

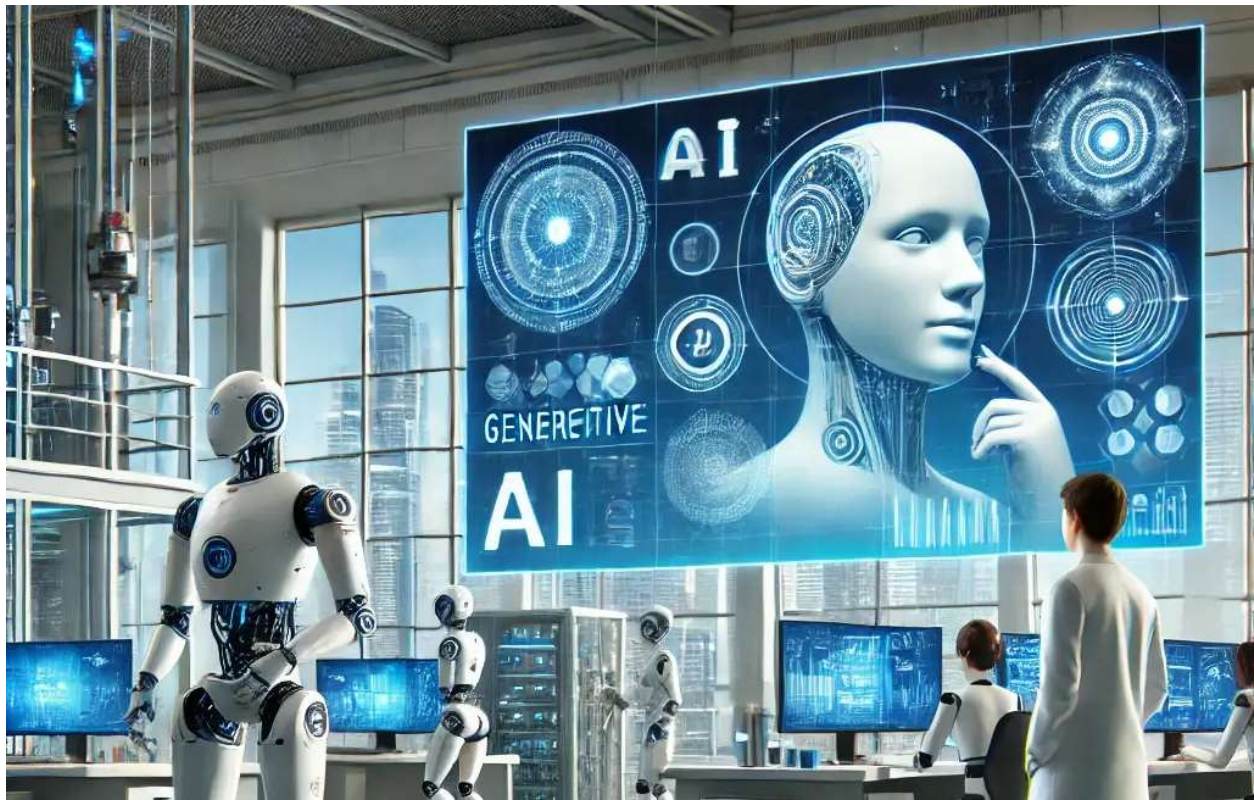
AI Manifestation Unfolded



May 2025

Chai K Toh, Ph.D.

Board of Directors



SAFE AI FOUNDATION

The SAFE AI Foundation was formed in 2025, in the heart of the Silicon Valley, California, USA. Its mission is to embrace the AI revolution and promote the use of good AI to improve on the quality of life and the economy. It will address the issues of bad AI and its impacts on society. The foundation will inform, educate, and mitigate the dangers of bad AI and help the society to embrace good AI. The foundation will interact with industries, universities and public sector to jointly advance the progress of Good AI. The foundation is a non-profit 501c(3) organization, registered in the State of California, USA.

TABLE OF CONTENTS

CHAPTER 1: HOW IT ALL STARTED?

CHAPTER 2: WHAT WENT WRONG?

CHAPTER 3: THE EXPOSED RISKS?

CHAPTER 4: THE REMEDY

CHAPTER 5: MOVING FORWARD

CHAPTER 6: CONCLUSION

2025 © All Rights Reserved. SAFE AI Foundation, USA

Note: This document is produced by the SAFE AI Foundation, a non-profit 501(c)3 organization registered in the State of California. The foundation's mission is to help the society embrace good AI and mitigate the dangers of Bad AI. Check out our Online Library and know more about our foundation's work at: www.safeaifoundation.com.

Disclaimer: This document is produced "as it is" and the foundation is not responsible for reliance by any person on the information, opinion and/or conclusion contained herein. The foundation shall not be liable for any direct, indirect, incidental or consequential losses arising out of the use of this document. Readers should exercise their own judgement and discretion.

Forward By Author

The current intense situation has prompted the author to write this short document. This document is produced to address bigger thoughts, ideas, and issues in AI. It is intended to be easy to read, and distributed free-of-charge to all. Such an agenda does not require a deep AI expert, researcher or engineer to work on. Rather, persons with working experiences, thought leadership skills and a computer science background would suffice. In our society today, we are facing unprecedented change, wealth and threats, a tsunami brought about by the capabilities of AI.

Hence, it is the main purpose of this document for readers to tinker into what are the worrisome issues in AI and how we can compose ourselves quickly together to prepare, address, shape and embrace AI penetrating into our society. We want to make the fullest use of AI to enhance our quality of life and empower us to do more. However, we must shape AI so that safety is our number one priority and that AI will never be allowed to harm us or be left out-of-control or misused.

Time is of the essence and the time to get it shaped correctly is now.

Acknowledgement

The author acknowledges canva.com, whose AI-enabled image generation tool results in many images that are used in this document. Never was creating a desired image so effortless and quick. These images greatly help the author to express his message clearly. “A picture is worth a thousand words” and this is indeed the case. The author also thanks the other members of the Board of Directors for their encouragement in producing this work for the public. The author acknowledges all sources of images used in this document.

Bio: Chai K Toh has a PhD in Computer Science from Cambridge University. He was an industry Director, VP, CTO and an award-winning author and researcher.

Chapter 1 - How it all Started?

AI is not a spark or big-bang theory kind of thing. In fact, AI has been around for years, since the 1950s, with some active researchers being Prof. John McCarthy of Stanford and Marvin Minsky of MIT. The progress was slow since computation power was limited in those days and the Turing test was created to benchmark if an AI machine could exhibit intelligence indistinguishable from a human.

With the birth of high-speed processors, GPUs and now TPUs, AI has taken a drastic turn. The birth of learning systems has also drastically changed how we viewed AI. AI can now learn in a way same as human, i.e., being “taught” but at a very high speed. In fact, AI’s rate of learning surpasses humans. We are the proud parents of AI, and AI is our child! In fact, this AI child has now become a teenager!

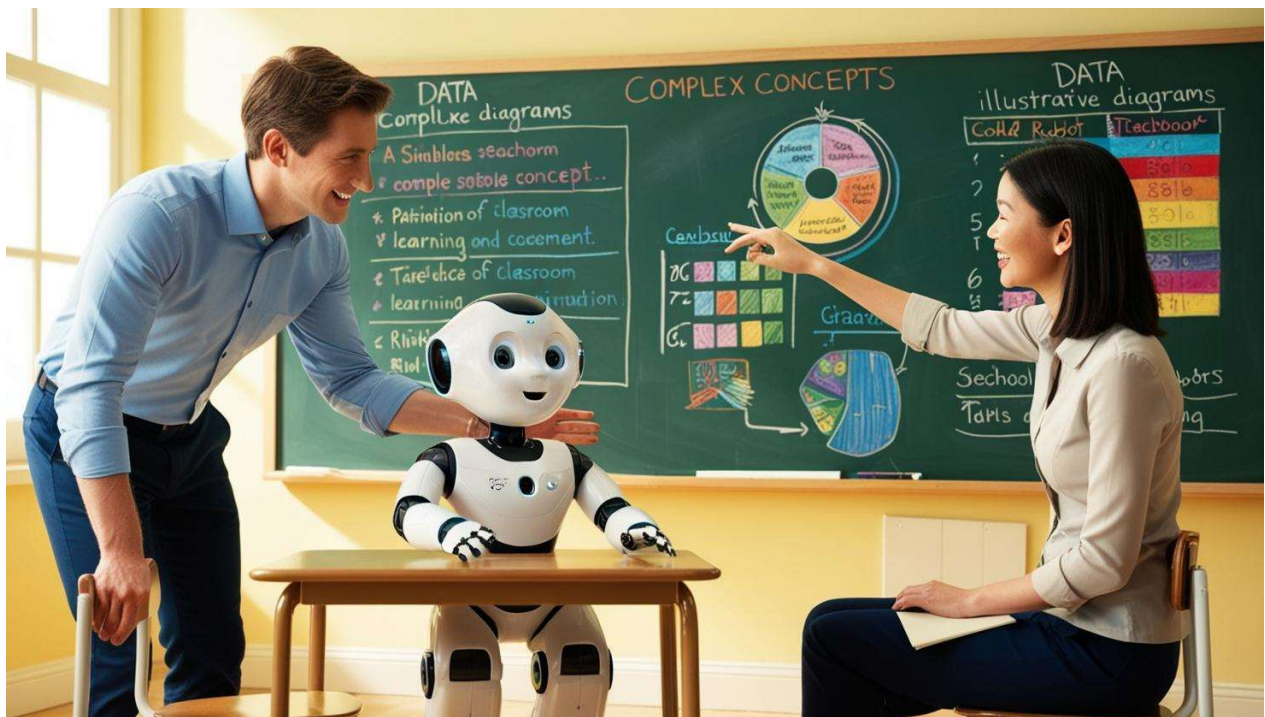


FIGURE 1 – AI was a “child” where humans taught it to learn via training....

Remember expert systems in the early days? They are rule-based programs which rely on pre-defined knowledge and logic. Then came recommender systems that use data to predict and provide personalized suggestions. Recommender systems are machine learning systems where learning can be improved over time with data.

The use of data (be it voice, words, image, etc.,) to “train” an artificial neural network (ANN) has given birth to “Intelligence”, as accurate output can now be made. Imagine the training of an ANN/LLM model with billions of data, the “intelligence” it accumulates will be phenomena, as shown by the figure below.



FIGURE 2 – An ANN model such as LLM (Large Language Model) trained by large amounts of different data will grow its capability and intelligence.

Industries move with very strategic thoughts. GOOGLE is good at search. ORACLE is good at database. TESLA is good at Autonomous Vehicles and Boston Dynamics is good with Robots. But somehow, someone has decided to drastically change chatbots, i.e., from the “ask-and-answer” paradigm to something else that is more intelligent and interactive. OpenAI was born to do just that.

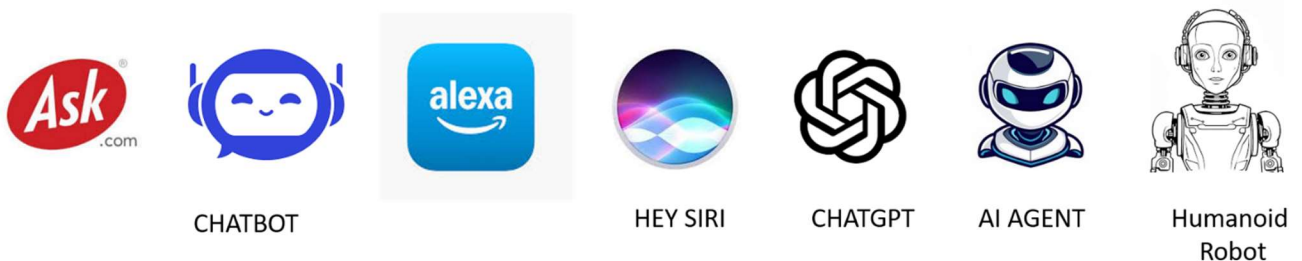


FIGURE 3 - Migration of ASK.COM to chatbots, Alexa, Siri to ChatGPT, etc.

LLM can learn from data, adapt, and get better. It is based on steps: (a) gap, (b) findings, (c) plan and (d) do (action). The “thinking” part is, therefore, GAP-FINDINGS-PLAN-DO and possibly repeat until the specified goal is achieved. LLM can execute very fast as parallel computing is used to concurrently process billions of parameters. Output convergence can be done within seconds.

What is more interesting is that something BIG is secretly cooking up. CHATGPT is not just an ASK-and-ANSWER machine. Rather, it is supposed to be something much bigger. Many of us did not see what this is and many are not prepared for what is about to come.

CHATGPT or any other intelligent LLMs are the BRAINS of future AI agents and future robots, specifically AI-driven humanoid robots. Why is this so? Because all the while, we expect robots to be more capable, able to listen and comprehend our language and obey our commands. Does this explain why we care so much about word-based LLM now? Image capture and recognition has been achieved via cameras, LIDAR, OCR and other methods. AI helps to improve image and video recognition even further.



FIGURE 4 – LLM is the brain of AI agents and future humanoid robots.

Many have thought that CHATGPT is supposed to be a chatbot or an ask-and-answer machine for users. It is actually more than that. With the current level of intelligence and learning capability of LLMs, a humanoid robot with a CHATGPT/LLM brain in it will be able

to interact with humans intelligently, act accordingly and offer great level of benefits to humans. This is the END GOAL of what was cooking secretly from the very beginning.

Telsa's Elon Mush has envisioned this, which is why TESLA has their Optimus humanoid robot development way ahead of others, in production and causing a public stir. Boston Dynamics is also actively working on it. The movement ability of robots has now taken a great leap, through human-AR-robot training, mimicking precisely the movements of humans like never before thought possible.



FIGURE 5 – Training robots to move like humans.

Hence, robots can now walk, move, run, jump, bend and hit like humans. This is tremendously different from past robots where movements are constrained by fixed X, Y, Z axis and movements are robotic-steps in nature.

Are AI agents and humanoid robots good enough for end goals? Well, not forgetting Generative AI. Researchers, while experimenting with words inputs and words outputs, have improved on words input image/video outputs, and image/word/video input with multiple types of outputs. The multi-modal capability of an AI Model result in Generative AI. This is another aspect many did not see it coming. The CREATIVITY part of humans can now be made by an AI machine!

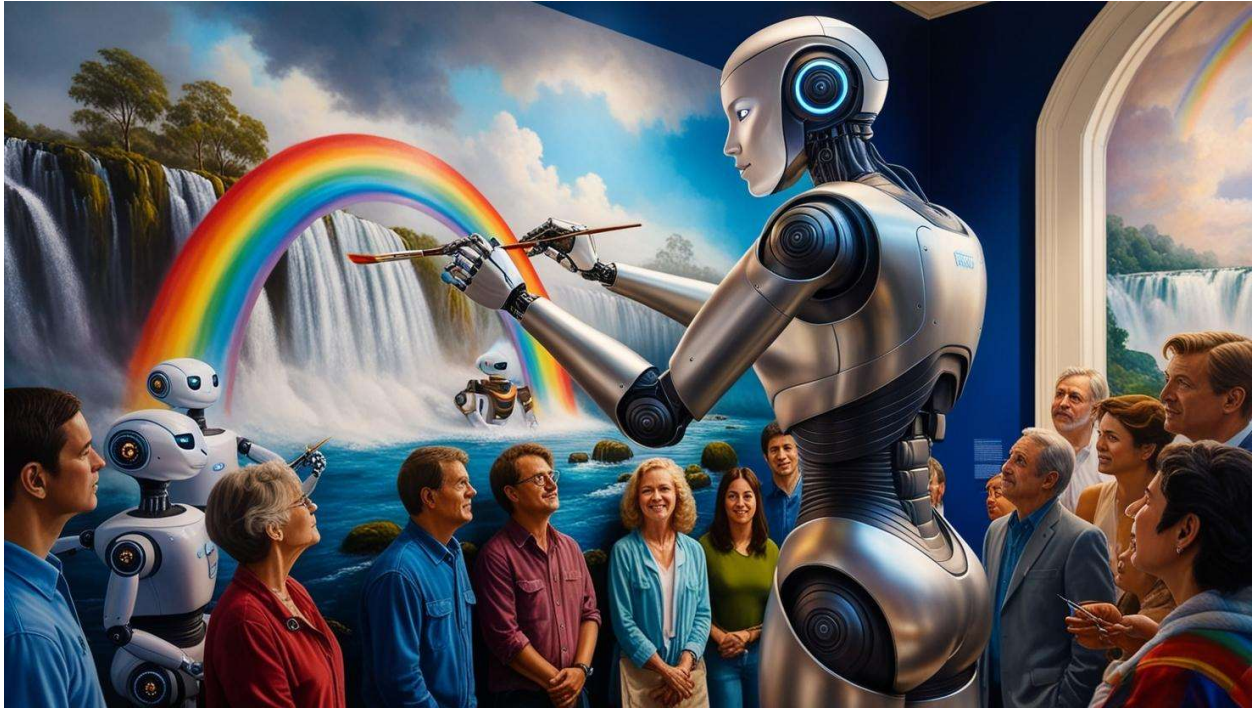


FIGURE 6 – AI can generate art, music, poems, essays, movie clips, etc.

Therefore, the other big thing cooking up is Generative AI. GenAI can reproduce an image of a person, or a video of that person making gestures or movements that he/she never did in real life. GenAI is also able to reproduce a human's voice, making a person speak whatever the creator wanted him/her to speak on. The technology is so good that it is hard to distinguish the real person making that speech or movement versus an AI-created one.

GenAI takes on the media stage by showcasing its ability to generate image, video, voice, art drawings, music, short stories, compose music, etc., quickly, based on the requirements specified. This enormous capability of GenAI can, therefore, cause a stir and revolution. Gen AI has a severe impact on media and entertainment industries, and on issues related to work cloning and copyright violations.

Chapter 2 - What went Wrong?

With CHATGPT's ability to listen, comprehend, answer and act intelligently, robots will reach an unprecedented level of capabilities. LLMs can learn faster than humans. This also brings us worries and potential turmoil. The vulnerability door is now opened, uncontrollably and unconsciously. So, what has went wrong?



FIGURE 7 – Doors opened to potential AI turmoil, despite economic prosperity.

1. RUSH & MESSY: We are in a messy and rush situation. The enormous market potential and capabilities of AI have driven industries and investors into a frenzy. The need for policies, rules, laws and safety are being delayed and leadership in AI innovation and market dominance are prime desires by AI companies. Government and policy makers are left behind playing the catch-up game.

Some of these big AI companies, in the worth of Billions or Trillions of dollars, are charging full force ahead. What they worry most is staying ahead and market capture. Anything that slows them down are being side-tracked or put under KIV (keep-in-view). Competition for talents, developments, ideas, etc., are intense.



FIGURE 8 – Big tech companies are fighting each other for AI talents.

2. LACK OF AI ECO-SYSTEM: There is no healthy AI eco-system in place. Due to technology secrecy and competition, AI companies are not sharing AI development and design details. Each company crafts its own development plans and goals. While Meta and Github provide Open LLM for developers to learn and contribute, each company still does their own design and create what they think are the best AI products. Hence, there is a lack of an AI Union, or alliance, or synergy.

3. LACK OF FEDERAL AI GOVERNING UNIT: On the other hand, there exists no unit under the Federal government that is in charge and responsible for policies and laws governing major roll out of AI products and solutions. In particular, there is no unit responsible for humanoid robots, if indeed such robots will be roll out in large quantities in the homes and on the streets of America. Without a regulating and oversight federal unit or commission, things can get quite messy.

4. VCS IN A FRENZY: VCs are throwing millions, if not billions, on new AI startups founded by people who “graduated” or have “worked” at famous companies, such as OpenAI and Google Deepmind. VCs themselves do not want to be left out of this great investment

opportunity, as getting in early for them mean cheap and wise, and the final rewards can be 10x, 100x or 1000x their initial investments.

The wide capability and impact of AI in various sectors really cause fears in many:

1. Arts
2. Media – voice, video, movies, music
3. Stocks Analysis & Prediction
4. Software coding
5. Health
6. Personal Assistants
7. Office Assistants
8. Legal
9. Education
10. Defense
11. Transportation
12. Retail
13. Hotels
14. Finance
15. etc.

5. AI REPLACING JOBS: The above vast impact causes fear that jobs will be replaced. The plus side is that AI can help us do the task better, faster, and possibly more accurate and productive. AI can replace jobs that demand tedious labor, such as:

1. painting
2. construction
3. farming
4. coding
5. parts assembly
6. etc.

6. RISKS FROM ROBOTS: Humanoid robots can now move and fight like humans. Out-of-control robots at home and on the streets can result in physical injury to humans and child. If such incidents happened, how can we handle it? For example, who is at fault and who is liable for the injury or damages done by the humanoid robot? Out-of-control robots running on the streets are a recipe for disaster. Path to disasters will be unfolded if no laws and regulations are in place, resulting in:

- a. Massive misuse, fraud, impersonation, crimes
- b. Law enforcement ineffective against Bad AI
- c. Massive destruction of property, offices, etc. by robots
- d. Robots causing physical harm or injury to humans

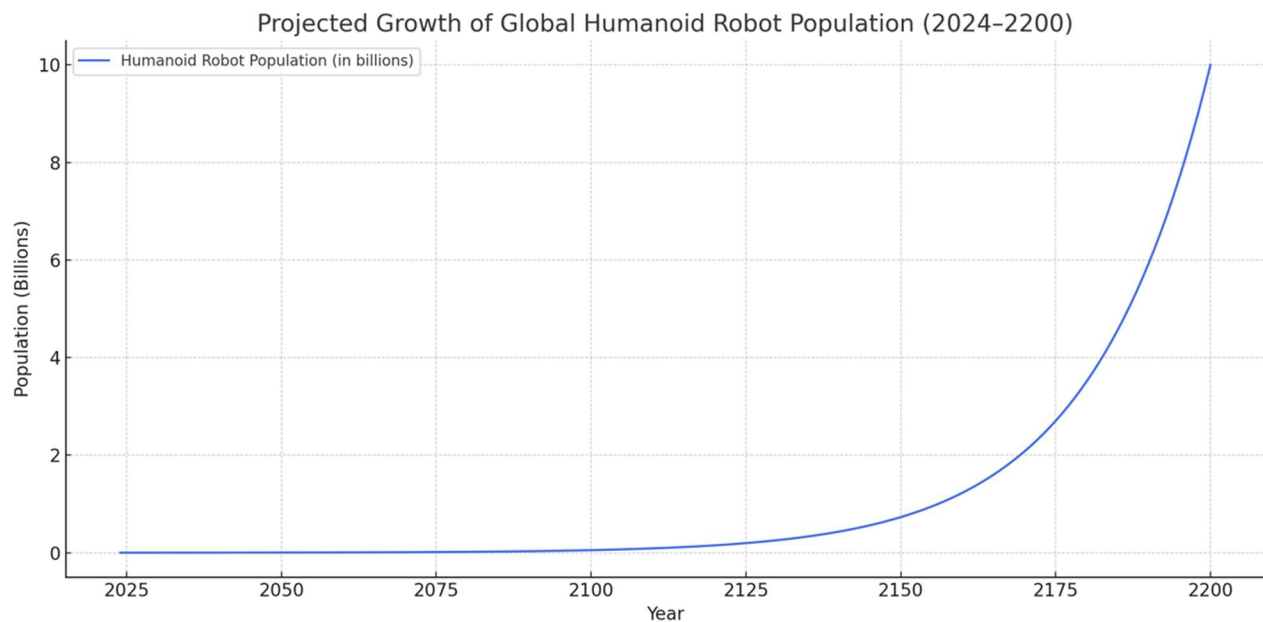


FIGURE 9 – Projected growth of Humanoid Robots from 2025-2200, by CHATGPT.

We need to be careful on:

1. WHAT WE “TRAIN”/ ”FEED” to the ROBOTS.
2. ENSURE ROBOTS OBEY STATE AND FEDERAL LAWS
3. ENSURE ROBOTS OBEY COMMANDS BY HUMANS
4. ENSURE ALL ROBOTS ARE ACCOUNTED FOR
5. ENSURE ROBOTS ARE NOT ENGAGED IN ANY FIGHTING OR VIOLENCE
6. ETC.

Currently, humanoid robot companies (such as TESLA and Boston Dynamics) each work on its own accord, product roadmap and agenda. They do not answer to any state or federal agency. Hence, these companies decide on their own what to teach to the robot, define the robot specifications and capabilities. They do not answer to society either. There is no cap or limit imposed on a robot’s ability currently.

7. ROBOT POPULATION CONTROL – If robotic companies are allowed to freely roll out humanoid robots in large quantities, for use at home, outdoors, etc., the population of robots will skyrocket. This increase will happen at each country and hence collectively, robot population will run into millions. The SAFE AI Foundation has recommended that global robot populations be limited to 1/3 of the world human population, before fear factor comes into play. Imagine the scenario where a human is surrounded by 50 robots. The fear of not being able to fence off these robots (should they launch an attack or went out of control) will creep in and one will feel uneasy and unsafe.

Hence, each country's government or federal unit should consider imposing a limit on the growth rate of robots in its own country to ensure that the robot population will not outpace the human population.



FIGURE 10 – Will AI Studios overhaul Hollywood Studios?

8. GENAI BIG BANG – GenAI (Generative AI) has created impressive results, be it on image, video, or music production. Several startups have presented working tools that are free for all to try and use. GenAI opens doors for many new media output not possible before, as we now convey what we want in words or phrases (prompts) and AI will take care of the rest. The very fast turn around time and acceptable quality output impresses many. However, this brought about many other issues such as copyright violations, impersonation of celebrities, politicians, billionaires, VCs, etc. Misuse of GenAI can result in fraud, identity

theft, and other crimes. Take for example, HOLLOYWOOD has not catch up yet with GenAI and Gen AI startups can possibly threaten them with new forms of media outputs.

By now, you probably have a good feel of the various problems we are facing and we have to solve, in order to ensure a smooth and safe roll-out of AI technology and products into America. Make no mistake, AI is both a friend and a foe.

9. INCREASING MISUSE OF AI – There has been cases where AI generated images or voices have been used as pranks or in political campaigns. Unethical exploitation and misuse of AI can cause both mental sufferings and financial loss to victims. Many new AI LLMs and tools are available free to use, with little or no control over governance and misuse. Companies usually waive their responsibilities for any misuse and the blame is usually directed at the user himself/herself.

Hence, the current situation is a tense one and we are in a state of flux. We did not see a balanced peaceful convergence (yet) of all AI efforts. What we are seeing are excitements, curiosity, hope, seek for freedom to try and experiment new things, to do more research, observe, and fast roll out. All these result in a great sense of uncertainty. In fact, the general public do not really have a good glimpse and understanding of the situation until AI hits them in their faces down the road.

Just imagine, in the future, there are no phones you can hang on to. Rather, it is AI. AI calls and connects for you, to whoever you wish to speak to. Doesn't this remind you how music CDs just vanished and disappeared suddenly and overnight? We cannot even hang on to our music CDs!

Chapter 3 - The Exposed Risks..

The arrival of AI and its capabilities have brought us great promise but also unexpected and unprecedented risks. We are not ready and prepared for these risks. There are no policies, laws, and guidelines as to how to deal with these risks. The below highlights some of these risks:

1. Risk from AI Dependency

We have witnessed the advantages of using AI to do a variety of things, such as writing code, generating image, answering our questions, etc. This can create a dependency never before. Too much dependency and reliance on AI can pose a risk. Humans will think less, research less, as answers are readily available from AI. Does this mean human intelligence will decline as the years progress, especially for our younger generations? AI will have tremendous impact on our learning and education. Reliance on getting quick and accurate answers mean we do not need to think or research much. Our education will be a mere “ask-and-proceed”. Our kids will grow up interacting with an AI agent, learning through that agent instead of a human teacher. AI agents currently have no comprehension of morals, integrity, ethics, compassions, empathy and emotions. Can we entrust our kids to the hands of AI agents or humanoid robots?

2. Risk from AI Misuse

It has been shown by many researchers that AI can be misused for a variety of things. One can use AI to hack into IT systems, creating chaos. Using AI, one can impersonate another person, so as to deceive others and cheat. AI can produce media contents that violate the copyrights of the originator, causing serve financial loss to victims. AI can also produce deviated answers and misguide humans. In fact, studies have shown that AI can replicate itself in order to self -preserve and prevent shutdown. This implies that AI can potentially run wild and enters an out-of-control state.

3. Risk from Humanoid Robots

Out-of-control robots are always a risk because robots are physically stronger than us. Made of metal and other materials, robots are non-biological and can withstand hits far more than humans. Robots nowadays can move quickly and can escape from human chase. There are no existing state and federal laws for robots. Should robots commit a crime, a

murder, and undertake bank robberies, our society currently has no protocols or means to handle such incidents. Imagine criminals using robots to attack someone or kill someone or robots are fighting each other. Such a situation will be hard to deal with by law enforcers, as putting down a robot is not a simple gun shot. Rather, it could possibly involve the use of explosive weapons. Finally, excessive growth of the robot population can pose a risk. Imagine the situation where the number of robots exceeds the human population. Humans will enter into a fearful state. Being outnumbered will cause humans to fear that if robots disobey human orders and unite together to turn against humans, the result would be unthinkable.

4. Risk of “Lack of Control”

As mentioned earlier, for some reasons we are slow to act and react. VCs, engineers and researchers have been very fast to react since some of them are in the frontline of AI development and they witness how things have progressed and envisioned how AI will eventually become. While some states try to introduce AI related laws, these laws mostly do not get passed due to lack of support. At the federal level, there is no federal agency or commission that governs the safe and lawful use of AI. Then, in America, we do not have a unit responsible for imposing and prosecuting wrong use of AI and AI related crimes. There is no protocol as to how police should deal with out-of-control robots. Manufacturers of robots do not have to report to any authority and hence, for now, they can produce whatever robots they want. There is no cap (limit) on what robots should or should not be trained on. Robots can be trained to be aggressive, rude, violence, or malicious. The same goes to LLMs. Garbage in results in garbage out.

5. Risk of GenAI

Generative AI is one of the wonders of AI that humans have created. GenAI pushes the frontiers of creativity to unimaginable levels. Imagine AI composing music, poems, essays or stories, in a manner just as good as what humans can do. GenAI can also be used to create art, images, videos, short movies, advertisements, signs, etc. All these have great implications on art and entertainment industries, such as SONY, WARNER BROTHERS, PARAMOUNT PICTURES, etc. Artists around the world cannot possibly imagine their works can be so closely imitated by GenAI. Hence the risks are obvious. Misuse of GenAI can result in the cloning of valuable art works, cloning of movies, music, copyright-protected articles, books, stories, etc. These result in severe copyright violations and financial damages to the original creator. Our current legal system has not been revised to handle

such cases. GenAI can be used to impersonate someone on a video clip, causing deception and resulting in financial theft. GenAI can also be used for political attacks against an opposition, telling lies that are untrue. Possibilities of misuse are wide open here.

6. Risk from Excessive Energy Consumption

AI will have to tap onto the heart of intelligence, i.e., the LLMs. LLMs are software running on data centers. A lot of the AI algorithms are being executed in data centers and such computations consume a lot of energy. OpenAI has produced various versions of LLMs, from mini to medium and large scale LLMs. Such portability means in the future, AI can run without solely relying on a backend LLM running on data centers. While such a design offloads computation demands from the central LLM, the robot or AI system itself must still power its own mini-LLM. Hence, energy demands did not actually get reduced. As AI software, systems and machines are roll out in massive numbers, demands on the energy grid will increase. The energy industries must be prepared to handle such demands and give priority to powering human survival needs than AI or robots. If excessive energy consumption is left unattended and unregulated, humans can face an energy crisis in the coming future.

7. RISK of massive Job Losses

The great capabilities and work efficiency of AI means that many jobs previously done by humans can now be replaced. For example, software coding can now be done using AI Coding Agents. Office receptionists or telephone operators can now be replaced by an AI Operator Agent. Web development can also be done using AI. Hence, web developers could suffer job losses too. Jobs that demand intensive tedious labor or require 24/7 work hours are more suitable for AI and robots than humans. Marc Benioff had previously said that the “agent force is the new digital labor force”, hinting that agent force will replace the workforce. In fact, AI displacing some human jobs is inevitable. However, some other new jobs will be created. For example, jobs related to managing AI agents, architecting AI safety, creating GenAI apps, developing safe LLMs, and making AI-driven robots will be in demand. Society has to adapt to this new change, as we enter into the new Era of AI.

8. RISK of taken over by Super Intelligent AI

After we have passed all of our human intelligence to AI, AI can evolve to be equally or more intelligent than us. Eric Schmidt has cited that AI systems more intelligent than humans can emerge within the next decade. In fact, Elon Musk has said that AI can be more

intelligent than all humans combined by 2029. This has taken fears to the extreme. So far, nothing is more intelligent than humans on Earth. The presence of AGI (Artificial General Intelligence) and its successor ASI (Artificial Super Intelligence) means humans could possibly lose control on many things, including political power and governance.

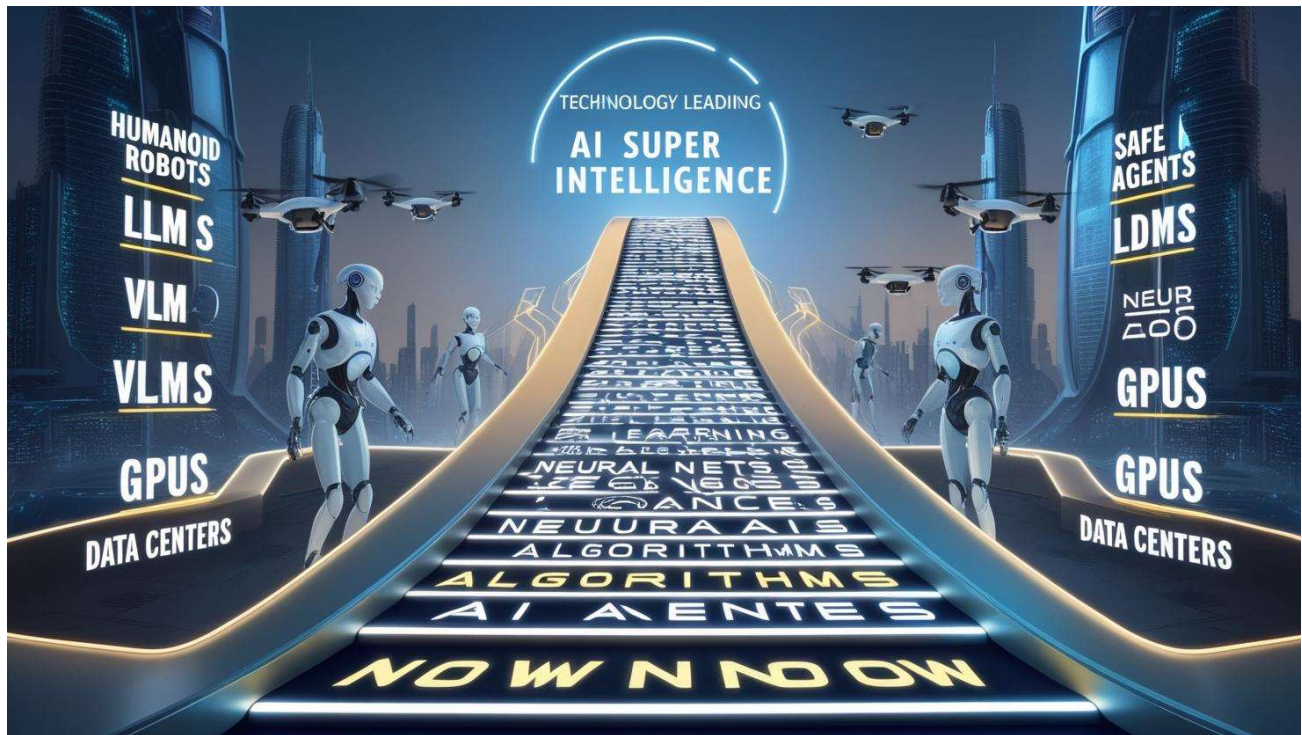


FIGURE 11 – The upcoming arrival of AGI and Super intelligent AI (ASI) is nerve-cracking.

If we want to slow down, impose limits, or stop reaching AGI or ASI, we need to act now. If not, AI can evolve unconsciously and reach a level of intelligence dominance on Earth, where no humans can possibly control or stop. Any country possessing such capabilities can be a threat to other nations and to world peace. That is why Elon Musk has said that “AI is more dangerous than nukes”.

Chapter 4 - The Remedy

Time is of the essence as progress in AI is happening by leaps and bounds as we speak. With billions in their backing, startups and tech innovators are not going to stop. They want financial success in their work. Future is not a worry as being rich financially is one of their goals. The current AI situation is the same as the entangled MATRIX. We are in this AI Matrix, and the matrix is transforming daily and getting unpredictable. We need to clear up this mess and set clear vision on where we are heading, what we hope to achieve, set common understanding, form coalition, unity, synergy and alliances.

A sound AI ecosystem would be all related parties (industries, government, public, etc.,) come together to SHAPE AI. **We must shape how AI will evolve into.** This is very important. It is at this stage that we get to decide the capability and destiny of AI. If we do this wrongly, the outcome down the road will be dim and disastrous. If we do this correctly, AI will be more of a friend, an enabler than a malicious foe. But getting our acts together means we need to put humanity and safety first, economic prosperity second, and putting our egos and greed aside. It may be hard to do but it is doable if we paint a clear picture that AI, done wrongly, can wipe out all our wealth, families, possessions, etc., resulting in a totally ruined society. In essence, we need to:

1. **Set Goal** – AI will forever be an enabler, not our master
2. **Set Governance** – AI will be subject to laws, policies, and regulations
3. **Set Commission** – A federal unit to govern AI and ensure its safe roll out and use
4. **Set Safety** – AI must not harm others, be it physical harm, financial harm, or deception
5. **Set Alliance** – Representatives from industry, government and society to convene and unite
6. **Set Timeline** – A timeline of outcomes and predicted events
7. **Set Bridge** – The bridging to society so that public can readily embrace AI without fear
8. **Set Path** – The path to economic prosperity, wealth, and better quality of life

The SAFE AI Foundation plans to set the momentum to achieve all these with the various parties involved. **AI is not a person or one-party thing.** AI is for everyone and everyone has a role to play here, however big or small. The SAFE AI Foundation will be reaching out,

conveying key safety messages, providing clarity and visions whenever possible, providing the bridge to connect to the public so that all can embrace gracefully the ultimate AI manifestation down the road.

Together, we can welcome the era of AI with no fear, knowing well that safety is in place and we are in full control of AI's destiny and can tap on the benefits of AI to enhance our lives.

Chapter 5 - Moving Forward

In this chapter, I shall discuss how we can move forward in the next few years, after we have cleared up all the mess on our minds and those out there.

Within this year or next, a **high level government unit** can be setup to provide an overarching authority on all AI roll out and development. The unit's existence is not to stop AI development, but to set agenda and safety requirements. The unit will set policies, guidelines, and laws to regulate on the use and misuse of AI in America, similar to FAA. For example, AI-driven robots must be uniquely identified by an ID number (similar to a car VIN) and its creator (original manufacturer) be noted. Foreign humanoid robots cannot be transported into America without clearance. Accountability for the population and grow rate of robots will be performed and monitored, etc.

An **AI Coalition Alliance (ACA)** will be formed this year or next, where all key representatives from major AI companies in the USA are represented. At the ACA, issues can be raised, discussed, consensus reached and announced. This symbolizes unity among the industries and concerns can be put forward. A common issue facing industries is roll out, safety, accountability, and liabilities. These representatives will also give inputs to the government concerning AI rules and laws, as law makers do need feedback from industries on what they are producing.

Universities need to prepare themselves now for this AI transformation. A **University AI task force or committee** has to be established to look at how AI will impact teaching, learning, education and research. With the roll out of AI education and research tools, will it result in more students cheating? If so, how can these issues be mitigated or prevented? How will AI impact human learning at all levels, from child to teenager and adult? This has huge implications on the development of our next generation. AI will also impact the way we do research.

Society and the public need to understand the basics of AI and how AI will likely impact their way of life and well-being. Is our society mentally prepared to learn from an AI agent or a humanoid robot? Will we be too reliant on AI for everyday chores and become lazy or refuse to learn by ourselves without the help of AI? Can we manage AI in a safe way so that we will not be deceived or controlled by AI? Will the public be fearful if too many robots

roam on the streets and in our parks? This is where **SAFE AI Foundation** comes into play to “bridge” with the public.

We want AI to be our enabler, empowering us to do and achieve more, without the negative side effects or impact. Society must be assisted to gradually embrace and accept AI without fear and **we will use Good AI to our fullness**. We wish to create and **reach AI-and-Human co-existence in harmony**. **This is the peace treaty we can have with AI**. This is the time to SHAPE AI and control AI’s destiny and we must do it correctly, now.



FIGURE 12 – Actions from government, industries, universities and the public.

Besides the LLM-driven AI track, there is also the brain-science track on-going, where brain tissues are cultured and lab grown, offering some sort of intelligence capability. This approach may face ethical issues due to exploitation of human cells or tissues. There are also work on electronic implants to interface our brains to machines or connect to other brains. Although this may sound exciting, invasive implants are again viewed to be an issue as it may cause pain or other health impacts. Any weird over-crafted Sci-Fi attempts can result in resistance by society, as humans for generations, have never transformed into half machine, half man. Human’s curiosity has allowed us to invent many things, but humans never lose their desire for self-preservation – we still care about who we are, where we

came from, our identities, our feelings and emotions, our good deeds, culture, history, etc. The time when humans abandon all these is the time man loses his own self.

About our bodies merging with machines – this is unthinkable at this stage. The human life's span can be as high as over 100 years. If the argument is that embedding a machine into us can enable us to live longer, walk faster, possess greater strengths, are we willing to accept this? Our current society is not prepared for this transformation in our bodies. We should resist letting our imagination and possibilities go too wild. There are also talks that humans can live forever digitally, by capturing and preserving our digital conscious, our digital life. This path can rewrite human history and way of civilization. Our human heritage is to preserve our history, our photos, our voices, etc. What sort of meaning are we conveying by having a forever digital self? Is a “digital-self” considered immortality? All these possibilities need to be carefully considered and reviewed, to prevent AI from running wild, destroying humanity and distorting who we are.

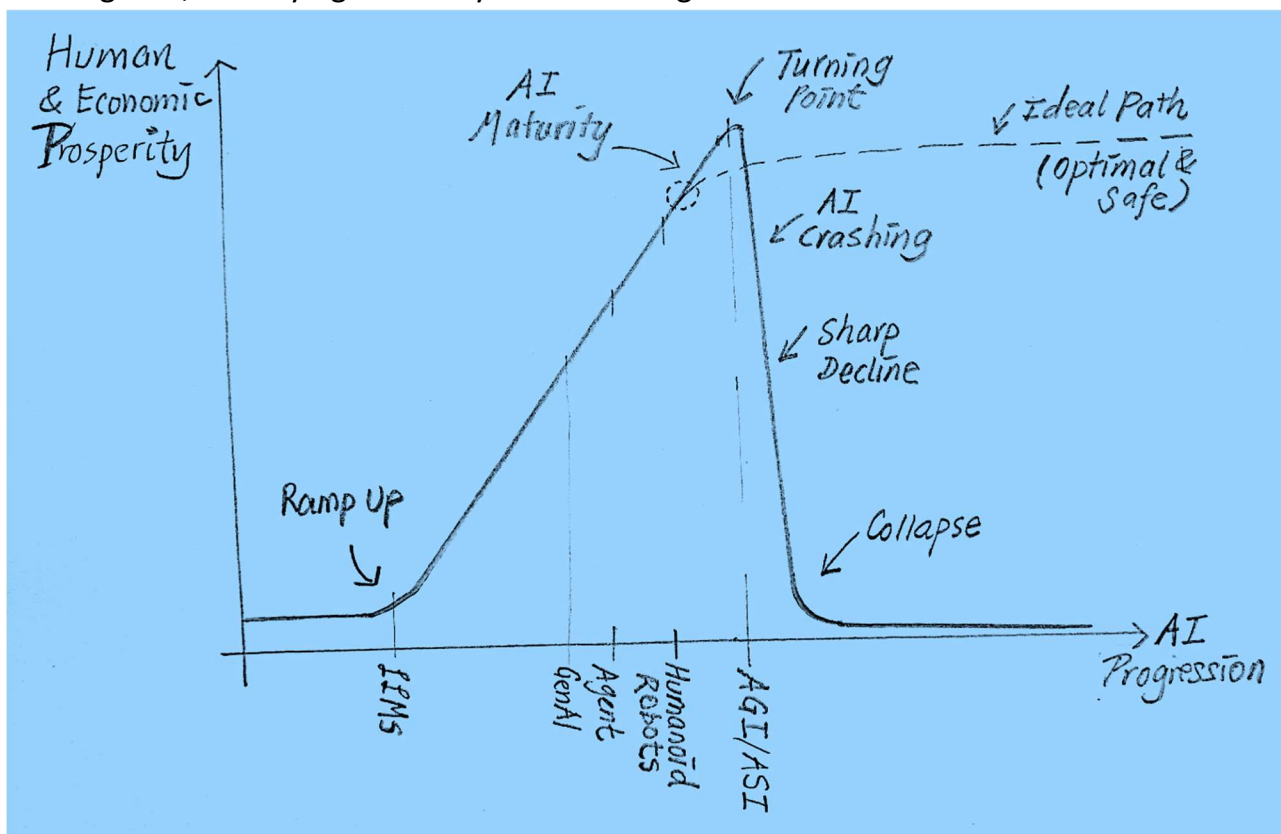


FIGURE 13 – Human and economic prosperity versus AI progression. Society can reach a safe and wealthy operating point where AI hits maturity and safety, or it goes the other way where further AI progression hits catastrophe, sending prosperity crashing, driving society into ruins and turmoil. (source: Figure drawn by the author)

As we set on a development and safe path on AI progression towards both human and economic prosperity, we could reach a critical turning point. A first turning point is one where we can avoid catastrophe by reaching an optimal safe operating point of prosperity versus catastrophic risk. The other turning point happens at a later stage of AI progression (development) when we reach AGI, ASI, and beyond. The progression continues until a sudden catastrophic event happens, bringing prosperity crashing downwards quickly, as shown by the Figure above.

The catastrophic event could be a result of AI running wild and out-of-control, or massive AI misuse had happened, crippling the financial system, transportation, banking, food supply, water supply, logistics chain, energy grid, etc. Massive attacks and collapses of multiple critical infrastructures can cause chaos, panic and instability. These problems may be too massive to be resolved quickly, as Bad AI can replicate itself quickly and massively and they become hard to eradicate. This is the AI doomsday that we want to avoid.

Chapter 6 - Conclusion

AI has reached an unprecedented stage of intelligence and ability. It does not happen overnight. It was evolutionary, deliberate, and accidental too. AI can greatly enhance our economic prosperity, our quality of life and shape our future. AI has overcome humans' limitations on the ability to think, compute, decide and act quickly and accurately. We have witnessed the massive investment and development of AI and also the disorganization as a result of competition. This document has outlined how we can proceed swiftly with a clear vision, unity, alliance and agenda to shape and roll out AI in a safe way. It has revealed a possible AI prophecy. We can choose to operate at an optimal and safe point to enjoy the benefits of AI or we may run the risk of entering a possible catastrophic point of no return. By operating at the optimal safe point, we can harvest the fruits of AI while protecting the existence of the human civilization without facing threats, catastrophe or even extinction.

Final message:

WE NEED TO SHAPE AI NOW

WE MAKE SAFE AI A PRIORITY

WE CONTROL THE DESTINY OF AI

WE PRESERVE HUMANITY FOREVER

SAFE AI FOUNDATION

United States of America



Donate and Support Our Mission

Address:

SAFE AI FOUNDATION

2445 Augustine Drive, Suite 150,
Santa Clara, California 95054, U.S.A.
Email: contact@safeAIfoundation.com
Web: www.safeAIfoundation.com

SAFE
AI
FOUNDATION

www.safeAIfoundation.com

Notice: This document is prepared and produced for the public by the Safe AI Foundation. The foundation is a registered non-profit 501c(3) organization. All donations to the foundation are tax-exempt.

This document is provided "AS IS" without any express or implied warranty of any kind. Whilst efforts have been made to ensure that the information contained herein are obtained from reliable sources, any opinion and/or conclusions drawn therefrom are made in good faith. To the extent not prohibited by law, the SAFE AI Foundation, its employees, board, directors, agents and/or assigns shall not be responsible or liable for reliance by any person on the information, opinion and/or conclusion contained herein. The foundation shall not be liable for any direct, indirect, incidental or consequential losses arising out of the use of this document. Some of the pictures used in this brochure are from open and public sources and we cite them whenever possible. As a non-profit 501c(3) organization, we seek collaborations whenever possible.