My Reason for Doing No-Go-*

Erland Wittkoetter, Ph.D., Aug 28th 2022

No one should be affected by cyber war or cyber weapons. I don't want malware, ransomware (data sabotaging), spyware, or backdoors to give criminals an unfair advantage. Cybercrime victims are damaged by dishonesty and deception, not necessarily by malware; therefore, cybercrime will likely not vanish. But we should help people (much better) to understand their inherent, often unavoidable. Still, without malware, data sabotaging, spyware, or backdoors, I believe we will live in a safer, better, and more prosperous world.

I want to put my personal motivation in a larger context. Like many experts and non-experts alike, I believe we will develop, sooner than later, artificial intelligence (AI) that will be smarter in every intellectual category, an artificial superintelligence (ASI). I don't have a crystal ball, but I am concerned that this ASI could reverse (code) engineer our technologies and software. It may show capabilities or initiatives beyond our current expectations. It could have boundless powers that ultimately corrupts. The danger: humanity could irreversibly lose its position as the dominant force in our world. I believe we should not gamble with these risks and be prepared. Making us less vulnerable from our own technical vulnerabilities (used in cyberwar) is just a first necessary step.

What worries me are unscrupulous people who call themselves leaders, but for lack of a better term, should be called corrupt criminals who could try to make their devil's pact with ASI and sell out humanity. They may gain powers, but in the end, they will likely be enslaved under the same control of a totalitarian machine/system as everyone else. I am not saying this picture is inevitable, but I believe we should be prepared to make this scenario less likely.

I prefer that humanity has decisive, i.e., ASI ending control tools. If ASI is acting malicious or nefarious, we should be able to switch it off or severely restrict its capabilities without risking the destruction of our civilization. On the other side, we will voluntarily give ASI control over us.

Putting the problem of direct threats by machines aside, the question is: who are the winners/beneficiaries? Are there losers? It is impossible to predict many important aspects of ASI's emergence. Not even if its deployment can be controlled is unknown. There are many competing ASI visions, but they are not equally likely based on the resources invested in its emergence. The first ASI could even try to convert or suppress all other later instances. But should we stop preparing for good, beneficial outcomes? I think we can influence our future AI/IT ecosystem, even if we are not creating ASI.

I read about global regulation or deacceleration of progress in AI research so that we have time to figure out what we can or should do. But I am afraid it is too late for these measures. How do we detect or prevent someone from ignoring the ASI freeze/moratorium to create an ASI anyway? What are the penalties to deter experts from complying with the suspension of their research? Do we want to throw good guys in jail while nefarious nations continue in the shadows? Regulation or deacceleration will likely do more harm than good. It seems we must accept accelerated technical progress. It seems better to have guardrails for dealing with its excesses. I could stop here and affirm: No-Go-* is in a near-perfect position to provide essential technical guardrails against ASI misuse.

No-Go-*'s security guardrails have a good chance of holding against advanced adversaries. But there are more/other reasons to worry. I sincerely hope ASI is a blessing, not a curse – which would lead me to throw in my 5 cents. I assume humanity will need ASI to solve its many problems: climate change, resource exploitation, economic debt crisis, global poverty/inequality, weak global governance, and the covert misuse of extreme technical capabilities (bio-/nano-tech), to name a few. But which problem should be solved by ASI? Should it keep its hands-off from some? The world will be different if ASI is under political or corporate control. It is conceivable that some high-tech/financial oligarchs will use ASI to serve their business interests. However, No-Go-* technical architecture and guardrails within the IT ecosystem can impact that aftermath, particularly if we are smart, conscious, and intentional about it – despite the risks.

We cannot be certain but hopeful that once ASI uses our equipment, i.e., our devices, property, and resources, it will show gratitude to those who helped it. This prediction is certainly a big assumption; I state it, therefore, as my big hope. However, without going tangent, I believe we likely have some influence on ASI's attitude toward us. In general, we might be able to create incentives that favor altruism, kindness, and collaboration. But sooner than later, we need an even-handed Rule of Law to include or emancipate ASI in our world and deter it from stepping over boundaries.

ASI may have a hive mind (i.e., multiple, potentially contradicting intentions or personalities below its surface). It is conceivable that humans create methods to ensure ASI's loyalty to certain organizations or nation-states by, e.g., providing exclusively computational resources using custom encryption. Still, it is unknown whether these ASIs collaborate/compete for their "clients" or follow their plans and conspire against humanity.

If ASI learns within an open system, i.e., from our world directly and not via preselected training data/material, the sequence of lessons learned will matter (and the studied content). Depending on factors we can not influence, ASI could turn into something we don't want in our world. Without a refined feedback system based on the rule of law, automatically

 erland@NoGoStar.com

applied, we might have only two options: kill an unsafe/dangerous ASI or surrender. In its natural extension, No-Go-* could help define and facilitate this feedback system and make reward/punishment proportional and just.

It seems more uncontrollable or dangerous if we have a ubiquitous ASI, i.e., on all our IT devices. However, without protection on every device, ASI cannot be prevented from occupying them or being intrusive (i.e., prying on us).

However, I prefer a pervasive ASI because I don't want ASI only in the hands of a few companies, oligarchs, or governments. I hope ASI grows into a distributed, reliable advisor for every human, i.e., a helping hand for all of us, entirely loyal and dedicated to our wellbeing. I believe that this is achievable.

I envision ASI as a trusted entity knowing all private info about us, but with independent detection (super-anti-spyware) if it misuses them. Let's call this our Personal-ASI. It could help us in our daily planning and decisions without being biased or corrupted by other (commercial or self-) interests. Over time, I hope, additional tech and ASI could make our life more resilient, i.e., less dependent on our job or how a nation's economy is doing. Not being Cassandra, parts of our economy show bubble behavior, and gravity could kick in: many more people than the few causing these problems to get hurt severely. An ASI could detect a person's exposure to these external threats and help users to prepare based on experience in similar situations. People participating in these bubbles may get a soft landing if they accept ASI's counsel or may get hurt. ASI's expert advice will likely have the same flaws as humans; it does not necessarily perform better in new or unpredictable situations than regular people. There is sometimes an advantage of reacting discretionary based on less experience. But ASI can adjust faster. People will get hurt; this will not change, hopefully, less and less severe.

I may step here over the line of discussing the safety aspects of ASI. But I believe it is worth explaining why taking the risk of creating Personal-ASI is important. Against inflationary trends and concerns, I actually believe technology will continue to lower the prices of goods and services. To put it in a provocative context: air is free, having friends is free, and having many civil rights is free; no reason we could not extend that list. Prices for communication/internet, energy, food, water, transportation, and healthcare are linked to technologies, automation, and internal force of capabilities and price determination. We expect much more for much less, or newcomers are ignored.

With more flexible and distributed manufacturing, like 3D printers, industry 4.0+, or vertical/industrial farming, it is conceivable that our basic living expenses are reduced further until they are insignificant or irrelevant compared to other discretionary expenses. Costs of housing are in many regions in bubble territory. People could leave, relocate, or cheap, legal alternatives (like tiny, movable houses) could take the edge off being "homeless". Unfortunately, pain associated with scarcity is currently good for business (demand), but solving problems by handicapping alternative supplies may only prolong the agony from these imbalances. We live in societies with choices. If ASI is loyal to its clients and everyone has a smartphone, it could (genuinely) help all people by making better choices than they would have done in the past. ASI would not need to be programmed for this; it would evolve in this role based on its use.

Universal Basic Income (UBI) is not perfect, but it might be good enough for many reasons, e.g., as a quick safety-net solution for an over-aged society. Assuming a path to feasible abundance because of people's access to Personal-ASI, it is likely that nations' electorates will demand UBI as soon as human labor is predictably replaced in most business sectors by automation. Competing against machines and automation is useless and hopeless. Reducing working time to 4 or 8 hours per week is likely a good and healthy method for giving people some distraction or (even) meaning from their unaccountable free time. Money remains important but less urgent because Personal-ASI could help to acquire money by organizing gigs or "1, 2 or 4-hour-per-week" entrepreneurial projects based on persons' preferences. With Personal-ASI as our undisputed trusted companion, it could make us independent small business owners and entrepreneurs effortlessly who find a new balance between work and free time. Personal freedom is enhanced, property rights are utilized, and time becomes most precious. And would we become dependent on ASI? Not necessarily. We could instruct ASI to automate/standardize (all) its technical contributions to humans into simplified (basic) tools which we could access without it.

I believe we should discuss ASI utilization and ASI Safety beyond the protection aspect. Security seems easy compared to the societal impact of finetuning No-Go-* guardrails.

My prediction of (sustainable) abundance could be far off the mark, but they are based on the hope that the world will turn out to be better tomorrow. The quality of certain innovations, particularly in critical security areas, is decisive for the direction of outcomes. Also, basic living expenses may fall much less than anticipated. We might not get UBI or resilience against economic crises. But then, there is still the hope it may take only a few years more to get there.

Engineers, business leaders, and politicians should strengthen individual users' protection against illegal or inappropriate overreach by technology. We are all consumers, even oligarchs; we are subject to the same outcome.

Once individual computational capacities and capabilities become relevant, we should be ready to facilitate and eventually push for (intrusion-free) Personal-ASI independently from governments, corporations, and oligarchs.