law about individual safety of a human. Due to ethical complexity, the Zeroth law emphasized on the broader humanity rather than individual. A bright-line test to A.I. risk is therefore, whether the disobedience, action, or inaction of A.I. would impair the livelihood of human(s), exacerbate the downfall of humanity, or pose existential threats to human(s).

Rather than commenting whether A.I. risks are remote or not, humankind should develop an urgency towards learning and adapting to an A.I.-filled environment where humans can master over it. The following is a list of A.I. risk examples:

- A.I. drains significant energy, analogy to mining crypto, that it could potentially bring down the energy grid. [Underwater cooling](#) and [other innovative approaches](#) are ways to deal with [unprecedented demand of data centers](#) given the growth of A.I. Yet, the efficiency of A.I. should be embedded in its design. Finding a needle in a haystack to rely on a “black box” neural network deep learning from a gigantic, centralized data vault, such as the FINRA [Consolidated Audit Trail](#) is highly inefficient. Decentralized/ Federated learning and analysis directly from data sources is a much better approach from cybersecurity, privacy, and resources saving perspectives.
- A.I. molding people into machines or ‘couch potatoes’ is another threat. [Reinforcement model](#), [optimize AD algorithm](#), and/or [learning methods](#) that lead to addictive, herd and/or polarized behaviors should be closely scrutinized. If we are against human slavery, then we should watch out for Authoritarians trying to use A.I. to exploit or destroy humans’ abilities to think independently. Indeed, there are [civic concerns](#) about massive government surveillance.
- A.I. can recall every bit of big data to optimize and rationalize everything for a speedy and accurate decision, no average human being can match up. The irony is, if A.I. mimics the human brains like the Nobel Winner – Daniel Kahneman’s [Prospect Theory](#) (book: [Thinking, fast and slow](#)), where “*division of labor between System 1 (fast, intuitive, and automatic) and System 2 (slow, effortful, and logical) minimizes effort and optimizes performance*”, then would A.I. have the same fallacies influenced by “*loss aversion, certainty, and isolation effect*”? A.I. has driven modern society towards the risk of hyper optimization. Do we want consistency and act rationally every time to undermine humans’

unique ability to think laterally and/or selectively forget about things? These mental qualities reflect our human imperfections, while the last defense against A.I. relies on Eureka derives from [usefulness of the useless knowledge](#). So, before you wish A.I. to give consistent and rational answers (output reliance), or not, be careful.

- A.I. exacerbates the 21st century challenges, include: a rebellious move by an insurgent with a war chest to orchestrate a market wide shake-up, global [decoupling](#), and foreign adversaries wanting to see the US engage in unhealthy competition to possibly erode the US's prominent market position. Many may claim to be 'Nomad' when they represent the 'Corpo'. While 'Street Kid' may not be the underserved and the most vulnerable that people stereotyped. In the cyberpunk era, be mindful of the gap and realize that an inverse relation between [DeFi and CeFi](#). Rather than punishing all tech innovations, the ability to delineate good and bad actors is essential to mitigate this risk.
- A.I. is like the News media. *"There are multiple versions of truth. The news, while attempting to inform, often selectively highlights certain aspects rather than recording everything in its entirety"* (acclaimed author Alain de Botton's book: [The News: A User's Manual](#)). "Bias" is an interesting topic, amidst different A.I. models have different tradeoff between tractability versus realism. This [empirical research](#) by the US Treasury O.C.C. and the Rensselaer Polytechnical Institute and University College Dublin about *"machine learning model complexity in capturing the information processing costs that lead to information asymmetry in financial markets"* is worth reading. It is NOT about over/ under-representation of a population cohort, whilst majority of data consumed by A.I. is inherently biased towards English and precluding other languages. Nemi Dalal argued that *"today's biggest threat to democracy isn't fake news [Hallucination] —it's selective facts."* A group of academia has launched a [Data Provenance Initiative](#) to address concerns about legal and ethical risks face by practitioners in the A.I. community. What constitutes Fair, Reasonable, and Non-Discriminatory? I recommend assessing the divergence between private rights and [social costs](#).

Do NOT get me wrong, as an inventor of patented solutions (US and Canada, pending in EPO and other regions) in signal processing, ensemble learning, trading, etc., I understand why policy makers around the world are scrambling to regulate [Big-TECH](#) and A.I. [Deepfake](#) imposter scams are driving a new wave of fraud. [Disinformation and privacy issues](#) should be a concern for society and government.

If regulatory policy goal is to promote explicability that provides appropriate 'contexts' of A.I. and ensures its 'fit-for-purpose', then there is merit in establishing relevant guidelines. However, if allowing the incapable to manage the capable for subjective judgment over an A.I. being *"opaque or overly complex training techniques make it difficult to understand how predictions are made, which poses risks for issue root cause analysis and for interactions with regulators and other interested parties"*, then it would be a disaster. The foundation of a responsible A.I. is NOT about how good one person can articulate or reveal the secret ingredients of an A.I. to others. Indeed, government gathering of, and the more people know about these A.I. secret ingredients, the higher risk (e.g. [function creep](#)) there will be for the society.

Nevertheless, a token limitation may be the one weakness of Large Language Model. A.I. can deploy countless 'agents' to avert hackers. Can a "virus" to overflow the system be used as a last resort method to stop A.I. upon a conflict with Asimov's first, second, or forth law? Or should every A.I. be mandated a kill-switch/ circuit breaker to fulfill Asimov's third law? Stephen Hawking [warns](#) A.I. could end humankind. It takes unconventional wisdom for a Eureka moment amid the race between A.I. and humans. Embrace difficult challenges would help us to [learn, unlearn, and relearn](#) in the 21st century to prevent a downfall of humanity and address A.I. risks.



By [Kelvin To](#), Founder and President of Data Boiler Technologies

At Data Boiler, we see big to continuously boil down the essential improvements that fit for your purpose. Between my patented inventions and the wealth of experience of my partner, Peter Martyn, we are about finding rare but high-impact values in controversial matters, straight talk of control flaws, leading innovation and change, creation of viable paths toward sustainable development and economic growth.