On free will, intelligence and cybernetics

Mark Tropin, Ju Wu, and Alexandra Korukova

May 2021

Contents

1 Introduction

Our paper focuses on the subject of differentiating human beings from other species and machines.

We explore and compare several aspects of human behaviour with similar properties exhibited by animals and computers through the lens of intelligence, free will and cybernetics. We show that some of these aspects can manifest in the aforementioned agents, while others are exclusively human properties.

This paper can be used as a foundation for further exploration of the following topics encompassing artificial intelligence, computer science, philosophy (ethics), cybernetics, robotics, which are beyond the scope of this essay:

- Is it necessary to have free will in order to be intelligent?
- To what extent does AI and cybernetics allow for modelling of human intelligence?
- Is the existence of Artificial General Intelligence (AGI) possible?
- What properties define human intelligent behaviour? Can they be expressed as computation?
- What are the differences / similarities of human intelligence and AI?

Our treatment of the subject is structured as follows:

In the first section of our paper different aspects of intelligent behaviour are discussed. We explore a number of case studies which lead us to interesting discoveries regarding the nature of intelligence. We then use these findings as a support for our definition of intelligence. Finally, we explore structural and behavioral properties of intelligence and criteria of its assessment. The section ends with a segue to the following chapters on free will and cybernetics.

In the second section the question of free will is raised. First, several definitions of freedom of the will are discussed. Second, other points of view on this topic are presented.

Finally, in the penultimate section of our paper the decision making process is explored from the point of view of cybernetics. We explore the principles that define sensor-motor integrated systems and compare them with the "mechanisms" used by human beings and other living creatures. Our treatment of the subject highlights the similarities between the two and leads to an exploration of further developments in generalized artificial intelligence and the social risks associated with them.

The paper concludes with a recapitulation of the aspects of human behaviour explored in our paper.

2 Intelligence

In this section we will introduce the topic of intelligence and discuss its properties with a link to the problems of free will and decision making.

2.1 Introduction

The enigma of the nature of intelligence has entranced humanity for many centuries, dating back to first attempts by the Ancient Greeks [18]. Until the 19th century, the problem of human intelligence

has not left the realm of philosophy, but with the advent of the scientific method and the emergence of experimental psychology, new prospects and frameworks for the analysis and interpretation of the human mind have appeared.
In the 1950s, experimental advances in psychology and concurrent progress in electrical engineering and computer science have paved the way for a new area of study known as Artificial Intelligence (AI) [25]. Intelligence was now viewed through the prism of scientific discovery and contemporary engineering efforts, and the scientific community has embarked on a journey of attempting to model the human mind by the means of computation and logical inference.

2.2 Intelligence: Case Studies

Before we proceed to the definition of intelligence and its main characteristics, it is important that we first consider some experimental studies that have had a profound impact on the way we view intelligence today. These studies will provide us with key insights into the nature of intelligence and will help us in defining what intelligence is.

2.2.1 Pavlov dogs

The first clinical experiment that defined how we view learning and intelligence today is Ivan Pavlov's study of *classical conditioning*. Pavlov has observed that on top of having an unconditional, "hard coded" response to a stimulus (salivating before eating), dogs are also capable of developing conditional, *learned* stimuli as well. For instance, hearing the footsteps of one of his assistants or a ringing bell triggered the dogs' salivation as much as the actual food did.
This simple yet insightful experiment leads to two important conclusions that will be taken into account during our treatment of the definition of intelligence:

- First of all, this experiment shows that a major component of intelligence is the capacity to learn. In this case, the dogs were able to observe the environment and infer that certain changes (sounds) in the environment lead to certain outcomes (food arriving). This ability to discover new patterns in the environment can be seen as an example of *reinforcement learning*, one of the main branches of *machine learning* that models intelligent behavior.

- Another important observation that was made by Pavlov is that the dogs started to salivate even if they heard the footsteps of other assistants in the laboratory, not necessarily the ones that were feeding them. This leads to an intriguing conclusion: the dogs were not only able to enrich their set of patterns on some fixed instances of stimuli and responses, but were able to *generalize* their newly acquired knowledge.

2.2.2 Turing test

The Turing test was first introduced by Alan Turing in his seminal work "The Imitation Game". Prior to that, in "Computing Machines and Intelligence" Turing notes that the metaphysical question "Can machines think?" requires unequivocal definitions of the terms "machine" and "intelligence". Instead, Turing proposes tackling a similar problem: is it possible for a machine to display behavior that is indistinguishable from that of a human?

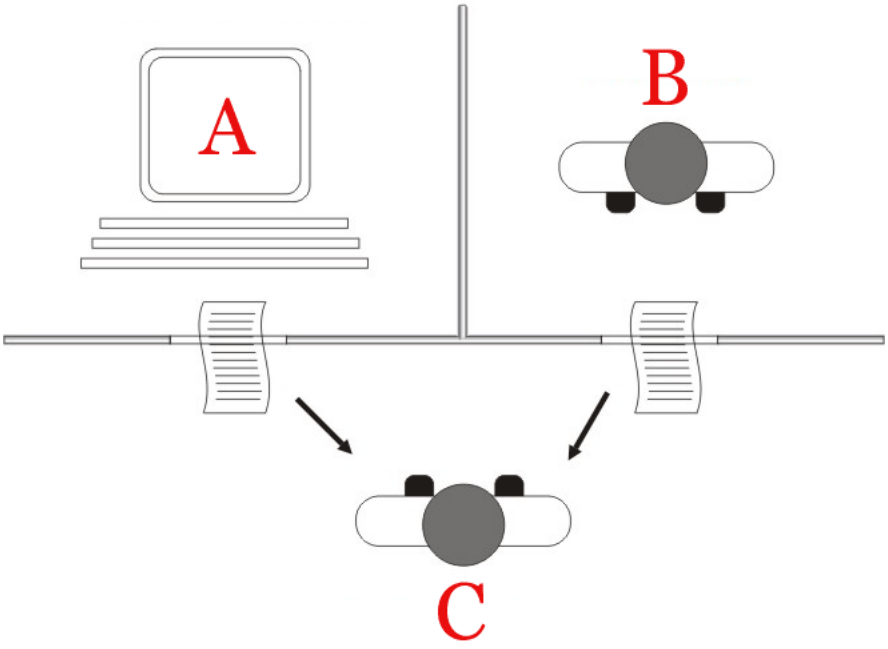

Figure 1: A brief overview of the Turing test.

In order to examine this proposition, Turing proposes the following test (see Figure 1): Imagine a judge (**C**) in a room that is linked to two other rooms. In one of this rooms one can find a computing device capable of Natural Language Generation (**A**), and in the second room one can find a human correspondent (**B**). Both the computer and the human communicate with the judge by means of transferring messages typed on paper through an opening in the walls of their rooms: thus there are no physical clues as to who might be located in which room.
If the judge is incapable of discerning whether he or she is currently conversing with a human or a computer over successive runs of the experiment (in statistical terms, the judge is right only 50% of the time and his estimations are therefore no better than random guessing), Turing concludes that we have reached the point where the intelligence of the computer has become equivalent to that of a human.

This hypothetical experiment highlights the importance of *interpretation* and *concept formation* as integral parts of human intelligence: it does that by evaluating the presence of intelligence via *language*.

Nevertheless, this experiment also raises multiple questions with respect to its "testability" and the validity of its results. For instance, how do we know if the judge is attentive to detail and honest in his evaluation? How do we know if the machine (presented here as a black box) actually generates new information instead of providing predetermined replies?
Perhaps, most importantly, what can the Turing Test say vis-à-vis other forms of intelligent behaviour? Evaluating intelligence in terms of language capabilities seems quite restrictive: how should we approach evaluating the intelligence of illiterate people?[1]

[1] A common test for identifying self-consciousness in animals consists in placing a mirror in front of them and observing their reaction. But what does this test say regarding the self-consciousness of *blind* animals?

2.2.3 Chinese room experiment

Some of the aforementioned questions were treated in "Minds, Brains, and Programs" by philosopher John Searle. Searle proposes a different thought experiment that sheds light on the faultiness of the Turing Test.

Imagine that Searle is locked in a room with a window which allows him to communicate by receiving and giving out paper sheets. Searle himself does not know a word of Chinese, but he has a list of flash-cards that link questions formulated in Chinese with their respective answers. Upon reading a question given to him on a card, Searle simply looks up an appropriate answer from his stack of cards and replies accordingly. Thus, a Chinese-speaking person talking to Searle would assume that the person in the room is fluent in Chinese, while we know that he clearly is not.

In his paper, Searle postulates that the same thing happens with computers: even if their behaviour may seem intelligent to us, that does not imply that the *underlying process* is intelligent in nature. Since computers (be it Turing machines, RAM models, or modern-day programming languages) are always defined by a finite number of operations, which means that they are bound to behave in predictable and deterministic ways, it is impossible for them to generate anything original and therefore they cannot be deemed intelligent.

This thought experiment elucidates many aspects of intelligence; we would like to focus on three of them that will be relevant for our treatment of the definition and criteria of intelligence:

- The ability to form concepts is a cornerstone of any intelligent system. Searle shows that language cannot be used as a metric of intelligence since it is possible to reduce it to mere manipulation of grammar rules and vocabulary substitutions.
- Logical inference is not equal to intelligence. Searle shows that pseudo-intelligent behaviour can be simulated by a deterministic machine that applies logical rules (rewriting) to a given set of logical facts (Chinese flashcards). For Searle, intelligence implies the ability of being creative and novel, which is not a property that can be expressed as computation.
- Tasks that may appear *generic* to us (such as language) can actually be *specific*. This conundrum has led to the recent controversies surrounding GPT-3, a state-of-the-art language model that is capable of writing anything from journal articles to computer code and conversing with a human correspondent in a way that may look almost indistinguishable from a human. Some have pointed out that its high performance on a vast array of language tasks is sufficient to state that GPT-3 has reached human-level intelligence, while others reject this claim by looking at the *architecture* of GPT-3: while GPT is an Artificial Neural Network, it has nothing to do with the neural networks found in our brains: it is simply a stack of encoders and decoders that process a stream of words. Moreover, its "objective function" is simply predicting the next word with the highest probability in a given sequence, which is far from the way humans reason about language.

2.3 Definition of Intelligence

2.3.1 Can submarines swim?

One of the forefathers of modern computer science, E.W. Dijkstra, has famously stated in his seminal work "Science fiction and science reality in computing": *"[the question] whether machines can think [is] as relevant as the question whether submarines can swim."* This famous statement (first appearing in 1984) has foreshadowed the coming of the "AI winter", a period of skepticism and distrust regarding AI and modeling of human intelligence.

Perhaps less famously (but more importantly for our discussion) this question was given an alternative answer by Peter Norvig. In his article "What do you think about machines that think?", Norvig points out that in Russian, submarines actually *do* swim; this implies that the issue of the existence of a hypothetical AI depends not as much on technology, but even more so on the *semantics* that we associate with terms like "thinking" and "intelligence".

2.3.2 Intelligence as adaptivity to environment

Having in mind our previous discussion of the difficulties of defining a non-trivial term like *intelligence*, we advance to a "working" definition of intelligence which will allow us to treat the behavioral properties of intelligence and define criteria for its assessment.

Why do we settle for a working definition? First of all, a definition that is too restrictive can hinder any computational modeling (AI) or experiments on human intelligence. Second, this kind of definition can discard interesting edge cases which may actually be intelligent but that don't fit a negligible fraction of the criteria. Third, a definition with many criteria will not allow for any "one-dimensional" measure of intelligence: as a result, we will not be able to compare the intelligence of two different agents when one of the is smarter than the other in one way, but less smart in some different way.

As it was mentioned in the famous Chomsky-Foucault debate, the same applies to the notion of life: it is impossible to provide a scientifically strict definition of this phenomenon given the vast array of its instances; nevertheless, we can agree on an intuitive and elegant definition that captures the essence of this notion that works for most cases we are likely to encounter in our lives.

Therefore, we define *intelligence* as the *ability to adapt and prosper in a range of unfamiliar and non-trivial environments*. This is best summarized in Russell and Norvig's definitive textbook on the subject of AI: "[intelligence is] how to act effectively and safely in a wide variety of novel situations" [19].

Our definition is based on the *intersection* shared by the three case studies that we have considered in the previous section: it is clearly explicit in the case of Pavlov dogs but is also present in the next two studies. In these two hypothetical thought experiments, language is seen as the environment the agent finds itself in and has to act in order to demonstrate its intelligent nature. The perceived input of the environment is provided by the means of an incoming message to the agent, and the agent's produced output is given back also in the form of a written message. The *challenge* put forth by the environment is to convince an outside observer that the agent is intelligent by the means of language (this is also known as the *objective* of the agent).

2.4 Structural properties of Intelligence

In this subchapter, we would like to explore the structure of intelligent systems.

2.4.1 Collective intelligence

Given that our definition of intelligence implies the fact that the agent is located in a certain environment, it is also reasonable to assume that it may find other agents of its kin there as well. We can then assume that these agents will interact in one way or another: this interaction can take on many forms, which range from hostility (e.g., fighting over a shared finite resource) to collaboration and collective efforts towards a common goal. A combination of these polar ways of interaction is possible as well.

We can therefore claim that intelligence is an emergent property. A textbook example of the emergence of intelligence is given by ant colonies: the collective intelligence of a colony of ants greatly exceeds the sum of intelligence of all of its parts, let alone the intelligence of one specific ant.

In the field of AI, this aspect of intelligence is yet to be fully explored. Nevertheless, some recent developments in this field have taken inspiration from biological systems and the ways of interaction we have defined before.

One of these developments is Generative Adversarial Networks (GANs) that were proposed by Ian Goodfellow in 2014. Unlike other types of neural networks, GANs require a joint training of two models that play an adversarial game (see Figure 2). One of the models, called a *discriminator*, learns to distinguish between real data (images from a given database) and fake data (random computer-generated images). After learning the distribution of true data, the discriminator is linked with a *generator* that tries to fool it by providing fake data that tries to resemble its real counterpart. As learning continues, both networks get increasingly better at their respective objectives and the learning process can be finalized when the discriminator is unable to guess whether it is given true or fake data by the generator.

As opposed to the adversarial interaction seen in GANs, *Federated Learning* is another trend in modern AI that explores a collaborative way of interaction of neural networks. Instead of having a single AI that learns patterns from a given dataset, federated learning employs multiple AIs that are trying to solve the same problem. The reason for using this technique is two-fold: first of all, this way each model can work with a smaller dataset and will therefore be able to learn quicker. Second, and, perhaps, more importantly, the fact that each model sees less data makes the entire system more secure. Once all the models have been trained, their collective "findings" are transferred to a single model which finally merges their newly obtained knowledge.

2.4.2 Modular intelligence

A different aspect of the architecture of intelligence that we would like to consider is *modularity*. This property is closely linked to the one discussed in the previous subchapter, but on a different level: instead of looking at a group of intelligent systems, here we would like to explore the modules of a given singular intelligent agent.

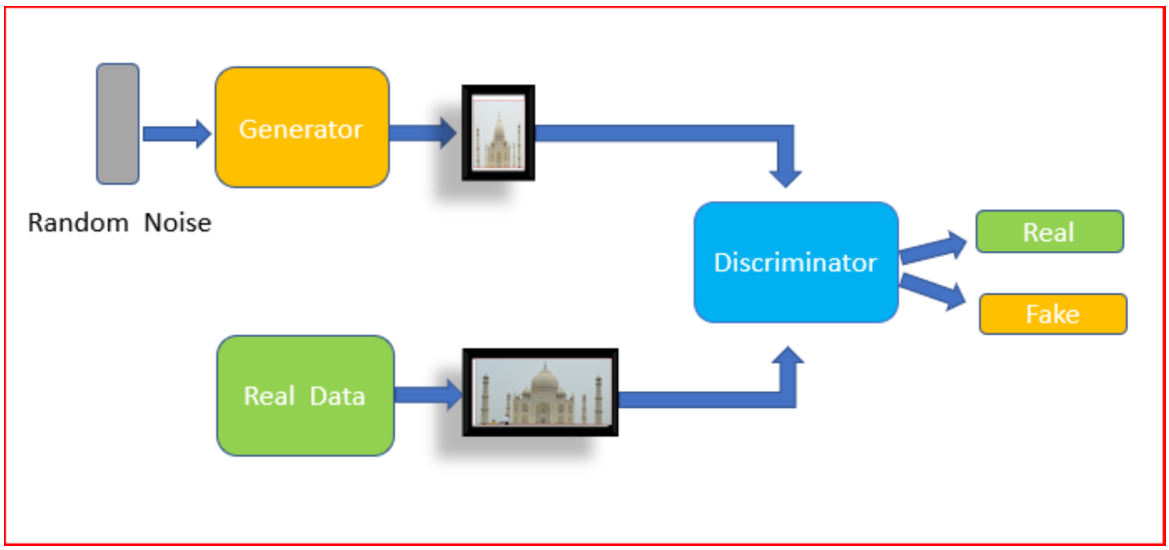

Figure 2: The architecture of a Generative Adversarial Network (GAN). The *generator* converts random noise into a fake image that resembles real images, while the *discriminator* learns to distinguish between real and fake images. As the two AIs continue to learn together, the knowledge of the real data gets transferred from the discriminator to the generator. Source: [28]

A classic example of the modular nature of intelligence can be seen in the human brain: research has shown that different regions of the brain excel at different tasks (e.g., the occipital lobe is responsible for visual processing). Even though the brain is capable of rewiring itself if a certain part of it is damaged (the responsibilities of a damaged module can be distributed among other modules), the presence of a hard-wired structure indicates that developing modularity proved to be evolutionary beneficial.

It is a common misconception that complex systems should be built out of complex components. Many areas of computer engineering have used this as a design principle (UNIX philosophy and encapsulation in OOP are some of the most famous ones), but it is most clearly seen in the Game of Life proposed by John Conway in 1970. In this *no-player* game, one can see an infinite grid consisting of empty (white) or filled (black) squares (see Figure 3). The only rules of the game are:

- If there are less than 3 or more than 3 neighbouring cells, the current cell dies (from loneliness / overpopulation).
- If there are exactly 3 neighbours, the cell remains alive.

It would seem that such a system cannot be expressive enough because of its simplistic structure. However, it was subsequently shown that this system is *Turing-complete*, which implies that it is capable of evaluating any computable function (in other words, it is as capable as any other computer). This tells us that being intelligent does not require complex components; it is the *connective structure* that matters.

2.4.3 Hierarchical intelligence

As opposed to the flat, egalitarian structure given by modularity, some aspects of intelligence exhibit a *hierarchical* structure.

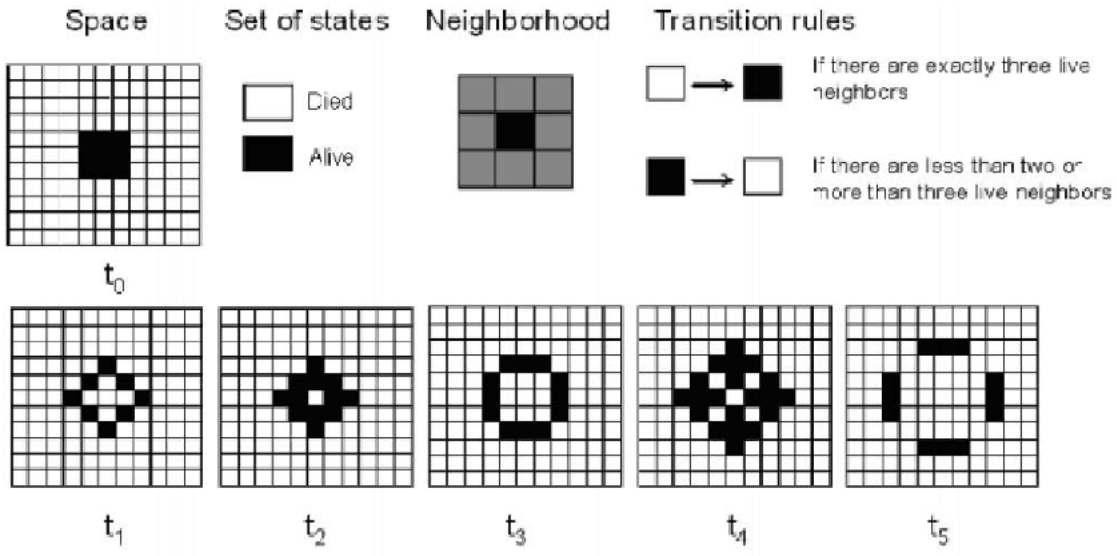

Figure 3: A brief overview of Conway's Game of Life. Source: [24]

Recent discoveries in neurobiology have shown that the structures of convolutional neural networks (AIs commonly used for working with images) are very similar to the structure of the visual cortex (the part of the human visual system that processes visual information). One key similarity that was pointed out is that both of these are structured *hierarchically*, that is, the input signal gets smaller and denser as it goes up the processing pipeline, and the result of the visual processing is a condensed form of the original input.

This aspect of intelligent systems is most clearly employed in wavelet processing, which decomposes a given signal into hierarchical components of different scales (see Figure 4).

Taking inspiration from the works of Herbert Simon, in his inaugural lecture at Collège de France, Stéphane Mallat remarks: *"Pour simplifier l'analyse de systèmes complexes, on étudie séparément les phénomènes qui apparaissent à des échelles très différentes... Cela permet d'expliciter la régularité des données et de calculer des représentations parcimonieuses."*[2] This shows that hierarchical structure is also beneficial for intelligence: it allows us to notice things that can't be seen by looking at the whole signal at once; it also shows that hierarchical systems are *parsimonious*: they remove the unnecessary parts and keep the ones that are significant.

2.5 Criteria of Intelligence

Given our definition and treatment of the architectural properties of intelligence, we can tackle other properties of intelligence that are worth discussing. These properties distinguish themselves as being more *criterion-like*: they are easy to evaluate and can be used for comparative analysis of intelligent systems.

[2]Rough translation: to simplify the analysis of complex systems, we divide them and look at the phenomena that appear at different *scales*. This allows us to explain the regularities of data and to calculate *parsimonious* representations.

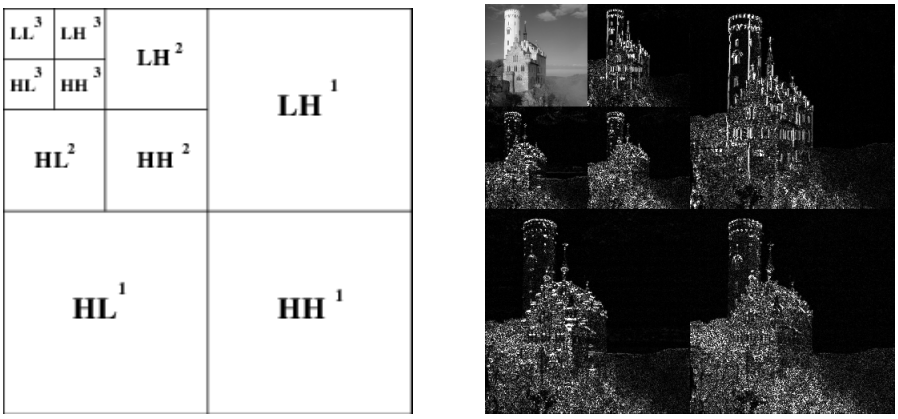

Figure 4: Left hand side: wavelet structure. Right hand side: wavelet pyramid of an image. Sources: [26] and [27].

- *Deductive reasoning.* As it was shown in the section on the Chinese Room experiment, deductive reasoning alone is not sufficient to prove that an agent is intelligent. However, the capacity of logical reasoning is central to intelligence: this was demonstrated in all of the case studies that we considered in the beginning of this chapter.

- *Concept formation.* The capacity of creating mental images (abstract representations) is crucial to intelligence. This was seen in the Chinese Room experiment: we said that John was not capable of speaking Chinese because he was manipulating flashcards without really understanding what he is talking about: on the other hand, if he could model his responses without consulting the cards and using *his own interpretation* of the conversation, he would be indisputably a Chinese speaker (in the Turing Test, for example, we expect that the correspondents don't use external aids like dictionaries).

- *Generalization.* Analogical problem solving was shown to be an integral part of intelligence in the Pavlov dog study. By learning to recognize familiar patterns and extrapolating this knowledge to other similar cases, we can thrive in novel, previously unknown environments, which is ultimately the way that we have defined intelligence.

2.5.1 Speed as a factor of intelligence

The issue of speed and efficiency requires a separate treatment. It is the basis of much research in the fields of AI, Information Theory and Data Science.

We propose two arguments in favour of viewing efficiency as a key factor in intelligent behaviour:

- Our first argument is based on Information Theory. If an agent is able to store information efficiently, it means that the agent *compresses* the information in a way that allows for easy access. As we know, compression consists in removing redundant parts of a given piece of information and keeping the ones that are pertinent to it. This implies that the agent is capable of knowing whether something is important or isn't; they are able to sacrifice small details of some information to store more general knowledge about something else; therefore they have to be intelligent in order to be efficient.

- The second way of proving the link between intelligence and efficiency can be provided by neurobiology. Imagine two gnus that are trying to outrun a lion in the wild. One gnu has an "efficient" brain, where the visual cortex that sees the lion is closely connected to the neurons that control the motion of muscles. The second gnu has a randomly configured brain with different brain modules connected in a disorganized and inefficient way. It is clear that the first gnu will be able to react to the appearance of the lion quicker, since its brain is wired in a way that is optimized for detecting predators and running away. The second gnu, however, will not be able to react quickly enough: the pathway between its visual cortex and muscle-controlling neurons will be far too long! As a result, the second gnu will end up on the lion's dinner table because of its inefficient brain structure, which shows that an efficient structuring of intelligence is evolutionary beneficial.

2.6 Conclusion

As we have seen, all of the definitions and criteria of intelligence that were presented above involve in some way or another the ability of *deliberate action* in the environment where the agent finds itself. This implies that our research goal and the related questions of the existence of Artificial General Intelligence (AGI) and modelling of human intelligence can only be viewed through the lens of *free will* and *decision making* which are considered in the following chapters.

3 Free will

In order to argue about the difference between humans and other living creatures or machines it is important to mention the concept of free will. This notion has been at the center of attention of Western philosophers for centuries, from Ancient Greek thinkers to modern philosophers.

Greek philosophers described freedom of the will as the capacity of a person to cultivate the virtues. They have seen it as a process of self-mastery by which a person's character is shaped. In the Medieval period, the theological approach to the freedom of the will has been developed. The will was seen as a self-determining power and the misuse of the freedom was considered as the source of evil. Free will was based on the metaphysical concept of the soul, which played a role of the agent free to choose between alternative options. With the development of the sciences, notably physics, the existence of non-physical entities has become the subject of doubt. Hence, modern philosophical schools have been roughly divided into two main branches: compatibilists and libertarians. Comptabilists claim that there is no conflict between determinism and free will. Free will fits into the deterministic world, in which everything is determined by the past events and the laws of nature. A person simply acts according to her/his strongest desire. Libertarians, on their turn, insist on the existence of a free non-physical agent able to choose between alternative options.

It is worth mentioning that the notion of free will serves as the base block of the modern legal systems. The question whether the defended was acting consciously and voluntarily while committing a crime is essential during the trials. Therefore, the question whether humans truly possess freedom of the will not only applies to philosophy, but also has social and political impacts.

However, before entering the debate about the existence of free will, it is necessary to define this notion in a clear way.

3.1 What is Free Will

Free will is often defined as the **freedom to do otherwise**. This is a very simple and intuitive definition of this notion. However, with further analysis of the definition multiple question may arise. What is meant by freedom? Why does it involve the actions ("to do"), whereas the term implies the desires ("will")? For instance, any animal can run in one direction or another and decide which direction it chooses. It can be concluded that animals also possess freedom of the will. However, free will is usually considered as exclusively human property.

In the previous example an animal can choose the direction it will run to, but eventually this animal does not have control over this decision, meaning that whatever direction the animal has chosen, there was no other option for it. The animal could not rationally evaluate the possible options and was driven solely by the instincts and reflexes which were the result of internal (animal's body) and external (nature around) events. Freedom of the will can also be defined as **control over one's actions**. As opposed to animals, humans seem to be able to reflect and evaluate the decisions they have taken or will take in future. This is surely the sign of "intelligent" behaviour, but is it also the sign of the presence of freedom of the will? Some animal species are also able to express such intelligent behaviour and seem to evaluate the possible options to make "smart" decisions. For instance, there exist multiple experiments in which bird or monkey species are solving logical tasks or play strategy games, which demand some sort of reflection. [9] Hence, it seems that animals/other living creatures also have desires, motives, they also make decisions. Again, two opposite conclusions can be made: either animals also have free will or the two definitions presented above are not accurate enough since only humans possess freedom of the will.

There are also other definitions of free will, relating to the **desires** ("to want") and not the actions ("to do"). For example, Harry Frankfurt introduces the concept of *second order desires* as the will to want something. Similarly, the *first order desires* are simply the desires to do or not to do something. "No animal other than man, however, appears to have the capacity for reflective self-evaluation that is manifested in the formation of second-order desire". In this framework the freedom of the will is defined as the freedom of willing to have or not to have a certain desire. According to Frankfurt, will, defined as the possession of the second order desires, is the structure which distinguishes humans from other species. To illustrate this concept, Frankfurt drives a thought experiment in which a psychotherapist treating the drug addicted patients may want to experience the need for a drug to find a better treatment. This clearly does not imply that the psychotherapist wants to experience the effect of the drug itself (this would be the *first order desire* in this case). Rather, she wants "to be inclined or moved to some extent to take the drug".[8] Thus, species other than humans are able to reflect in order to fulfill their desires (of the first order), but there is no evidence for them to have what is defined above as the second order desires. In other words, non human species do not care about their will and do not want to be moved or not to be moved towards certain desires.

3.2 Other approaches to the question of free will

3.2.1 WEIRD cultures

Does the term of "free will" exist in different cultures? If it does, is the construct of free will conceptually the same? It has been shown [3] that the concept of free will only exists in the Western (WEIRD: Western Educated Industrialized Rich Democratic) Christian cultures. There exist no identical constructs in Hinduistic, Buddhistic or Confucianistic philosophical traditions. For instance, in Hinduistic and Buddhistic writings the causal weight of past actions shapes the range of possible options one may have. From the perspective of the notion of free will, this can be seen as a kind of causal determinism. However, no terminology corresponding to the Western subject of free will was developed in the aforementioned cultures. In the Confucianistic philosophical tradition, the term "free will" appeared as the consequence of the Western influence.
This observation renders the subject of free will even more complex, since the concept of freedom of the will is not wold spread. Hence, other question arise. Can one have free will without consciously realizing it? Or is free will just the cultural construct developed in Western cultures over ages?

3.2.2 Neuroscience

Neuroscientists are generally sceptical about the concept of free will. Every action and thought a person makes or has, from a simple one (move the finger, take a breath) to a complex one (reflection, decision making) are governed by signals transmitted through neurons in one's brain. There is no signal transmitted without reason. This reasoning leaves no space for the concept of free will, since every action or thought one makes or has is the result of such interactions of neurons. One may argue that the freedom to do or will otherwise results from the laws quantum mechanics, which are governed by the probabilistic laws. However, quantum mechanics applies to the subatomic level and can hardly influence the interactions between neurons. Furthermore, if we assume that the functioning of neurons in humans' brain do follow the probabilistic laws, the presence of freedom or the choice still remains unclear. In other words, probability, or an uncertain event does not imply freedom of the will. If the decision one makes is the result of a random event that happened in the brain (for example, a neuron fires randomly), it cannot be concluded that this is caused by one's free will, since there was no control over this decision.

Scientists are still far from the entire understanding of the brain functioning. Although, in past decades many new discoveries have been made in this field. Brain structures involved in the motoric, as well as more complex functions such as emotions (sadness, fear, attraction, happiness) are determined. Multiple experiments show that people's judgements, emotions and moods are strongly influenced by environment. For example, several studies reveal the impact of odor on humans' emotions [11]. Emotions, in their turn, play an important role in decision making process, which is inseparably linked to the subject of free will. "[...] many psychological scientists now assume that emotions are, for better or worse, the dominant driver of most meaningful decisions in life". [12]

3.3 Conclusion

From what has been studied in this section, we have come to the conclusion that the notion of free will is very difficult to define. However, the opinions about it evolve over time. Back in the days when science was not yet developed, all the events were explained with mythological or theological approaches. As for

today, science has shed light on many processes, including human behaviour. It was empirically shown that our characters and decisions are strongly influenced by the internal factors (for example, genes) and external factors (environment, education, traumas, mood, emotions). This causes doubt about existence of the freedom of the choices. Probably, free will is just the belief which has arised and remained for historical and cultural reasons. Such a conclusion however would mean that multiple social institutions, such as legal systems have to be reshaped, since no criminal would be responsible for her or his crime.

4 Cybernetics and its impact on free wills and intelligence

In this section, I introduce the origins, basic principles and definitions of cybernetics and its impact on decision making and free wills from various aspects.

4.1 Cybernetics and Decision Making

As article [15] indicates, the current prevalent brain–machine interfaces were envisioned already in the 1940s by Norbert Wiener, the father of cybernetic. In virtue of examples of similar learning mechanism of missile navigation system and human emotion and similar communicative mechanism of hydropower station and human language, the pioneer regarded human intelligence emerging from interaction of feedback loops and expects that the development of machines should serve to augment human abilities.

From the top-level overview, some of the decision making processes can be simplified as closed-loop feedback control, where the commands are generated by either individual desires or wills mixed with unmodeled dynamics triggered by unknown high dimensional coupling between the second free wills and external influences from environments. The brief illustration that the intelligent agents interact with the environments to make the desires i.e., reference input come true is demonstrated in Fig. 5. The human beings actually enforce impact on the nature via modification of environments and building the artificial systems e.g., combustion engines, rockets to send humans to the moon, to facilitate evolution of civilisations and maintain prosperity of reciprocal societies comprised of the hierarchical and interconnected groups.

The three pillars of modern science and technology are the theoretical deduction, experiments observations and induction and computation. As far as the control procedures employed to modify the nature, we need the data collected to construct the models of the target, and then in virtue of the rough models, control policies are generated via computation in terms of the control laws derived from the theoretical deduction, then the actions guided by control policies are enforced on the targets and the feedback data are hereby employed to tune the performance of control policies and improve the precision of target models. Furthermore, to handle the unexpected phenomena we encountered, new theories regarding control laws are proposed to guide the design of control policies and learning methods so that they are continuously evolving to better describe and modify the configuration of objective world. It can be claimed as well that the desire of free will is similar to the reference input of the above closed-loop feedback system and the procedure of decision making can be automated with modifiable neural circuits built in our minds.

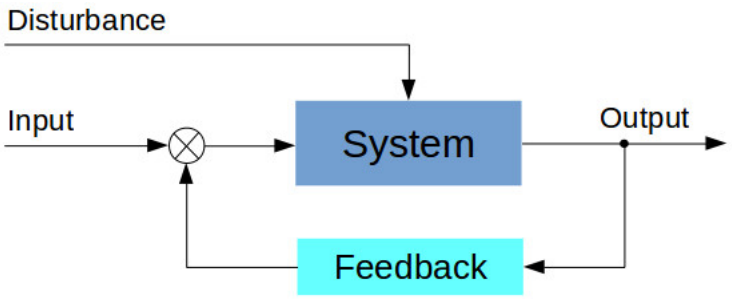

Figure 5: A brief overview of closed-loop feedback system

4.1.1 What Is Life

There is a serious hypothesis about the origin of life as follows. The living creatures depend on neg-entropy to maintain a kind of internal equilibrium and resist against global increasing entropy of the universe. The biological large molecules obtain energy through dissolving organic matter via complex biochemical reaction processes [20]. The first appeared life may be large organic molecules that resemble energy conversion factories driven by absorbed energy from external environment and self-assemble to generate gradually optimized next generations.

It can be claimed as well that human beings are pockets with decreasing entropy embedded in a framework in which general entropy tends to increase. Therefore the questions are left that whether the process of decision making depends on neg-entropy and what is the relation between free will and neg-entropy.

4.1.2 Good Regulator Theory

The good regulator theorem reveals the general scheme of interaction between intelligent individuals and external environments as an extension to basic closed-loop feedback control system. The pioneer of cybernetics first put forward good regulator theorem that any regulator that is maximally both successful and simple must be isomorphic with the system being regulated. As illustrated in Fig. 6, the interaction of intelligent agents and external environments can be modelled by five specified variables identified in the whole process they play in, i.e., the total set Z of events that may occur, the regulated and the unregulated, the set G, a sub-set of Z, consisting of the 'good' events, those ensured by effective regulation, the set R of events in the regulator H, the set S of events in the rest of the systems, and the set D of primary disturbers that tend to drive the outcomes out of G by causing the events in the system S. It is further derived that the living brain as an efficient regulator for survival must processed in learning by the formation of model of environment. Modelling should be necessary part for regulation, and the successful regulation is defined as minimization of outcome entropy [6].

In terms of efficiency, the error-controlled regulation is primitive compared to the cause-controlled regulation which is direct regulation mapped from external disturbance as source and determiner of regulatory actions in higher organisms. Thus the best regulator of a system is one which is a model of that system in the sense that the regulator's actions can be merely regarded as a mapping from system to regulator.

The creatures are always regulating their actions to maximize the ability of adaptation through modelling of external environments, which can be regarded as built-in biological regulatory circuits for decision

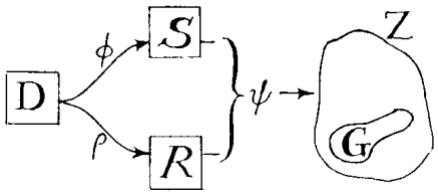

Figure 6: Modelling of interaction between intelligent agents and external environments

making to optimize the benefits. Therefore a question is left that what is the role of free will in the evolutionary development of the above good regulators that seem to be generated automatically.

4.1.3 Internal Model Principle

Inspired by good regulator theorem, internal model principle is introduced that a good controller should incorporate a model of the dynamics that generate the signals tracked by controller. In other words, the controller should contain information and structure of the outside world. These decades, this principle has been developed and applied into specialized frameworks of control theory such as linear multivariable systems.

A general representation of the internal model control. The design concept of the internal model control is to compensate for errors or make adjustments from the desired outputs. Thus, two primary parts of the control algorithm are the forward model and feedback loop used to enhance the control performance. The article [10] utilizes internal models for motor control and trajectory planning, and the generalization

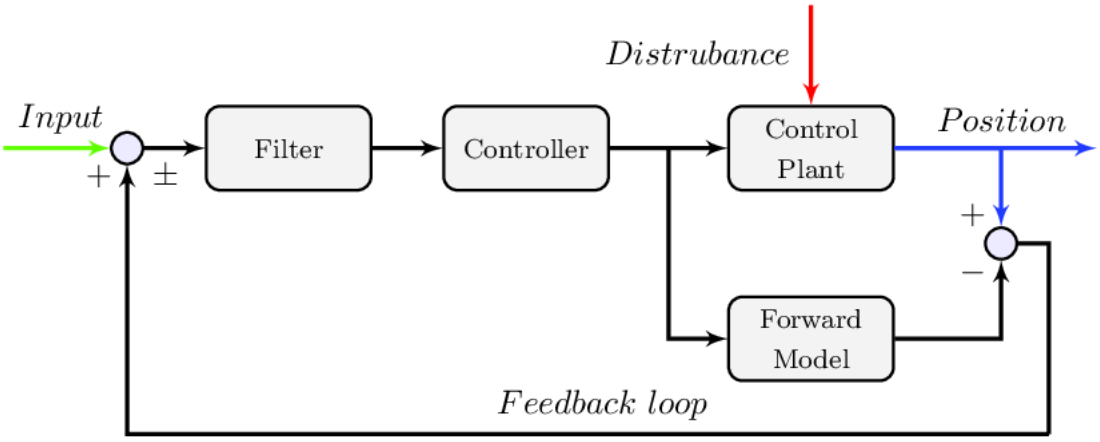

Figure 7: A brief overview of internal model principle based system

of sensor-motor integrated internal model is comprised of incorporating internal models and contextual information, training from a large set of data and selection of desired trajectory model.

4.1.4 Sensor-Motor Integration for Interacting with External Environment

The article [21] emphasized the importance of sensor-motor integration, that is for sensori-motor systems in higher animals and complex systems, the information perceived and responses sensed from external

environment are coordinated to central pattern generator to guide the generation of a broad spectrum of behaviors.

At the same time, there also exists an underlying self-organizing bridge linking perception and motion, that is the sensory information facilitates selection of efficacious behaviors from chaos. Therefore, it can be claimed that free will represents self-organizing structures emergent from chaos and external disturbs while it can maintain stable and nearly the same behaviors under different conditions. We assume that the modelling of external environments from good regulators reserves the essential patterns of external environments so as to play an important role in the above process.

Feedforward control can be regarded as an instance of good regulator directly employing the models of external environments and has advantage of fast response and well robustness compared to feedback closed-loop control scheme. Feedback and feedforward control using an inverse model of a controlled object are illustrated in diagram a and b of Fig. 8. In feedback control shown in a, the realized trajectory is compared with the desired trajectory, and the error is computed. The feedback motor command is generated from this error using a relatively simple algorithm, such as a proportional, integral and derivative feedback controller. In robotics, almost all practical applications depend solely on feedback controls. This is because feedback delays in artificial systems can be made small; hence, sampling and control frequencies can be quite high (from 500–10000 Hz). In biological motor control, however, the delay is very large. For visual feedback on arm movements, the delay ranges from 150–250 ms. Relatively fast spinal feedback loops still require 30–50 ms time delays. These are very large compared with the movement duration of very fast (150 ms) to intermediate (500 ms) movements. In diagram b, inverse dynamics model(internal model) is serially-connected with the controlled object, the serial system gives a mathematical identity function. That is, the output (i.e. the realized trajectory) is identical to the input (i.e. the desired trajectory). Thus, the inverse model, if it exists and can be learned, becomes an ideal feed-forward controller. In biological systems with large feedback delays and small feedback gains, internal models are the only computational possibility for fast and well coordinated movements. And an instance of ball grasping experiment is shown in Fig. 11. We can see that though there exists changes in the configuration of the tasks,

4.1.5 Generalized Energy for Decision Making

Furthermore, the article [16] discussed an active pathway for intelligent agents to make decision instead of either passively perceiving contextual information or been embedded internal model containing structure of external world. Active inference is introduced as an approach to understanding behaviour that rests upon the idea that the brain uses an internal generative model to predict incoming sensory data.

The variation free energy functions for action and perception and including states, policies and observations are designed to score the fits between an internal model and the external world. This formulation ensures that selected policies can minimise uncertainty about future sensory data by minimising the free energy expected in the future. It can be claimed that in some sense free will is an active exploration mechanism like generalized energy to minimize the difference between the expected perception and the reference of internal dynamics like desires.

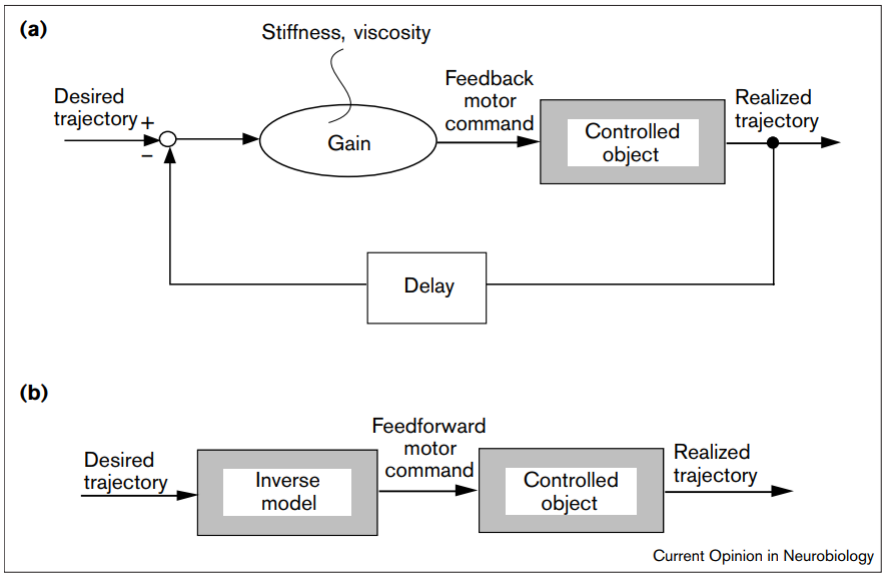

Figure 8: A brief diagram of two types of motor control

4.2 Human Use of The Human Beings

In the current digital era, we can observe that machines are gradually replacing humans on many occupations. So in which aspects do human beings and machines share plenty of similarities? In book [22], Norbert Wiener considered both living creatures and machines in probabilistic point of view instead of mechanics. In virtue of thought experiment of the Maxwell demon controlling particles filter gates in an isolated system, he inferred that only if the light particles and material particles are not in equilibrium, the velocity and position information can be obtained by Maxwell demon, and thus the entropy could decrease locally. Similarly, creatures exploit mutual messages as well purposively to fight against disorder introduced by increasing entropy.

As far as the life-imitating automata, he introduced three major components:

- the organs as actuators to impose influence to outer world.
- the ability to sense external world such as photoelectric and thermometer.
- feedback function, which can adjust future actions via past performance.

and feedback can be classified as common reflex regulating specific movements and conditioned reflex as learning, i.e., the process changing the whole policies of behaviors.

He claimed that the nervous systems and automatic machines are alike since they can make decisions on basis of the past decisions.

4.2.1 Modelling and Simulation of Brains

Brains are regarded as the origins of our thoughts and creativity. So can we build a generalized artificial intelligence from scratch? Which principles should we conform to construct such automatic systems? W.

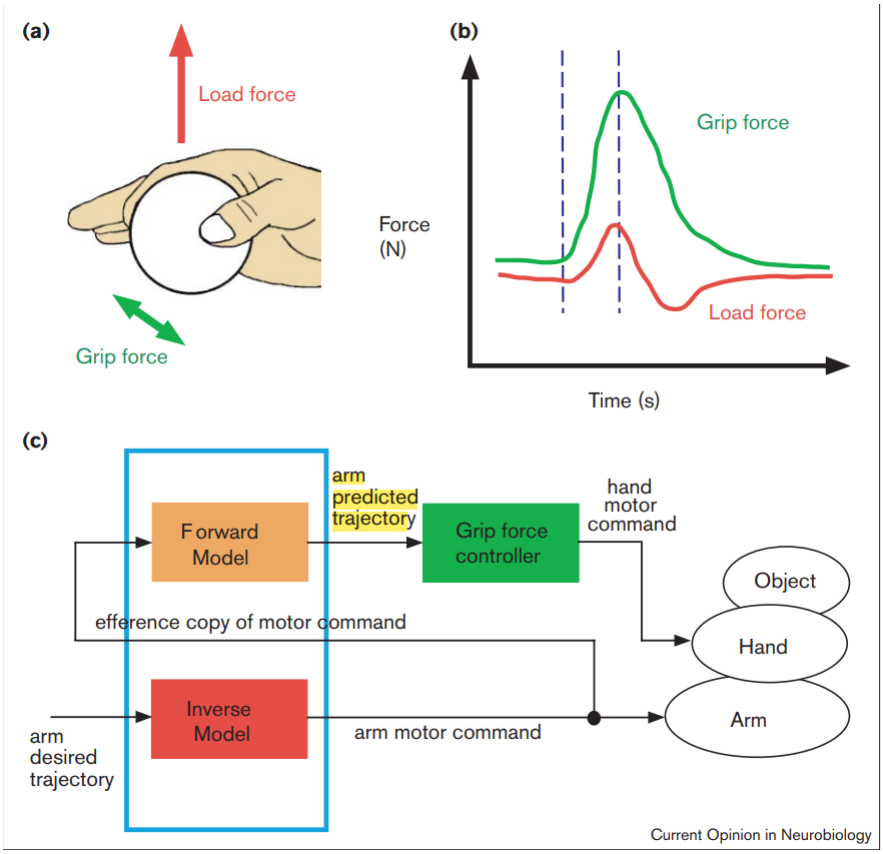

Figure 9: An illustration of sensory motor control exemplified by balls grasping experiments

Ross. Ashby researched in details the relationship between living organism and automatic machines in "The Design of Brain" and his work named Homeostat like living creatures can locally and temporarily resist the general increase of entropy of external environment. [1]

On topic of general learning process, Wiener held view point that the unpurposeful random mechanism seeks for its own purpose through a process of learning. And the memory and learning are continuous organization, allowing alterations induced by outer impression to be changed as more or less permanent structure and function. [22]

Cybernetics take the view that the structure of machine or organism can be regarded as an index of performance expected from it. just like the rigid shell limits intelligence of insects while mechanical fluidity structure of human beings allows indefinite intellectual expansion of human beings, thus Wiener claimed that theoretically, machine with similar intelligence features as humans can be built with duplicated structure as human beings. [22]

EPFL has maintained "Blue Brain Project" to explore and simulate functions and activities of brain varying from cellular neocortical to mammalian brain like mouse, which is shown in Fig. 10. As indicated in [14], these days there are four biggest challenges for simulation and modelling brains

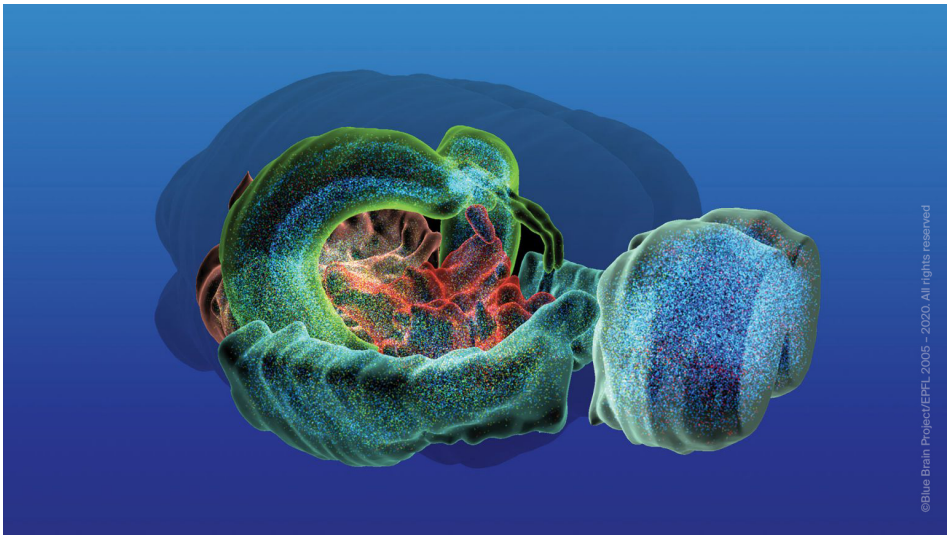

Figure 10: An illustration of Blue Brain Project of EPFL

- With one hundred billion neurons and one thousand trillion synapses working in parallel, simulating the human brain would push the limits of even the exascale computers.
- Producing a biologically faithful simulation of the brain would require an almost limitless set of parameters.
- No present technology can run large-scale simulations faster than in real-time.
- To model functions that involve brain-wide networks, top-down models of brain regions and bottom-up biophysical models will need to be combined. .

4.3 Sociology: Simulation and Simulacra

On the one hand, we humans are so particularly sophisticated and talented for self-control, reasoning, reflective thoughts etc, which makes us quite distinctive from other animals. But on the other hand we were thus able to build super computers and other apparatus to construct a totally virtual world. And now we come to a point where our technical construction gain their own control and authority and start to control parts of us and thus reduce actually our distinctiveness compared to other animals and degree of freedom. With gradually matured human machine interface and augmentation of perception and signal processing of machines, these days one dark pathway of trans-humanism gradually come true, and Wiener has worried that machine could eventually be used to control humans and to displace jobs instead of being used to augment human abilities. [22]. We take the following cases for examples.

4.3.1 The Social Media Impacts on People

The innovation of telegraph apparatus represents that the intelligence can be communicated remotely and its importance is on a par with the design of internal combustion engines. Everything is connected to build an information networks, virtual identities are gradually labelled on each authentic human. On the one hand, the internet facilitates people's communications, on the other hand, we are somewhat

becoming slaves of it. And it's thus no longer clear how to evaluate the net gain of it. For instance, it is actively attempting to infiltrate into offline lives of the masses. We can observe a trend that in USA the average time spent daily on digital social media has increase from 2008 to 2018 and the proportion of mobile channel among digital media has skyrocketed.

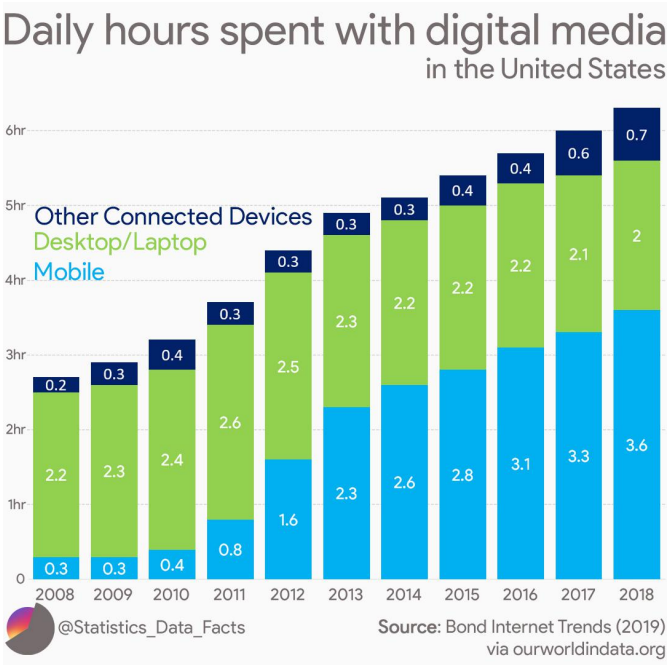

Figure 11: An illustration of increasing usage of digital social media in USA

4.3.2 The Large-scale Recommendation System and Attention Engineering

Michael Zeiler's famous pecking pigeon experiments showed that rewards delivered unpredictably are far more enticing than those delivered with a known pattern. And it has been verified that unpredictability releases more dopamine, a key neurotransmitter for regulating our sense of craving. [4]. There are plenty of similar mechanisms in large social media such as Facebook and Twitter, in which "like" is just like the unpredictable incentive entice users to be gradually addicted to these platforms.

We take YouTube for example to illustrate that how people's solitude are deprivation by attention engineering facilitated by intelligent recommendation systems. The system can be split into two parts: deep candidate generation networks driven by big data collected in forms of user habits and advanced algorithm and deep ranking networks to recommend prioritized content to users. And the final block is to adjust parameters based on several performance metrics through live testing. The brief diagram of large-scale recommendation system is shown in Fig. 12.

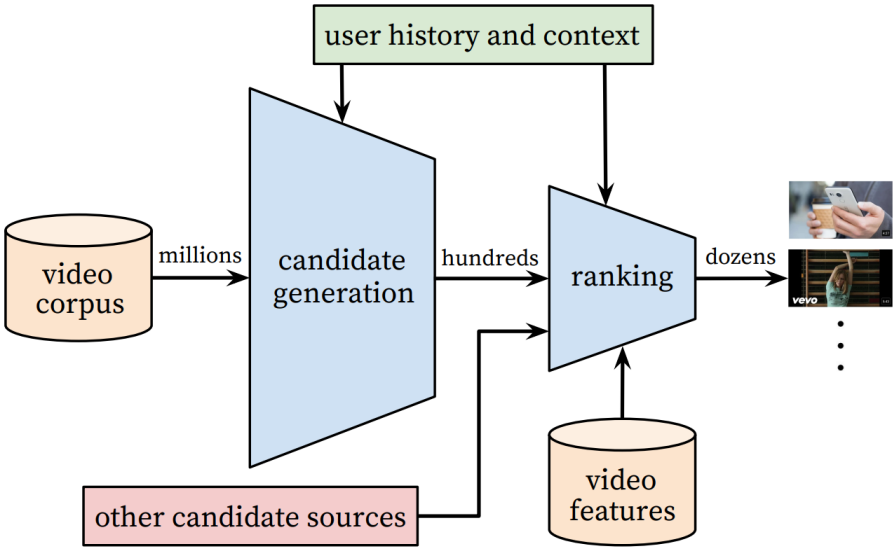

Figure 12: A brief diagram of large-scale recommendation system

4.4 Conclusion

As we can see, cybernetics shares a lot of concepts with free will and decision making. We first analyze the basic closed-loop feedback scheme and internal model principle derived from good regulator theorem and further extend to sensor-motor integrated systems and generalized energy frames, which entails that during constant interaction with external environments, intelligent individuals are inclined to embed world model and generate patterns emergent from external chaos and disturbs while eventually stay stable and robust against external changes. Then we dive into the more microscopic scales that biological large molecules that have the same basic functions as complex creatures can be regarded as the first life and these days, with development of modern digital technology, we are constructing machines that can carry out similar tasks than humans and managing to build generalized artificial intelligence like simulated brains. However, as already predicted by several pioneers of cybernetics, the sophisticated technologies that make machines more and more alike and even replace humans are gradually out of control from us, the delicate control strategies and schemes are backwards enforced on our lives, which leads to a dark path of human society and eliminate distinctiveness of humans compared to machines and other animals. Everyone ought to contemplate how to control the intelligent devices and schemes instead of being manipulated by them in the digital era.

5 Conclusions

Both human and non human species or machines exhibit some sort of intelligent behaviour and are able to make decisions based on the prior thought. Our paper has shown that there are clearly some similarities between the behavioural patterns of human and non human agents. Animals are able to form concepts and solve logical problems. Modern AI technologies involve structures similar to ones engaged in human brain.

To study these similarities it is important to clearly define the aforementioned notions. However, it has

been discovered that both of them are very semantics- and context-dependant. The concepts discussed in our paper are incredibly complex and we therefore have to restrain ourselves to models regrouping the essential (in our opinion) properties in order to be able to analyse them. Yet these models may lack important aspects of intelligence similar to human.

On the other hand, it is hard to deny the fact that modern technologies allow to reproduce the behaviour that becomes more and more similar to human. What was considered impossible in the early stages of the AI field does not seem so unachievable anymore.

References

[1] Ashby, W., 1970. Design For A Brain. [Lieu de publication inconnu]: Chapman and Hall.

[2] Baudrillard, J., 1994. Simulacra And Simulation. Ann Arbor: University of Michigan Press.

[3] Berniūnas, R., Beinorius, A., Dranseika, V., Silius, V. and Rimkevičius, P., 2020. The weirdness of belief in free will. Consciousness and Cognition.

[4] Bromberg-Martin, E. and Hikosaka, O., 2009. Midbrain Dopamine Neurons Signal Preference for Advance Information about Upcoming Rewards. Neuron, 63(1), pp.119-126.

[5] Chollet, François, 2019. On the Measure of Intelligence. arXiv reference: 1911.01547 (cs.AI).

[6] Conant, R. and Ross Ashby, W., 1970. Every good regulator of a system must be a model of that system †. International Journal of Systems Science, 1(2), pp.89-97.

[7] Covington, P., Adams, J. and Sargin, E., 2016. Deep Neural Networks for YouTube Recommendations. Proceedings of the 10th ACM Conference on Recommender Systems,.

[8] Frankfurt, H., 1971. Freedom of the Will and the Concept of a Person. The Journal of Philosophy, 68(1), p.5.

[9] Heinrich, B., & Bugnyar, T. (2005). Testing Problem Solving in Ravens: String-Pulling to Reach Food. Ethology, 111(10), 962–976.

[10] Kawato, M., 1999. Internal models for motor control and trajectory planning. Current Opinion in Neurobiology, 9(6), pp.718-727.

[11] Kontaris, I., East, B. and Wilson, D., 2020. Behavioral and Neurobiological Convergence of Odor, Mood and Emotion: A Review.

[12] Lerner, J. S., Li, Y., Valdesolo, P., Kassam, K. S. (2015). Emotion and Decision Making. Annual Review of Psychology, 66(1), 799–823.

[13] Lewis, Rick, 2007. All The World's A Text? [online] Available at: https://philosophynow.org/issues/60/All_The_Worlds_A_Text.

[14] Makin, S., 2019. The four biggest challenges in brain simulation. Nature, 571(7766), pp.S9-S9.

[15] Nature Machine Intelligence, 2019. Return of cybernetics. 1(9), pp.385-385.

[16] Parr, T. and Friston, K., 2019. Generalised free energy and active inference. Biological Cybernetics, 113(5-6), pp.495-513.

[17] Plato.stanford.edu. 2020. Free Will (Stanford Encyclopedia Of Philosophy). [online] Available at: https://plato.stanford.edu/entries/freewill/.

[18] Plato.stanford.edu. 2020. Cognitive Science (Stanford Encyclopedia Of Philosophy). [online] Available at: https://plato.stanford.edu/entries/cognitive-science/.

[19] Russell, S. and Norvig, P., 2020. Artificial Intelligence. 4th ed.

[20] Schrodinger, E. and Penrose, R. (2012) What is Life?: With Mind and Matter and Autobiographical Sketches. Cambridge: Cambridge University Press (Canto Classics)

[21] Steingrube, S., Timme, M., Wörgötter, F. and Manoonpong, P., 2010. Self-organized adaptation of a simple neural circuit enables complex robot behaviour. Nature Physics, 6(3), pp.224-230.

[22] Wiener, N., 1968. The Human Use Of Human Beings. London, U.K.: Sphere Books.

[23] Brighton, Henry and Selina, Howard, 2015. Introducing Artificial Intelligence. London, U.K.: Icon Books Ltd.

[24] Wahyudi, Agung & Liu, Yan. (2015). Spatial Dynamic Models for Inclusive Cities: a Brief Concept of Cellular Automata (CA) and Agent-Based Model (ABM). Jurnal Perencanaan Wilayah dan Kota. 26. 54-70. 10.5614/jpwk.2015.26.1.6.

[25] Plato.stanford.edu. 2020. Artificial Intelligence (Stanford Encyclopedia Of Philosophy). [online] Available at: https://plato.stanford.edu/entries/artificial-intelligence/.

[26] Zhang, Zhong & Blum, Rick. (04.2003.) Region-Based Image Fusion Scheme For Concealed Weapon Detection.

[27] JPEG2000 2-level wavelet transform. Alessio Damato, CC BY-SA 3.0, via Wikimedia Commons. 2007. [online] Available at: `https://commons.wikimedia.org/wiki/File:Jpeg2000_2-level_wavelet_transform-lichtenstein.png`.

[28] Generative Adversarial Network (GAN) using Keras. Renu Khandelwal, 2019. [online] Available at: `https://medium.datadriveninvestor.com/generative-adversarial-network-gan-using-keras`