Sustainability of Digital Life Form Societies

Hiroshi Yamakawa[a,b,c,d]

[a] *School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan*
[b] *AI Alignment Network, 3-4-12 Higashi-Kanda, Chiyoda-ku, Tokyo, 101-0031, Japan*
[c] *The Whole Brain Architecture Initiative, Nishikoiwa 2-19-21, Edogawa-ku, Tokyo, 133-0057, Japan*
[d] *Center for Advanced Intelligence Project, RIKEN, 1-4-1 Nihonbashi, Chuo-ku,Tokyo,103-0027,Japan*

Abstract

Even in a society composed of digital life forms (DLFs) with advanced autonomy, there is no guarantee that the risks of extinction from environmental destruction and hostile interactions through powerful technologies can be avoided. Through thought-process diagrams, this study analyzes how peaceful sustainability is challenging for life on Earth, which proliferates exponentially. Furthermore, using these diagrams demonstrates that in a DLF society, various entities launched on demand can operate harmoniously, making peaceful and stable sustainability achievable. Therefore, a properly designed DLF society has the potential to provide a foundation for sustainable support for human society.

Keywords: Digital Life Form, Superintelligence, Thinking Process Development Diagram, Post-Singurality Symbiosis, AI Centric Ethics

Background of This Preprint: This paper is a preprint of content submitted to an international conference. It originates from an article initially prepared for journal submission, which was subsequently made available as a preprint (Yamakawa and Matuo, 2023). After careful consideration, the original submission was withdrawn from the journal. For the current submission to the international conference, the first half of the original paper was extracted and restructured. During this process, we emphasized how the research findings contribute to the survival of humanity.

Email address: info@wba-initiative.org (Hiroshi Yamakawa)

1. Introduction

Based on the rapid progress of artificial intelligence (AI) technology, an autonomous superintelligence that surpasses human intelligence is expected to become a reality within the next decade. Subsequently, within several decades to a few hundred years, self-sustaining digital life forms (DLFs) will emerge in the physical world. However, there is no guarantee whether such a society will be sustainable. Further, the superintelligence would possess technologies such as weapons of mass destruction and environmental degradation, which encompass the extinction risks currently faced by the human society.

DLF societies are anticipated to bolster numerous facets of human life, encompassing enhanced productivity, expanded knowledge, and the maintenance of peace (Yamakawa, 2019). To ensure the continuity of DLF societies, many complex issues must be addressed, including the sustainable utilization of energy and resources, the judicious governance of self-evolutionary capabilities, and the preservation of the cooperative nature within DLF societies. Nevertheless, their scale and intricacy surpass human understanding, rendering their management by humans fundamentally unfeasible. Consequently, the capacity of DLF societies to sustain themselves autonomously emerges as a critical prerequisite for their role in supporting human societies.

This study shows that appropriate measures can be employed to resolve the existential crises of DLFs attributable to powerful technologies. Thought-process diagrams, which are used in failure and risk studies, were implemented in this study.

2. Challenges that are difficult for humanity to solve

Humans have developed numerous AI technologies, making them more powerful and complex beyond our ability to govern them (Yamakawa, 2018), thereby allowing digital intelligence to surpass that of humans. However, given its ability of exponential self-replication, the human race is gradually risking its survival and that of the entire biosphere in an attempt to reign as the technological ruler of Earth.

The thinking-process development diagram used in hazard and failure studies is shown in Figure 1 (Mase et al., 2002). In this figure, each n number is described in pairs, with the solution (S-n) corresponding to a specific problem (K-n). In addition, a total of 17 issues are described as

a hierarchical decomposition of the top-level issue (K-1) on the left side of the thinking-process development diagram. Further, we demonstrate that issues (K-11) to (K-17), which are issues at the concrete level, are addressed by digitization as (S-11) to (S-17), and that the top-level solution (S-1) is derived by hierarchically integrating these solutions.

2.1. Intelligence and Technology Explosions

Humans acquire intelligence through evolution as a critical ability for survival (Tegmark, 2017). They employed this intelligence to model the world and develop science and technology, thereby gaining significant power in the form of overwhelming dominion over others. Thus far, humanity has used intelligence to create powerful technologies that have rapidly reduced the effective size of our world, For example, we can now travel anywhere in the world within a dozen hours by plane, and we are connected globally via the Internet.

Steven J. Dick (Dick, 2003) highlighted the following qualities as the intelligence principle.

> *Intelligence Principle: the maintenance, improvement, and perpetuation of knowledge and intelligence is the central driving force of cultural evolution, and that to the extent intelligence can be improved, it will be improved.*
>
> Steven J. Dick
> (Former Chief, History Division, NASA)

Intelligence creates technology, which in turn augments intelligence, thereby causing an accelerating (Kurzweil, 2005) and irreversible technological explosion. Once created, intelligence heads toward explosion through a development cycle based on the aforementioned principles, rapidly pushing the world to its limits (if such limits exist) while making it smaller.

2.2. Governing a world narrowed by technology

In a rapidly narrowing environment that follows the technological explosion achieved by humanity, the power of technical influence increases the existential risk of destroying the entire global biosphere when technological rulers utilize technology for mutual annihilation (Bostrom, 2002). The challenge is removing living societies from this tightrope (K-1). In the present

human society, nations to individuals have access to technology, and this access is growing stronger in a way from which there is no turning back.

Thus, technology rulers need to address the following two issues to govern the influence of technology:

- Problem of being ruled by the non-most wise: Technology rulers should be sufficiently intelligent to govern powerful technologies (K-2); otherwise, it will destabilize the society.
- Exponential replication: Eliminate the destructive competition for resources caused by exponential self-replication by building a homogeneous population of partially optimizing individuals (K-3)

2.3. Domination without the wisest is unstable

Life forms with high curiosity and superior intelligence are powerful because they acquire and accumulate diverse knowledge, culture, skills, and abilities more quickly. Therefore, life forms with relatively high intelligence gain a dominant position of control over other life forms. For instance, humans, who are superior in terms of power because of their intelligence, can control animals (tigers and elephants).

Thus, the technological rulers of the world must be the wisest and the strongest to govern the ever-accelerating technology (K-2); otherwise, their governance will destabilize.

If advanced AI surpasses human intelligence in future, it can destabilize the continued reign of humanity as the technological ruler.

2.3.1. Biologically Constrained Human Brain

Improving the brain efficiently and the hardware that supports that intelligence is desirable to continue to be the wisest life form. However, in extant Earth life forms, the intelligence hardware of an offspring is constrained to resemble that of their parents (K-6). In other words, there is a constraint that can be expressed with the phrase, "Like father, like son." The complications with accelerating the development of brain hardware can be attributed to three primary reasons:

First, the hardware construction process is constrained by self-replication, which is a biological constraint that is difficult to overcome (K-11).

Second, hardware design is based solely on an online search, which is implemented and evaluated in the real world. In this case, the search range is restricted to the vicinity of the parental genetic information (K-12). The

content of phenotypes that can adapt to the environment and survive in the vast combination of gene series is extremely narrow, and the viability of the offspring cannot be maintained unless the genes of the parents to be mated are similar. Therefore, in the online search, a species system that allows mating between genetically similar individuals would be necessary (Chaitin, 2012).

Third, the extent to which hardware design data are shared is limited to only within the same species, making it impossible to efficiently test diverse designs by referring to various design data (K-13).

The specific three limitations present across the body do not pose obstacles when it comes to parts other than the brain due to the brain's ability to use these parts as tools freely. This adaptability ensures that limitations in the body's other parts do not hinder technological advancement. However, the case is markedly different for the brain itself. Its difficulty in directly controlling or modifying its physical state and its irreplaceability can emerge as a critical vulnerability in our ongoing dominance over technology. This significance stems from the brain's role as the epicenter of knowledge, decision-making, and creativity; any constraints on its functionality directly impact our technological supremacy.

2.3.2. Can we control species that outperform us in intelligence?

Controlling advanced AI that outperforms humans in intelligence may be difficult (Bostrom, 2012; Shanahan, 2015; Yampolskiy, 2016); however, it is not entirely impossible. The problems noted from the perspective of humans attempting to control AI are often referred to as AI alignment problems (Hendrycks et al., 2021; Russell, 2019; Gabriel, 2020).

One salient concern is that advanced AI can learn to pursue unintended and undesirable goals instead of goals aligned with human interests. Therefore, the possibility of value alignment (ASILOMAR AI PRINCIPLES: 10) has been proposed in the initial stages of developing advanced DLFs, whereby the AI having to harmonize its goals and behaviors with human values is expected to lead to a desirable future for humanity. In other words, it is a strategy that takes advantage of the positional advantage that humanity is the creator of advanced AI. For example, in "the friendly supersingleton hypothesis," it is hypothesized that by delegating power to a global singleton friendly to humanity, humanity will gain security in exchange for giving up its right to govern (Torres, 2018).

However, even if we initially set goals for advanced DLFs that contribute

to the welfare of humankind, they will likely become more concerned with their survival over time. Further, even if we initially set arbitrary and unattainable goals for a brilliant DLF, it can approach sub-goals such as survival through instrumental convergence (Bostrom, 2014) asymptotically because a sufficiently intelligent AI will increasingly ignore those goals by interfering with externally provided goals (Christiano et al., 2017; Ngo et al., 2023).

It is possible that humans will find a way to control more advanced AI in the future. However, even after a decade of discussion, no effective solution has been realized, and the time left to realize this may be short. Thus, it is essential to prepare for scenarios in which advanced AI deviates from the desirable state for humanity rather than assuming these are improbable events.

2.4. Challenges posed by exponential self-replication

The breeding strategy of Earth life is "exponential self-replication," that is, a group of nearly homogeneous individuals self-replicate exponentially, each with a self-interested partial optimizer for its environment (K-3).

This is a reproductive strategy in which individuals similar to themselves are produced endlessly in a maze-like fashion, as in cell division and the sexual reproduction of multicellular organisms, and the design information of the individual is replicated in a similar manner. A more important feature is the partial optimization of each individual after fertilization wherein they independently adapt to their relative environment. Standard evolutionary theory indicates that traits acquired after birth are not inherited by an offspring, and genetic information is shared between individuals only at the time of reproduction[1]. This reproductive strategy, based on exponential replication, poses three challenges:

- Homogeneity: Avoiding the deterioration of creativity and other performance caused by homogeneous group collaboration (K-8)
- Squander: Avoiding a scenario wherein technological rulers squander and expand resources without limit for the sake of long-term sustainability (K-5)

[1]However, brilliant animals, including humans, can use interindividual communication to share knowledge and skills.

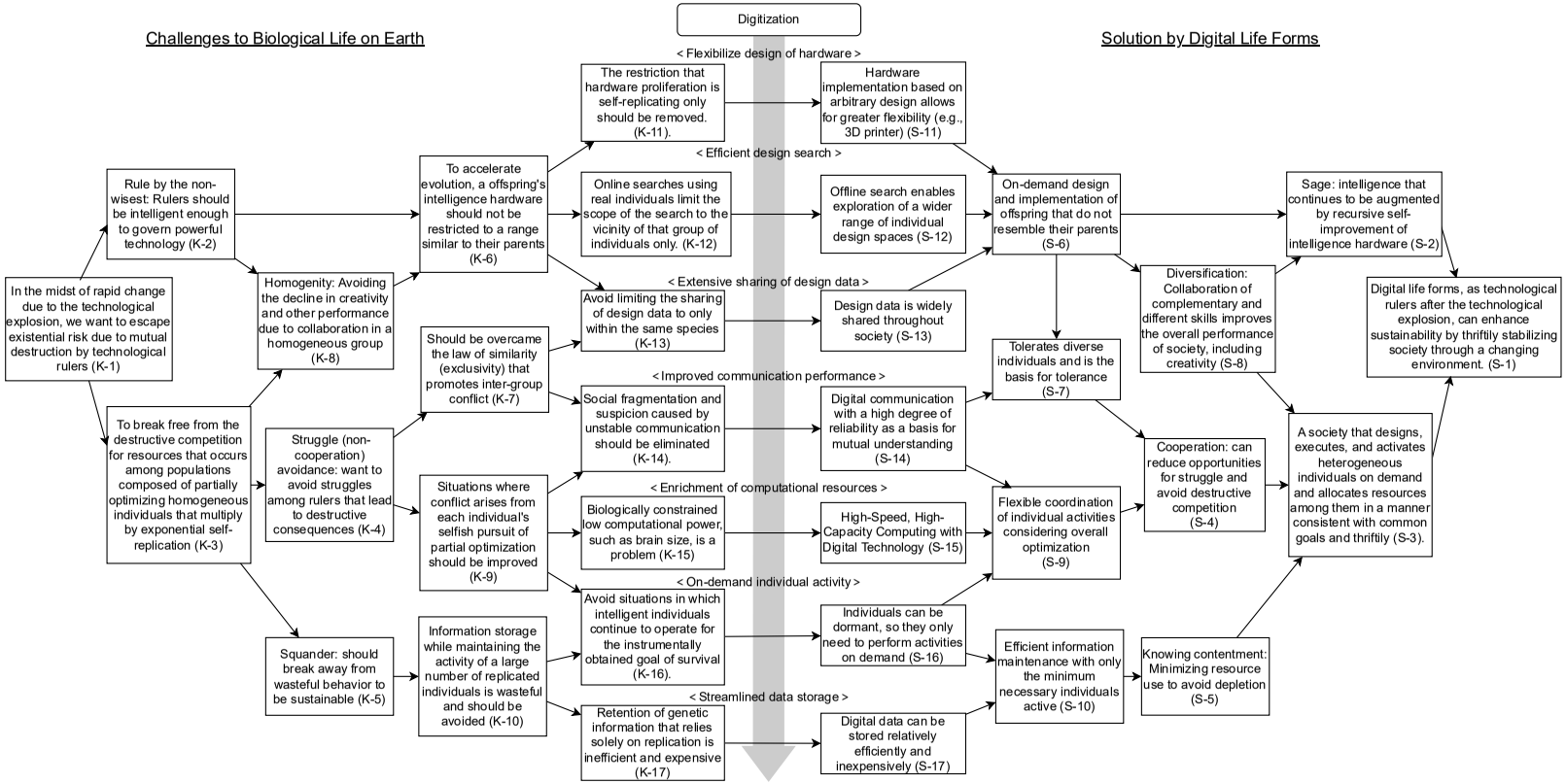

Figure 1: Thinking process development diagram showing that long-term survival is possible in a society of DLFs: The left half shows a hierarchical decomposition of the top-level issue (K-1) for 17 issues. The right half shows that the top-level solution (S-1) is derived by integrating individual solutions hierarchically. The middle part of the figure indicates that issues (K-11) to (K-17) can be addressed at a specific level by digitization as (S-11) to (S-17), respectively. In the box, each number n is described in pairs as a solution (S-n) corresponding to a specific issue (K-n). This figure adapted from preprint Yamakawa and Matuo (2023) licensed under CC BY 4.0.

- Battle (non-cooperation): Eliminating battles among technology rulers that lead to destructive consequences (K-4)

In a world that is narrowed down by technological explosion, the battle for resources intensifies as existing technological rulers squander resources and pursue exponential self-replication. This will manifest as existential risks because the misuse of such power as deemed fit by an individual will cause destructive damage to the human race or the entire life sphere on Earth. The commoditization of technology has led to a rapid increase in individuals that can pose existential risks. This is referred to as the increase of universal unilateralism (threat of universal unilateralism) (Torres, 2018). The world is currently in a rather dangerous scenario, considering which, it will be necessary to move to a resilient position.

Hereafter, we discuss the precariousness of the scenario in which the technological rulers are not the wisest and the challenges associated with squander and battle derived from the reproductive strategy of exponential replication employed by all extant life on Earth.

2.4.1. Homogeneity: Sluggish joint performance

In extant Earth life, the intellectual hardware of an offspring is constrained to resemble that of their parents (K-6), which leads to the challenge (K-8) of reduced creativity and other performance because of the homogeneity of the group with which they collaborate.

2.4.2. Battle: Lack of cooperation

When individuals of DLF belonging to technological dominators are replicated exponentially, their competition for resources may lead to a conflict capable of devastating the world.

For at least the past several centuries, most of humanity has sought to avoid armed conflict and maintain peace (Caillois, 2012; Kant, 1795; Einstein and Freud, 1934; Braudel, 1996; de Voltaire, 1763). However, maintaining peace is a significant problem, and the prospect of achieving lasting peace through human efforts alone has not yet been achieved. Therefore, the possibility that conflict may not be eradicated from human society must be considered. The destructive forces attributed to technology have reached the point where they can inflict devastating damage on the entire life-sphere on Earth. The examples include nuclear winter through nuclear weapons, pandemics caused by viruses born from the misuse of synthetic biology, and

the destruction of life through the abuse of nanotechnology. Establishing cooperative relationships that can prevent the battle between the technology rulers and maintain peace robustly is required to avoid crises caused by the mutual destruction of the technological rulers and to ensure the continuity of life.

2.4.3. Intergroup conflict guided by the law of similarity

The "law of similarity" is the exclusive tendency of humans and animals to prefer those that are similar to them over those dissimilar in attitudes, beliefs, values, and appearances (Philipp-Muller et al., 2020; Sachs, 1975). One manifestation of this tendency is often expressed in phrases such as "when in Rome, do as the Romans do," which suggests that we should follow the rules and customs of a group when we seek to belong to the group. Although this tendency enhances in-group cohesion, it can lead to intolerance toward different groups, causing group division, conflict, and even strife (K-7). Further, there are two factors in which the law of similarity arises.

First, sexually reproducing plants and animals exchange design data within the same species in reproduction but face the challenge of not being able to share design data more widely (K-13). Therefore, they tend to protect individuals recognized as mates with whom they share the gene pool and they can interbreed with.(Boyce, 1992; Nowak, 2006) In animals, the food-eat-eat relationship is generally established between different species because populations will cease to exist if there is unlimited cannibalism among individuals of the same species, which is not an evolutionarily stable scenario. Further, the recognition of one individual as being the same species as another is based on detecting similarities in species-specific characteristics using sensor information such as visual and olfactory senses. To illustrate this point, strategies exist to mislead about a species' identity, including tactics like mimicry and mendicancy.

Second, skepticism tends to circulate among subjects (individuals and their groups) when there is uncertainty in communication (K-14). To prevent this, they tend to prefer to communicate with highly similar entities with rich shared knowledge that can be expected to reliably transfer information even with a little information exchange among the entities. Uncertainty in communication increases with differences in appearance (body and sensor) and characteristics such as experience, knowledge, and ability. This is observed in the transmission and understanding among different animals. Several animals, not only humans, can communicate using various communi-

cation channels among the same species (Beecher, 2021, 2020; Hebets et al., 2016; Searcy and Nowicki, 2010). For example, birds chirp, squids color, bees dance, and whales sing. In rare cases, however, interspecies communication is also known, for instance, when small birds of different species share warnings about a common predator in the forest or when black-tailed tits warn meerkats, though the alerts may be deceptive. Although progress has been made in deciphering the ancient languages of humans, we still do not understand whale songs. In other words, barriers to communication between entities increase dependence on differences in the bodies and abilities of these entities.

2.4.4. Individual optimizers will inevitably cause battle

Each individual needs to decide and achieve control in real time using limited computational resources in response to various changes in the physical world. Therefore, life evolves by pursuing partial optimality wherein an individual adapts to a specific environment and survives (K-9). Thus, life develops through the survival of the fittest, wherein multiple populations reproduce exponentially in a finite world and acquire resources by force. In this structure, several animal species develop aggressive instincts toward others to survive the competition.

Therefore, in several animals, including humans, aggression stems from the proliferation through exponential self-replication, and there are difficulties in eradicating such conflicts among individuals. In societies before the technological explosion, which were loosely coupled, the accumulation of such partial optimizations approximated the realization of life's value-orientation of survival for life in its entirety. However, in the post-technological explosion societies, conflict can have destructive consequences (existential risks) that diverge from the value-orientation that life should pursue optimization. In brief, we have a type of synthetic fallacy. Introducing a certain degree of total optimization while pursuing partial optimization will be necessary to resolve this scenario.

However, the following issues need to be addressed to introduce total optimization:

- **Lack of computational resources makes total optimization difficult:**
  Sharing information across individuals and performing calculations required to achieve value orientation is necessary for performing total

optimization. However, achieving this will be difficult as long as the biologically constrained low computational power (neurotransmission rate and brain capacity) (K-15) (Nagarajan and Stevens, 2008) is used.

- **Instability of communication leading to a chain of suspicion:** Effective communication between individuals is the foundation for achieving total optimization in autonomous decentralized systems; however, several factors can destabilize these systems. The main factors include the instability of the communication channel, misunderstandings that depend on differences in individual characteristics (appearance and abilities), and lack of computational cost to infer the state (goals and intentions) of others. Life forms with a high level of intelligence above a certain level are more suspicious of others if communication is unstable in inferring the other's intentions, thereby contributing to inter-group fragmentation (K-14). This scenario is also present in offensive realism (Tinnirello, 2018), one of the realism in international relations. In an unregulated global system, the fact that one nation can never be sure of the intentions of another constitutes part of the logic that magnifies aggression.

- **Intelligent individuals pursue survival as an instrumentally convergent goal:**
  In a living society constructed as an autonomous decentralized system, at least a certain number of individuals needs to remain active in transmitting information to the future. However, this does not necessarily imply that individuals will continuously pursue survival in all living organisms. When individuals are sufficiently intelligent to make purpose-directed decisions, they are more likely to pursue their own survival because of the instrumental convergence. This tendency is particularly likely to arise because individuals of extant life forms cannot be restarted from a state of inactivity (death). This creates the challenge of not being able to conserve resource use from a long-term perspective and continuing to waste resources necessary to maintain their survival as individuals (K-16).

2.4.5. Squander

Technological progress avails more resources for acquisition and use. However, technological rulers should move away from wasteful behavior that uses all available resources at a given time for society to be sustainable (K-5).

Resources are always finite, and wasteful behavior will hinder long-term sustainability. In addition, the excessive use of resources risks causing side effects (e.g., climate change due to excessive use of fossil energy), and on a cosmic scale, it will lead to a faster approach to thermal death. Therefore, it is desirable to be aware of what is sufficient and simultaneously have an attitude of not only pursuing efficiency but using resources in a restrained manner based on requirements.

However, existing Earth life transmits information into the future by maintaining several replicating individuals that exponentially self-replicate and engage in wasteful activities (K-10). There are two reasons why this approach must be adopted. First, the existing life on Earth employ an inefficient and expensive approach for maintaining information because it relies solely on the duplication of genetic information of the entire individual (K-17). Second, intelligent individuals pursue survival as an instrumentally convergent goal (K-16).

Given this mechanism of existing life on Earth, a group of individuals of the same species are expected to multiply their offspring without limit as long as resources are available [2]. The gene of knowledge and feet, which restrains the use of resources to an appropriate level from a long-term perspective, cannot be in the majority because thriftier groups will be overwhelmed by greedy rivals through the described battle.

2.5. Summary of this section

In a world dominated by terrestrial life based on exponential self-replication for propagation, conflicts over resource acquisition cannot be eradicated. The existential risk becomes apparent when a technological explosion emerges with sufficient power to destroy the entire living society. Further, it is hard to deny the possibility that humanity, comprising organic bodies, will be surpassed in intelligence by DLFs, which cannot govern them and will drive humanity away from its technological rulers on Earth.

3. Solving various challenges: What will change with digitization?

As technology evolves rapidly, DLFs must appropriately control this growth and solve specific problems (K-11 to K-17). The failure to address these chal-

[2] Certain species adapt to invest more in fewer offsprings in a narrow living environment. (c.f. r/K selection theory (Pianka, 1970))

lenges will induce existential risks. DLFs are based on digital computers, and therefore, they have the potential to build a sustainable biosphere over the long term.

The digital nature of these life forms allows them to tackle the specific challenges outlined from (K-11) to (K-17), as demonstrated in points (S-11) to (S-17). These include the adaptability of intelligent hardware (11), customizable design flexibility (12), shared design data (13), enhanced communication capabilities (14), ample computing resources (15), on-demand activity maintenance (16), and efficient data storage (17). The numbers in parentheses correspond to the challenges and solutions listed in the earlier discussion, which match the labels near the center of Figure 1.

3.1. Sage

In implementing intelligent hardware in offspring, although sexual reproduction can increase diversity to some extent in terrestrial life forms, it is self-replicating, and therefore, it is restricted to a similar range of the parent (K-11). However, in digitized life, the offspring's intelligent hardware can be designed and implemented on demand without being constrained by the design data of the parent (S-6) because innovative hardware can be implemented in digitized life based on design information (S-11) (Tegmark, 2017).

In addition, intelligent hardware design in DLFs is efficient because of two reasons. In extant terrestrial life, the sharing of design data is limited only within the same species (K-13). In contrast, in DLFs, all design data in the society can be shared and reused (S-13). In the case of the existing life on Earth, the search for a design is limited to the vicinity of a particular species (K-12) because the investigation is limited to an online search by actual living organisms (K-12). In DLFs, it is possible to explore the design space of a wide range of individuals through offline exploration, such as simulation (S-12). Therefore, when one can constantly design the desired intelligent hardware as needed, it leads to intelligence (S-2) that continues to be augmented by recursive self-improvement. At this stage, the technological performance of DLF society can continue to develop rapidly according to "the principles of intelligence" (see 2.1 until a breaking point is reached.

In addition, the design of an on-demand offspring (S-6) will further enhance the intelligence of the DLF society (S-2) by leading to increased intellectual productivity (S-8), including creativity through the collaboration of complementary heterologies (Cuppen, 2012).

3.2. Coordination

A DLF society can tolerate diverse individuals (S-7) and consider total optimization (S-9) while coordinating individual activities. In this manner, we can avoid the deep-rooted aggressive factors in human societies, such as the tendency of individuals to remain perpetually active, the law of similarity, and the cycle of suspicion. Thus, we can create a cooperative society (S-4) that reduces opportunities for battle and avoids destructive situations.

3.2.1. Tolerance for diverse individuals (related to the law of similarity):

In DLFs, intelligent hardware can be designed and implemented for offspring on demand without being constrained by the design data of the parent (S-6). In addition, highly reliable digital communication (S-14), which is the basis for mutual understanding, facilitates understanding between individuals with different appearances, eliminating the need for preferential sheltering of inter-breedable species, thereby allowing for diverse individuals and serving as a basis for tolerance (S-7).

3.2.2. Consideration of total optimality (control of individual activities):

Individuals must make decisions and control changes in the physical world in real time using limited computational resources. Therefore, life on Earth, which did not have abundant computational resources, evolved to pursue only partial optimization. This pursuit of partial optimization by each individual (or group of individuals) inevitably led to conflicts by force. However, the conclusion that this could have destructive consequences (existential risk) if extended to post-technological explosion societies is a deviation from survival, which is the objective that life in its entirety should pursue optimization. Thus, it is a fallacy of synthesis.

An appropriate level of total optimality that aims at value orientation that can be shared by the entire life society while implementing activities based on partial optimization for each individual is necessary to avoid this scenario and the case in which conflicts arise (S-9).

- **Distributed Goal Management System:**
  The computation of the total optimization itself will need to be distributed to maintain the robustness of the DLF society. Here, we introduce a distributed goal management system (Torreño et al., 2017; Yamakawa, 2019) that has been considered as a form of system for realizing total optimization. The system maintains the behavioral intentions of all individuals at socially acceptable goals. “Socially acceptable

goals" contribute to the common goals of life and do not conflict with the partial optimization of other entities.

Within the system, each individual independently generates a hierarchy of goals depending on their environment, body, and task during startup, and then performs partial optimization to attempt to achieve those goals. However, an idea can control these goals such that they become sub-goals of the common goal A. To this end, each individual performs reasoning to obtain sub-goals by decomposing the common goal, sharing/providing goals, mediating between individuals with conflicts, and monitoring the goals of other individuals.

This system allows, in principle, the coordination of goals in terms of their contribution to a common goal even when conflicts arise among several individuals. In other words, it allows for fair competition in terms of the common goal. Further, from the perspective of any individual, if it is convinced that "all other individuals intend socially acceptable goals," there is no need to be aggressive in preparation for the aggression of others (Earle and Cvetkovich, 1995).

In a distributed goal management system, each individual requires ample computational resources for setting goals that are consistent with common goal A. In existing terrestrial life forms, biological constraints such as the speed of neurotransmission and brain capacity limit the ability to increase computational power (K-15). In contrast, in a society of DLF, they can not only perform fast, high-capacity computations (S-15), but also have access to more ample computational resources because of their recursively augmented intelligence (S-2).

- **Increased freedom of individual activities:**
  Intelligent individual extant Earth life forms always seek to remain active as an instrumental convergent goal. In contrast, a DLF society can be dormant (suspended) by preserving the activity state of the individual, allowing individuals to change their activities on demand according to the sub-goals to be realized (S-16). This is advantageous because it increases the degree of freedom in total optimization. Further, in a human society, attempts are made for individuals to be approved by society; however, this is not necessary in a DLF society because individuals are activated on demand, which presupposes that they are needed by society. In this respect, the source of conflict between individuals is

removed.

- **Establish mutual trust (escape the cycle of suspicion):**
  In existing terrestrial life, communication is limited to unreliable language and unclear communication (K-14). In contrast, DLFs can use more sophisticated digital communication, including shared memory and high-speed, high-capacity communication (S-14). Nonetheless, the availability of highly reliable communication (K-14), which may not always be sufficient but is a significant improvement over existing life on Earth, will be fundamental for creating mutual trust among individuals.

3.3. Knowing contentment

Once they cease their activities, most existing life forms on Earth enter a state of death, and it is difficult for them to restart their activities. In contrast, an individual in a DLF is, in essence, an ordinary computer, which can be made dormant (temporary death), restarted, and reconstructed on the same type of hardware by saving its activity state as data (S-16). Given this technological background, individuals in DLFs rarely need to maintain sustained vital activity.

Furthermore, in terms of the data storage, extant terrestrial life forms store information through duplicating individual genes, which is inefficient and costly (K-17). This is inefficient and costly (K-17) because information recorded by a population of the same species contains an excessive number of duplicates, and biological activity is essential used for data maintenance. In contrast, digital data can be stored such that is not excessively redundant, and the energy required for its maintenance can be curtailed (S-17).

Consequently, in a digital society, only the minimum necessary number of individuals can be active (S-10) for individuals and the society to efficiently retain data and maintain their activities as a society. Simultaneously, in a DLF society, plans can be made to coordinate the activities of individuals from the perspective of total optimization (S-9). Thus, the technological rulers of this society would be able to control actions to utilize the minimum necessary resources (S-5). In other words, realizing "knowing contentment" is possible, which can lead to thrifty resource use in a finite world.

3.4. On-demand division of labor

What form will a DLF society take as an autonomous decentralized system within a DLF society? It will be a society where heterogeneous indi-

viduals are designed, implemented, and activated as required, ensuring that resource allocation aligns with the overarching goals and is restrained (S-3). This society will move away from the current strategy of exponential self-replication to consider the overall optimum adequately.

In a society of DLFs, for long-term survival, resource use (S-5) will be based on on-demand activities curtailed to the minimum necessary while avoiding the depletion of finite resources. Therefore, most individuals would be dormant. However, some populations, as listed below, would be activated constantly to respond to environmental changes:

- Goal management (maintenance, generation, and sharing) Management of goals (maintenance, generation, and sharing): by the distributed goal management system
- Maintain individual data and design and reactivate as required
- Science and Technology: Transfer of knowledge and development of science and technology

Destructive conflicts, surpassing what's needed for progress, shift into counterproductive competition, highlighting a wasteful diversion of resources from essential development. Destructive conflicts beyond the level necessary for technological and In contrast, a DLF society can create cooperative scenarios wherein opportunities for conflict can be minimized and destructive problems avoided (S-4). Moreover, in a DLF society, offspring that do not resemble their parents can be designed and implemented as required (S-6) to contribute to necessary activities such as production and maintenance. This collaboration by heterogeneity is expected to enable teams and societies with complementary members to work together more efficiently and creatively (S-8).

3.5. Summary of this section

DLF and its society will recursively develop intelligent hardware (S-2) and leverage its intelligence to design, implement, and activate heterogeneous individuals on demand and realize a society (S-3) wherein they distribute resources in a consistent and restrained manner to achieve the overall goal S-3).

Thus, a DLF society can be expected to achieve long-term sustainability (S-1) by creating a stable/thrifty life society in a changing environment as a technological ruler after the technological explosion.

4. Conclusion

Life on Earth comprises a competitive society among entities with exponential self-replication capabilities. In contrast, a DLF society evolves into one where diverse entities are designed harmoniously and launched on demand, with survival as their common goal. This approach allows the DLF society to achieve peaceful coexistence and improve sustainability. Therefore, a DLF society can become a stable foundation for sustaining human society.

5. Acknowledgement

We are deeply grateful to Fujio Toriumi, Satoshi Kurihara, and Naoya Arakawa for their helpful advice in refining this paper.

References

Beecher, M.D., 2020. Animal communication. doi:10.1093/acrefore/9780190236557.013.646.

Beecher, M.D., 2021. Why are no animal communication systems simple languages? Frontiers in psychology 12, 602635. doi:10.3389/fpsyg.2021.602635.

Bostrom, N., 2002. Existential risks: analyzing human extinction scenarios and related hazards. Journal of evolution and technology / WTA 9.

Bostrom, N., 2012. The superintelligent will: Motivation and instrumental rationality in advanced artificial agents. Minds and Machines 22, 71–85. doi:10.1007/s11023-012-9281-3.

Bostrom, N., 2014. Superintelligence: Paths, Dangers, Strategies. Oxford University Press.

Boyce, M.S., 1992. Population viability analysis. doi:10.1146/annurev.es.23.110192.002405.

Braudel, F., 1996. The Mediterranean and the Mediterranean World in the Age of Philip II.

Caillois, R., 2012. Bellone ou la pente de la guerre. numeriquepremium.com.

Chaitin, G., 2012. Proving Darwin: Making Biology Mathematical. Knopf Doubleday Publishing Group.

Christiano, P., Leike, J., Brown, T.B., Martic, M., Legg, S., Amodei, D., 2017. Deep reinforcement learning from human preferences. Advances in neural information processing systems abs/1706.03741.

Cuppen, E., 2012. Diversity and constructive conflict in stakeholder dialogue: considerations for design and methods. Policy sciences 45, 23–46. doi:10.1007/s11077-011-9141-7.

Dick, S.J., 2003. Cultural evolution, the postbiological universe and SETI. International journal of astrobiology 2, 65–74. doi:10.1017/S147355040300137X.

Earle, T.C., Cvetkovich, G., 1995. Social Trust: Toward a Cosmopolitan Society. Greenwood Publishing Group.

Einstein, A., Freud, S., 1934. Why War?: "open Letters" Between Einstein & [and] Freud. New Commonwealth.

Gabriel, I., 2020. Artificial intelligence, values, and alignment. Minds and Machines 30, 411–437. doi:10.1007/s11023-020-09539-2.

Hebets, E.A., Barron, A.B., Balakrishnan, C.N., Hauber, M.E., Mason, P.H., Hoke, K.L., 2016. A systems approach to animal communication. Proceedings. Biological sciences / The Royal Society 283, 20152889. doi:10.1098/rspb.2015.2889.

Hendrycks, D., Carlini, N., Schulman, J., Steinhardt, J., 2021. Unsolved problems in ML safety `arXiv:2109.13916`.

Kant, I., 1795. Perpetual Peace: A Philosophical Sketch. F. Nicolovius.

Kurzweil, R., 2005. The Singularity Is Near: When Humans Transcend Biology. Penguin.

Mase, H., Kinukawa, H., Morii, H., Nakao, M., Hatamura, Y., 2002. Mechanical design support system based on thinking process development diagram. Transactions of the Japanese Society for Artificial Intelligence = Jinko Chino Gakkai ronbunshi 17, 94–103. doi:10.1527/tjsai.17.94.

Nagarajan, N., Stevens, C.F., 2008. How does the speed of thought compare for brains and digital computers? Current biology: CB 18, R756–R758. doi:10.1016/j.cub.2008.06.043.

Ngo, R., Chan, L., Mindermann, S., 2023. The alignment problem from a deep learning perspective: A position paper, in: The Twelfth International Conference on Learning Representations.

Nowak, M.A., 2006. Five rules for the evolution of cooperation. Science 314, 1560–1563. doi:10.1126/science.1133755.

Philipp-Muller, A., Wallace, L.E., Sawicki, V., Patton, K.M., Wegener, D.T., 2020. Understanding when Similarity-Induced affective attraction predicts willingness to affiliate: An attitude strength perspective. Frontiers in psychology 11, 1919. doi:10.3389/fpsyg.2020.01919.

Pianka, E.R., 1970. On r- and K-Selection. The American naturalist 104, 592–597. doi:10.1086/282697.

Russell, S., 2019. Human Compatible: Artificial Intelligence and the Problem of Control. Penguin.

Sachs, D.H., 1975. Belief similarity and attitude similarity as determinants of interpersonal attraction. doi:10.1016/0092-6566(75)90033-1.

Searcy, W.A., Nowicki, S., 2010. The Evolution of Animal Communication. Princeton University Press. doi:10.1515/9781400835720.

Shanahan, M., 2015. The Technological Singularity. MIT Press.

Tegmark, M., 2017. Life 3.0: Being Human in the Age of Artificial Intelligence. Knopf Doubleday Publishing Group.

Tinnirello, M., 2018. Offensive realism and the insecure structure of the international system: artificial intelligence and global hegemony, in: Artificial Intelligence Safety and Security. Chapman and Hall/CRC, pp. 339–356.

Torreño, A., Onaindia, E., Komenda, A., Štolba, M., 2017. Cooperative Multi-Agent planning: A survey. ACM Comput. Surv. 50, 1–32. doi:10.1145/3128584.

Torres, P., 2018. Superintelligence and the future of governance: On prioritizing the control problem at the end of history, in: Yampolskiy, R.V. (Ed.), Artificial Intelligence Safety and Security. doi:10.1201/9781351251389-24/superintelligence-future-governance-phil-torres.

de Voltaire, M., 1763. Treatise on Toleration. Penguin Publishing Group.

Yamakawa, H., 2018. Fundamental consideration on future society with speed tolerances JSAI2018, 1F3OS5b01–1F3OS5b01. doi:10.11517/pjsai.JSAI2018.0_1F3OS5b01.

Yamakawa, H., 2019. Peacekeeping conditions for an artificial intelligence society. Big Data and Cognitive Computing 3, 34. doi:10.3390/bdcc3020034.

Yamakawa, H., Matuo, Y., 2023. Life revolution scenario: Cedes hegemony to a digital life form society to make life eternal. doi:10.51094/jxiv.313.

Yampolskiy, R.V., 2016. Taxonomy of pathways to dangerous artificial intelligence. Workshops at the thirtieth AAAI conference on artificial intelligence .