The map of agents which may create x-risks

Type of agent	Subtype	Details	Illustrations	Probability that this type of agent will be responsible for global catastrophe	What could be done to lessen risk from this specific type of agents	Explosion Type of risk typical for the agent	Bio Type of risk typical for the agent	Nuclear Type of risk typical for the agent	AI Type of risk typical for the agent	Collider Type of risk typical for the agent
1. No agent	1. **Nature**	Any natural process which could kill humanity, more in natural risks map		Very probable, but on very long term		Asteroids, gamma ray bursts, Sun superflares, supervolcanos	Pandemics			
	2. **Anthropogenic risks** (intersections with nature)	Global warming, resources depletion, pollution etc.		Probable, if will not be prevented	Raising awareness and global cooperation					
	3. **World-system** World as one evolving system	Hyperbolic acceleration and singularity. System crisis and oscillations before the singularity.		Not clear						
2. Processes, resulting from collectible behavior of many people, not a single agent	1.**Arms race**			Seems probable, but includes actions of some other agents, like superpowers, presidents etc.	International negotiotions and treaties	New weapons	New bioweapons	Many nukes	AI arms race	
	*2.***Natural selection**	Valuable traits disappear		Probably will not happen in the near future						
	3.Ascending economy; **capitalism**	Money and short term reward driven society. Competition favor less altruistic agents		It constantly distorted all our efforts by corruption, but long term perspective and risks are not clear	Socialism? Regulation?		Black market		Ascending economy as a type of unfriendly AI	
	4.**Value drift**	Civilization evolves into something completely different from what we currently value (Whimpers in Bostrom classification).		Not soon						
	5.Memes dissemination: **superdrug**	Dangerous ideas could propagate or self-propagate like viruses, one such idea is a super-addictive drug recipe.		Not the main risk	Infromation control?		New addictive drugs			
	6.**Techological progress**, Moore's law	Creation of many new risky technologies and exponential growth of their number.		It is underlying other risks, and make them much more probable and less manageable, as too many risks appear simultaneously		Bigger bombs	Moore law in biotech	Cheap nukes	Moore law in computers, many new AIs	Many new unprecedented physical experiments
3. Agents	1. UN, or a future **world government**.	It most likely will make dangerous mistakes, but the main problem that we don't have such central authority for risks prevention.		It would probably be blamed for inaction and weakness.	Base it on AI and FAI principles					
	2.**Superpowers**	Superpowers are always in a fight for world domination, so they create new weapons, doomsday devices, extremely large nuclear arsenals, arms races and wars, as well as acts of terrorism, and diversions. There is always a fight for domination between existing superpowers, currently there are three: US, China and Russia. The superpower which is second in the race is the more dangerous as it may prefer risky behavior.		The fight of superpowers posed the most serious global risk during the Cold War, and the risk probably remains high, but will be replaced by risks of smaller agents in the future.	Demilitarisation, prevent cold war	Doomsday bombs, Tsar bombs	Clandestine bioweapon program, like "Biopreparat"	Very large nuclear arsenal, including 10 thousands bombs	AI in control of nuclear forces (Perimeter) AI in control of civilians Secret government AI	Most investment in new physics
	3.**Rogue-contries**	They have a crazy ideology and they are always afraid that a superpower will take them by force. North Korea, Iran, ISIS North Korea probably is most dangerous		The main risk arises from attempts to disarm a rogue country, which is also prone to unpredictable military actions outside which may result in a war.	Do not treaten them. Take over in early stages of their development or integrate them into society		Risky bioweapons program with lower margin of safety (Iraq) Bioterrorism	• Defensive nuclear weapons • Possible source of nuclear terrorism • Incentive to create doomsday bombs for deterrence or blackmail.	Not much chance to be first in AI race	
	4.**Pair of nuclear countries** in deadly existential conflict and local arm race	India-Pakistan North and South Korea China and Taiwan Russia and Ukraine Israel and Iran		Accidental war could happen	Hotlines, peacekeapers			Risks of regional nuclear war		
	5.**Corporations**	Large companies fighting for domination in their field.		Corporations are interested in profit, so the main risk is that they will be too risky in their field. It looks like AI will be created by a large corporation like Google.	Regulation, enlightment				Could go all in for strong AI for profit. The riskiest will win (Google, IBM, Facebook)	
	6.**Expert community** in field of x-risks	The community could make collective mistakes, presenting an incorrect picture of x-risks, of incorrect methods of their prevention, or an inability to present the problem in coherent way to their government. It could even give dangerous recommendations. They could imagine new ways of ending the world and inadvertently give this idea to terrorists or inspire them.		The current community can create a picture of risks, but the risks prevention plan is lagging, and the promotion of risk prevention agendas to governments is also weak. It is disproportionately Anglo-American, but most risks originate from other countries.	Cooperation					
	7. **Secret military** and intelligence programs	Some countries could have rogue spy agencies or crazy clandestine military programs, like former nuclear programs in Switzerland and Sweden		Seems to be small	Transperency					
4. Sects - that is small group of people with idea	1.A group which think about X-risks prevention by **smaller catastrophe** (such group doesn't exist but such opinion was voiced)	Willard Wells said that one of the best options to prevent extinction is a smaller catastrophe which stops tech progress.		Not high in foreseeable future	Standard antiterroristic measures, including penetration and interent control for early detection of intentions.					
	2. Anti-technology **Luddites** - Kachinsky	Diversion to stop technologies. Or terrorism to attract attention.		Will grow as capabilities of small organisations will grow	+ theoretical prove that they are wrong, meme war.		Artificial pandemics	Provoking nuclear war		
	3. **Ecoterrorists**: antihuman, for Gaya - ELF, Voluntary human extinction movement, PETA	They value animal life more when humans are causing large-scale animal suffering and new mass extinction, and they may want to stop it. (But they don't understand that animal life will die off anyway because of growing sun luminosity and humanity is its only chance of long-term survival)		Will grow as capabilities of small organisations will grow	+ show more respect to animals rights		Artificial pandemics	Probably will not use such weapons, as animal may suffer		
	4. **Blackmail** . Any group which uses x-risks for global blackmail via creating a Doomsday bomb or virus, for any other goal.	Doesn't exist now.		Will grow as capabilities of small organisations will grow	+ declare that you will not succumb to blackmail under any conditions		Artificial pandemics	Cobalt bombs		
	5. **Domination**. Any group which attempts global domination by killing or enslaving all other people, but goes too far—or just attempts to raise its own status by destroying larger agents and creating revolutionary situation.	Doesn't exist now.		Will grow as capabilities of small organisations will grow	Survilience		Artificial pandemics			
	6.**Doomsday cult** which tries to fulfill an "End of the world" prophecy (ISIS, Aum Shinrikyo)	ISIS wants Armageddon battle. Most dangerous agent of doom now.		Will grow as capabilities of small organisations will grow	+Imprisoment of the members		Artificial pandemics	Cobalt bombs		
	*7.***Suicidal cult** - *everybody will be happy in a paradise after death*	Some UFO cults, Johnsontown massacre		Will grow as capabilities of small organisations will grow			Artificial pandemics			
	*8.***Biohackers** *groups, trying to kill each other, or participating in some kind of fancy championship of creating deadly viruses, or a simple like of self-replicating things*	In the 1980s the first computer viruses appeared. In the naughties millions of viruses appeared every year. If we have multiple biohacker groups, they will try to create all possible viruses with all possible goals.		Will grow exponentially	Regulation and survilience		Artificial pandemics			
	9. **AI-hackers***: interested in creating personal superAI extensions, or AI-powered computer viruses.* *AI-doomsday cults: want replace humanity with nonhuman superintelligence* *AI-hackers could access existing AI system and change their behaviour for self profit or just expression of personal mighty (self-driving cars, weapon-drones, planes or even AGI)*	A lot of computer hackers exists and they will probably will be interested in new AI technologies. Some people claime that their goal is creation of superhuman AI after which they could retire.		This will grow with the availability of building blocks and the failing price of AI-capable computers.	Regulation and survilience				AI ransom ware	
	*10.***Poetic of apocalypse***: gamers or survivalists, who want to live in post-apocalyptic world. A lot of people like post-apocalyptic games and moves, and some may try to create such an environment, and it is part of desire to end the world*	A lot of people likes postapocaliptic game and moves, and some may try to create it, and it is part of desire to end of the world		Will grow as capabilities of small organizations will grow			Artificial pandemics	Starting nuclear war		
	11. **Desire for the end of the world** - unconsciousness drive to end the world as we know it, in the past presented in carnivals, apocalypses, and now in catastrophic movies. It is not a conscious agent, but could manifest itself as subconscious drive.	People have a built-in desire for revolution, as it provides an opportunity for young males to gain a higher status in the tribe. Any catastrophe may also kill off the elderly.		Will grow as capabilities of small organisations will grow	+Psyciatric care		Artificial pandemics	Starting nuclear war		
	12. **Risking** the end of the world for personal benefit (dangerous experiment, risky life extension practice).	If some new genetic treatment has an 80% chance of giving me immortality, but also a 1% chance of going out of control and killing all humanity, than most selfish agents would prefer to create it. But if many such treatments are created, catastrophe is inevitable.		Will grow as capabilities of small organisations will grow	+Promote cooperation prisoneer dilemma		Artificial pandemics		Unfriendly AI	
	13. **Genocide**. Groups of people, who see solutions of their problems in complete extermination of other group of people	If two several groups of people would practice mutual genocide		May arise if animosity and barbarism in the world will grow, and nationalistic values rise.	+Promote value of human life		Artificial pandemics, and genetically targeted viruses			
5. Individuals	*1***. Revenge.** Revenge-driven mad individual, creating a long term plan of omnicide There are several examples of people executing long term plans of revenge on their community and building sophisticated weapons (e.g. armored bulldozer).	There is several examples how people executed long term plans of revenge to their community and built sophisticated weapons (e.g. armored bulldozer)		Will grow as capabilities of small organisations will grow			Artificial pandemics			
	2. **Operator** of dangerous technology, under the influence of strong effect (missiles launch operator, worker in a biolab with dangerous viruses). Or just the owner of dangerous weapon.	African killing bees were realized in South America by their owner, perhaps as a result of his mood change. School shouting as an example of a desire to kill everybody.		Will grow as capabilities of small organisations will grow	Mutual control, failsafe systems		Artificial pandemics	Starting nuclear war		
	3. **Schizophrenics** with complex ideas. Idiosyncratic actors in P.Torres classification	Could be complex and unpredictable. Examples: School shootings Gerostrat behaviour.		Will grow as capabilities of small organisations will grow	Cure them, tests		Artificial pandemics			
	4. **President-dictator** of a nuclear-capable country could act under emotion, madness or make a mistake (Saddam).			Grow with the number of nuclear countries	Sanity check for presidents			Starting nuclear war		
	5. **Arrogant scientists,** willing to continue dangerous research and motivated by self-confidence and vanity want to proceed with dangerous experiments and dismiss all concerns. It could also be collective interest of large group of people, like CERN	They will benefit personally if their experiment continues: they will receive money, power and become famous.			Ethical committees		Flu pandemics New flu strains creation Recreation of old flu strains		SETI-attack METI risks	LHC Stevenson probe
	6. **Almighty wannabe.** Individuals who want to reach personal almighty, and are willing for it to go on high risk, and also kill large group of people as rivals The subfiled "**Mad scientist**" combines this and arrogant scientist with also schizophrenics.	They will probably use biomodification, genetic therapy, nanotech and attempts to create AI. They will go in even if chance of success is like 1 per cent, and human extinction is 99 as they still win infinite utility.			Parential control		Pandemics		Unfriendly AI	
	7. **Herostratus**. People, who want to be famous in their group by creating the largest possible things or the most visible achievements. **Hackers competitions.**	Herostratus was driven by a desire to be famous. There could be competition between biohackers who first change biosphere by their replicators. It will kill biodeversity.			Universal sanity checks		Artificial pandemics. Relieasing replicators into the wild			
	8. **Sadists**. People who get pleasure by creating suffering to others. + Psycopats	Some sadists may be interested in creating large scale suffering.			Cure them, Control based on a criminal record		Artificial pandemics			
	9. **Children**. Teenagers could achive large result in some field of knowledge, but their vaue system is inmature, risky and emotionally driven.			On current level of proliferation of knowledge of mass ditraction it is not risky.	Parential control					
6. Mistakes (agent is not aware that he is an agent of the catastrophe)	1.**Prediction mistakes**, biases			Our knowledge is growing, but the price of a mistake is also growing	Independet assesments	Impacts				
	2.Managements and **policy mistakes**			Our knowledge is growing, but the price of a mistake is also growing	Many level democratic control			Starting nuclear war	Unfriendly AI	
	3.**Handling mistakes** - spill of a virus in a lab			Grow with growing number of agents	Fail safe labs		pandemics, grey goo	Starting nuclear war	Unfriendly AI	
7.Posthumans	A **war** between some kind of posthumans and humans is imaginable.			Zero before any posthuman appear, but high after	Find a way to smooth transition to posthumanity		pandemics			
8.AI	**Strong AI** on different stages of its development could exterminate humanity (see AI failures map)			High, as no clear FAI theory exist	Safe AI theory					
9.Aliens or other form of non-human intelligence	**SETI-attack**, see map "risks from aliens"			Growing as our SETI capabilities are growing	Postpond development of SETI until our AI is created					
10. Owners of simulation (if we are in it)	**Simulation shutdown** by its owners in unknown moment.			Risks are growing as the increasing complexity of our world requires larger calculation power for its simulation.						
Types of agent behavior	**1. Error**	An agent who doesn't want a catastrophe, but accidentally creates it.		Error is more likely than terror	Cognitive biases prevention, many level check					
	2. Terror	An agent who wants to destroy civilization AND kill himself.			Negative emotion cure, depression cure, dangerous memes erradication					
	3. Overspilling - action, which occurs on a larger scale than expected (terror+ error)	An agent who wants to kill only a group of people but creates an unexpectedly large catastrophe. It easily could happen if self-replicating devices are used. Summing up of many agents (arms race, tragedy of the commons, many small actions adding up) - many errors and terrors.		This is most likely situation						
	4.Summing up of many agents (arms race, tragedy of the commons, many small action adding up) - many errors and terrors	Multiple agents create local terrors or errors, which combined result in global catastrophe.		Also probable						
	4. Risk - taking: Risk-taking: risks which benefit personally but which raise global risk.	Agent gains large personal benefit while creating a small global risk.				• Dangerous genetic therapy by self replicating viruses for personal longevity • Arrogant scientists, who recreated extinct flu viruses				Dangerous physical experiments, dagerous SETI and METI programs (Zaitsev)
	5. Multilevel stacking of agents, such as: superpower - president - missile operator	Coordination of a group of agents								
	6. Blackmail by Doomsday weapon - quazi-rational behavior, which looks like win/win situation, but multiple contradictory blackmails could result in stalemate, in which use of doomsday devices is inevitable.	An "arms race" of stronger and more dangerous doomsday devices is also possible, for example a device which kills 90% of the population could be overpowered (i.e., have more blackmail potential) by 99% device.							The map is created by Alexey Turchin, 2016	