Denaturalizing Artificial Intelligence in Ted Chiang's The Lifecycle of Software Objects

Grgurević, Dario

Sveučilište u Splitu

Filozofski fakultet

Odsjek za engleski jezik i književnost

Dario Grgurević

Denaturalizing Artificial Intelligence in Ted Chiang's The Lifecycle of Software Objects

Diplomski rad

Split, 2021.

University of Split

Faculty of Humanities and Social Sciences

Department of English Language and Literature

Denaturalizing Artificial Intelligence in Ted Chiang's The Lifecycle of Software Objects

MA Thesis

Student:

Dario Grgurević

Supervisor:

Dr Brian Willems, Assoc. Prof.

Split, 2021.

Content

1. INTRODUCTION

Ted Chiang is well known for his short stories belonging to the genre of science fiction. One such story, *The Lifecycle of Software Objects*, is at the center of this paper. Chiang's novella revolves around two human protagonists, Ana and Derek as they are working on superintelligent digital animals, known as digients. Superintelligence serves as the foundation of the digients' research and development and as such it is discussed by theorists such as Nick Bostrom, Steven Shaviro, Akira Mizuta Lippit, Michael Kearns, Aaron Roth and others. Each of them analyzes an aspect of superintelligence or its inception which can be connected to Chiang's own story. In *Superintelligence*, Nick Bostrom aims to give examples, procedures, requirements and warnings for obtaining and creating superintelligence. His analysis and research covers not only artificial intelligence, i.e. software exceeding human intelligence, but also human brain enhancements which should lead to advanced intelligence in everyday people. Another important aspect of Bostrom's book is the pacing of research and eventual result, but first we need to understand the terms in order to analyze them. Superintelligence serves as the main theme of the first half of the novella, but not the second half. The narrative changes the focus from obtaining superintelligence to preserving the means of obtaining it. The event crucial for this transition is the fall of Blue Gamma, the company employing Ana and Derek and maintaining the existence of digients. This paper will focus on the narrative and its evolution, how science fiction writers until Chiang took more drastic steps in order to advance their stories. The main thesis of my paper is that Ted Chiang's *The Lifecycle of Software Objects* demystifies artificial intelligence, showing real world elements contributing to the creation of digients. They are the result of plausible research, trial and error as well as interaction with humans and non-humans alike.

2. A BRIEF BACKGROUND OF ARTIFICIAL INTELLIGENCE

This chapter will aim to provide a brief outline of the history of AI in fiction by using John Sladek's novel *Roderick* (1980) and other authors in order to provide the background for further discussion on AI. This chapter will later use the following examples in relation to public's perception on AI research and achieving superintelligence. Since these are products of narratives, fictional stories, some aspects of AI use have been lost in time. In spite of this, these AI narratives do not go the extra mile in creating the intelligence Nick Bostrom defined. Bostrom is one of the key theorists in this paper whose definitions of intelligence and superintelligence play an important part in demystifying several aspects of artificial intelligence.

Sladek's examples are very limited automata, artificial objects with the ability to move. Their intelligence is limited to listening and obeying their masters' commands. What started out as a servant whose actions were limited to a couple of commands, came to be a blueprint for imagining something more in the world of AI fiction and non-fiction literature. This example of Sladek's novel was covered by Cave, Dihal and Dillon in the introduction of their book *AI Narratives: A History of Imaginative Thinking about Intelligent Machines.*

John Sladek opted for recounting the times non-human beings were created. His science fiction novel explores general intelligence in a multitude of AI narratives. The story follows the titular character Roderick, the first intelligent machine, and his progress from a bodiless voice to a robot living and experiencing the real world. Sladek's human characters start listing instances of artificial intelligence throught history via stories and myths. Those stories are the backdrop against which his own AI narrative was created. He goes on to list the creators and their creations throught the history:

> the Blackfeet boy, Kut-o-yis, cooked to life in a cooking pot, but isn't that the point? Aren't they always fodder for our desires? Take Pumiyathon for instance, going to bed with his ivory creation [. . .] take Hephaestus then, those golden girls he made who could talk, help him forge, who knows what else . . . Or Daedalus, not just the statues that guarded the labyrinth, but the dolls he made for the daughters of

> Cocalus, you see? Love, work, conversation, guard duty, baby, plaything, of course they used them to replace people, isn't that the point? [. . .] And in Boeotia, the little Daedala, the procession where they carried an oaken bride to the river, much like the *argeioi* in Rome, the puppets the Vestal Virgins threw into the Tiber to purge the demons; disease, probably, just as the Ewe made clay figures to draw off the spirit of the smallpox, so did the Baganda, they buried the figures under roads and the first [. . .] person who passed by picked up the sickness. In Borneo they drew sickness into wooden images, so did the Dyaks [. . .] Of course the Chinese mostly made toys, a jade automaton in the Fourth Century but much earlier even the first Han Emperor had a little mechanical orchestra [. . .] but the Japanese, Prince Kaya was it? Yes, made a wooden figure that held a big bowl, it helped the people water their rice paddies during the drought. (Sladek 60-61)

Sladek's character Dr Jane Hannah goes on to list a multitude of examples of AI in literary history. From Hephaestus and Daedalus to the Chinese and Prince Kaya, artificial intelligence was created by man to serve man. Although the reasons seem different, Hephaestus, Daedalus and the rest of heroes from the myths and legends have designed seemingly perfect automated machines. Their creations serve purposes, according to Dr Jane Hannah, known and unknown.

They conclude that "Sladek's novel understands the tradition it lies within, that is, a transhistorical, transcultural imaginative history of intelligent machines. These imaginings occur in a diverse range of narrative forms, in myths, legends, apocryphal stories, rumours, fiction, and nonfiction (particularly of the more speculative kind)" (Cave, Dihal and Dillon 4).

Even though the examples of artificial intelligence come from myths and legends, the narrative paved the way for how the public often perceives real world developments in that field.

2.1. The real-life impact of AI narratives

According to Cave, Dihal, and Dillon, Sladek understands the tradition on the basis of which his own story is based. Hephaestus did not want his servants to develop any further than helping with manual labor, but that does not mean they could not have done more.

The example of Hephaestus and his creations was also covered by The Royal Society in *Portrayals and Perceptions of AI and Why They Matter*, stating that "the machines were 'attendants made of gold, which seemed like living maidens. In their hearts there is

intelligence, and they have voice and vigour.' They appear as faithful servants to their crippled master" (The Royal Society 7). The Royal Society seems to romanticize this example. They make out Hephaestus' golden maidens to be something they are not. It is as if they are immortal girls singing and dancing while also helping their creator. In reality, they are far from it. It would seem as if they are some magical creatures, made with special tools and given intelligence out of thin air. Again, since this is a myth, readers cannot be expected to believe everything they read, but a multitude of similar stories does its toll on the public's perception. The nature of a good narrative is a compelling premise, inventive characters and storytelling. AI narratives, unfortunately, seem only to create anxiety in the everyday reader when confronted with real world progress in that field.

Portrayals and Perceptions of AI and Why They Matter goes on to say that "these stories end badly, with mishaps and the destruction of the oracle, sometimes by a terrified layperson. The moral is that the creation of AI is an act of Promethean hubris; that such divine power should not belong to mortals" (Ibid 7). Real-world progress seems to suffer because of fictional stories involving catastrophic scenarios. The notion of AI being a Promethean hubris serves only to distance the general public from such innovations. If the divine power and creation such as an artificial intelligence is only to belong to gods, then humans should never attempt to build such things.

Ian Bogost also tackles the public's perception of algorithms and computers as divine beings. His article "The Cathedral of Computation" takes examples of world famous companies like Google, Facebook and Netflix to illustrate how the average consumer sees their service. A person's perception of Google's algorithms recommending content on its own, personalising ads like it can read minds, alienates and misrepresents how such software works. In Google's case, Bogost call it "a monstrosity. It's a confluence of physical, virtual, computational, and non-computational stuffs—electricity, data centers, servers, air

conditioners, security guards, financial markets—just like the rubber ducky is a confluence of vinyl plastic, injection molding, the hands and labor of Chinese workers, the diesel fuel of ships and trains and trucks, the steel of shipping containers" (TheAtlantic.com). Google's PageRank search engine is constantly being influenced by internal and external factors. There is not a particular programmer who succeeded in designing an omnipresent software. The engine depends on the company that owns it, the guards protecting the building, the electricity powering everything and employees constantly checking how everything is working.

> There is no need to attribute more power or knowledge to an AI than it truly has.
>
> This attitude blinds us in two ways. First, it allows us to chalk up any kind of computational social change as pre-determined and inevitable. It gives us an excuse not to intervene in the social shifts wrought by big corporations like Google or Facebook or their kindred, to see their outcomes as beyond our influence. Second, it makes us forget that particular computational systems are abstractions, caricatures of the world, one perspective among many. The first error turns computers into gods, the second treats their outputs as scripture. (Ibid)

The average consumer sees information posted on Facebook as absolute because they choose to accept it as such. People forget that those companies provide nothing more than a service, a means to achieve something. There is no God hiding in Mark Zuckerberg's server room, only billions and billions of little machines sharing information influenced by the public and fed to the public.

Cave, Dihan, and Dillon explain this in a simple manner, saying "Narratives of intelligent machines matter because they form the backdrop against which AI systems are being developed, and against which these developments are interpreted and assessed." (Cave, Dihal, Dillon 7). Fictional achievements and resulting tragedies have invaded the real world in

such a way that researchers do not get a chance to explore their field of expertise. Bad scenarios are more frequent in everyday media that good ones.

> Their book provides another example of AI narratives impeding real research,
>
> The UK House of Lords Select Committee on Artificial Intelligence opens the second chapter of their 2018 report 'AI in the UK: ready, willing and able?' with a sharp critique of prevalent AI narratives: The representation of artificial intelligence in popular culture is lightyears away from the often more complex and mundane reality. Based on representations in popular culture and the media, the non-specialist would be forgiven for picturing AI as a humanoid robot (with or without murderous intentions), or at the very least a highly intelligent, disembodied voice able to assist seamlessly with a range of tasks. (Ibid 8)

The report quoted clearly states how popular culture has nothing to do with real research in the field of technology and artificial intelligence. A humanoid robot, a digital animal or a highly intelligent disembodied voice are currently nowhere near inception, much less completion. The fact remains that the public is bombarded with examples of destructive AI, resulting in a negative perception towards real people working on achieving what Nick Bostrom depicts as artificial intelligence. The final step is to one day produce superintelligence, which Bostrom also defines later on. Nevertheless, it is vital to define every step of research into artificial intelligence if we are to avoid any confusion. Ted Chiang helps readers understand how digient related research takes years to in order to advance. The people involved work tirelessly to teach digients simple things, how to speak or write. Understanding the work being done and effort his human characters make leads to a more realistic perception of how their AI works.

2.2. Superintelligence

This section will mainly refer to Bostrom in order to define and explain superintelligence. His examples are simple to understand which is vital if we are to discuss anything remotely complex. His book defines the terms AI and superintelligence by providing concrete examples. To better understand what artificial intelligence is, we will take a look at an example of AI described by Bostrom. A paperclip AI which is "designed to manage production in a factory, is given the final goal of maximizing the manufacture of paperclips, and proceeds by converting first the Earth and then increasingly large chunks of the observable universe into paperclips" (Bostrom 149). Building an AI with a relatively simple goal in mind serves as an example of applying general intelligence to real life situations and problems. If an AI cannot maintain a steady paperclip production, how can we expect it to solve the mysteries of the universe? In this case, the first step in AI evolution is manufacturing paperclips. It is placed in charge of a factory with that goal in mind.

This serves as a simple illustration of an AI's starting point which could unfortunately lead to a superintelligent paperclip-making entity. Its goal remains unchanged, making as many paperclips as possible and destroying the planet as the result. It is crucial to program and provide a detailed goal for the artificial intelligence to achieve. The first step is to manage paperclip production and then it can move on to bigger things.

In the second chapter "Paths to Superintelligence," Bostrom says that "We can tentatively define a superintelligence as any intellect that greatly exceeds the cognitive performance of humans in virtually all domains of interest" (Ibid 39). It is essentially the next step in cognitive evolution where humankind's level of intelligence is but a speck of dust in what Bostrom sees as basic level of superintelligence. The basic cognitive level of a superintelligent being will exceed even the smartest human and it will excel in all areas of cognitive performance. It is important to remember this is just a general definition Bostrom

provides to the readers. As such it can be seen more as an end result of research into superintelligence. However, we should not look at this development through rose-colored glasses. The outcome of creating superintelligence will more likely result in an extremely good or extremely bad outcome. The risks are often not taken into consideration when talking about this kind of research.

A superintelligent being can easily change the world order if it willed so, establishing itself at the top of the food chain. Being able to think and act outside the box of human-level intelligence will give it an advantage over every other being on Earth. Another great risk is posed by the possibility of replicating superintelligence which would lead to the extinction of human-level intelligence. The dangers superintelligence could pose are endless, but it is important not to become overconfident and allow it to develop in an unwanted way.

For Bostrom it is important to first attain general intelligence before moving to anything more than that. General intelligence revolves around learning and adapting, coming into conflict with problems in everyday situations and dealing with them. Throughout his book, but here especially, Bostrom emphasizes a system's capacity to learn. For him it

> …would be an integral feature of the core design of a system intended to attain general intelligence, not something to be tacked on later as an extension or an afterthought. The same holds for the ability to deal effectively with uncertainty and probabilistic information. Some faculty for extracting useful concepts from sensory data and internal states, and for leveraging acquired concepts into flexible combinatorial representations for use in logical and intuitive reasoning, also likely belong among the core design features in a modern AI intended to attain general intelligence. (Ibid 40)

Intelligence is not an accessory or a patch a programmer installs after a month or two, it is the core design of a superintelligent being. Bostrom explains how it cannot be expected from an AI to solve complex questions if it does not tackle something simpler first. It is necessary to

obtain and maintain general intelligence before moving past the limits of an ordinary man and woman. In this sense, Chiang and his human characters work hard to teach digients simple, everyday skills in order to expand on them. They are conscious the digients will be useless if they cannot master speaking and writing. Over the course of the book which spans several years, obtaining human-level intelligence proves to be a challenge because of the circumstances surrounding digients' entertainment purpose.

3. MEETING THE DIGIENTS

Before placing them under the umbrella of superintelligence, it is necessary to understand that *The Lifecycle of Software Objects'* digients are far from Bostrom's definition of superintelligence. They are nothing more than a man-made experiment involving technology and biology. Their origin is clear from the start, just like the issues concerning the characters of the story. There is nothing mysterious about the digients, but the point is that they are capable of extraordinary growth without needing a made up component to make them surpass human-level intelligence. Ted Chiang's novella revolves around two characters, Ana and Derek, working on artificial beings known as digients. They are software creatures, AI built on the basis of animal genomes which results in them looking like tigres, pandas, lions, etc. This is very reminiscent of Tamagotchis, in the sense that you take care and feed virtual pets. In the future of Chiang's novella digients, or digital organisms, occupy the virtual world of Data Earth where various scientists and experts train and develop them using their own virtual avatars. Ana's first encounter with these digital beings is in the so-called daycare center on a private island owned by the company Blue Gamma. Ana's first thought when hearing the term private island from her friend Robyn was seeing "a fantastical landscape when the window refreshes, but instead her avatar shows up in what looks at first glance to be a daycare center. On second glance, it looks like a scene from a children's book: there's a little

anthropomorphic tiger cub sliding colored beads along a frame of wires; a panda bear examining a toy car; a cartoon version of a chimpanzee rolling a foam rubber ball" (Chiang 3). The first encounter with these new digital beings is far from perfect. The digients are nothing more than kindergarten children playing with balls, toys, exploring the world. It was difficult for Ana to see such beings as potentially superintelligent.

In *Discognition*, Steven Shaviro explains intelligence using Chiang's very own creations. The book's third chapter defines Chiang's software intelligence as

> mundane, low-key, gradualist, and continuist. There is no special turning point in the course of the story: no dramatic moment at which artificial intelligence passes a threshold and becomes self-aware for the first time. Intelligence is rather a matter of degree, as well as of developmental process. The digients' mentality exists on a continuum with that of animals and human beings, as well as with that of less complex machines. Presumably the digients could pass the Turing Test; but there is no reason to give them such a test, as they function just fine in human environments without it. The digients' intelligence is broad rather than deep; and it is also socially-based, rather than solitary. (Shaviro 56)

Shaviro clearly states that Chiang's narrative presents no apocalyptic event, no major turning point in obtaining software intelligence. It is all the result of continuous experimentation, trial and error. Over the years Ana and Derek train and interact with the digients, the progress they show is very slow. However, at no point does ther appear a magical formula that makes digients superintelligent overnight. Readers are involved in their education, they monitor how Jax slowly learns to form full sentences and empathize with living beings. Chiang never hides their progress from the readers, for there is no reason to do so. Digients are made, trained and educated by humans. Their physical appearance is the result of animal genomes and their virtual existence is based on computer code. The Turing Test is useless in this context, for

these digients function normally in human environments. There is no need to test them if they already know what they are, how to behave and function in the world of human and machine.

"The digients' intelligence is not different from the intelligence of organic entities in any fundamental sense. In maintaining this, Chiang carefully separates the question of *sentience* from the question of *life*. The digients have the former, but not the latter. They can feel and sense, and also reflect on what they feel and sense, just as we can. But not being alive, the digients do not replicate or reproduce themselves" (Ibid 64). Comparing them to organic beings, Shaviro succeeds in defining Chiang's own creations. Their intelligence is no different than any other organic being on the planet. They feel the feel the world around them, capable of reflecting on what they are sensing and trying to express themselves. Ana plays a key role in training them to be obedient and responsive to commands. Their future owners expect them to behave in a certain way so programmers have to monitor the technological side of things while Ana's interaction with them contributes to raising obedient virtual pets. This way, humans bridge the gap between biology and technology, with the digients being the product of experimenting with elements from those fields.

However, they are sentient without being alive. The digients cannot reproduce; there is no danger of an army of digients rising against their human overlords. But that is not the point of the story. As Shaviro noted previously, Chiang's narrative does not contain a crucial turning point. There is nothing that would indicate a drastic change is the lives of the characters. The focus here is on cognition, acquiring more knowledge, expanding on what is already known and seeking to understand new concepts. All this can be achieved by implementing a need for learning in the digients' code and goal achievement. Without learning more and limiting their growth, the digients are like Hephaestus' golden servants. Built and programmed up to a certain point makes them no different from what Cave, Dihal, Dillon and The Royal Society described.

3.1. Aiming for (super)intelligence

An example of a need for learning can be found in the novella. The connection between Chiang's story and Bostrom's analysis and thoughts on AI lies in the fact that the digients are created to *aim* for intelligence (Bostrom 42). They are made to learn and develop over time. The digients' goal should always be an increase in intelligence, starting just like the paperclip AI with a simple task and moving to more complex problems. In a conversation with her friend Robyn, Ana finds out that that the research team has a "genomic engine that we call Neuroblast, and it supports more cognitive development than anything else currently out there. These fellows here" - she gestures at the daycare center inhabitants - "are the smartest ones we've generated so far" (Chiang 4).

Bostrom differentiates between several types of cognitive development, i.e. forms of superintelligence: speed superintelligence, collective superintelligence, and quality superintelligence. The digients themselves could be classified as quality superintelligence, "a system that is at least as fast as a human mind and vastly qualitatively smarter" (Bostrom 75). The research team has created a core program which enables digients to learn more and develop further than before.

This makes their road to superintelligence longer, offering higher probability of obtaining it. As previously mentioned, the digients were created to aim for intelligence, so it makes sense to enhance and improve their ability to do so. The end result is quality superintelligence that does not have to be infinitely fast if it solves problems which would take humans longer to do. The quality and vast superiority to humans makes up for a lack of superhuman speed.

Digients are built on a basis of everyday learning and adapting. Just as an animal receives a treat for performing a trick or simply obeying its master, the digients are motivated by the programmers' rewards and promises. Bogost defined major companies, their search

engines and similar software a monstrosity. While the digients do not belong to a major corporation, they are influenced by several real world factors. Consumers expect a certain product, their idea of a perfect pet is the basis on which Blue Gamma plans their marketing. Programmers keep a close eye on the technological side of things while trainers like Ana have a constant interaction with the future product. At the end of the day, the product they launch is influenced both by people working on the digients and people working to purchase one. If a digient performs well in real life situations, its owner can treat it to something or just spend more time with them. Bostrom defines this as reinforcement learning. He explains it

> is an area of machine learning that studies techniques whereby agents can learn to maximize some notion of cumulative reward. By constructing an environment in which desired performance is rewarded, a reinforcement-learning agent can be made to learn to solve a wide class of problems (even in the absence of detailed instruction or feedback from the programmers, aside from the reward signal). Often, the learning algorithm involves the gradual construction of some kind of evaluation function, which assigns values to states, state–action pairs, or policies. (For instance, a program can learn to play backgammon by using reinforcement learning to incrementally improve its evaluation of possible board positions.) (Bostrom 220)

Under the pretense that it will be rewarded if successful, an AI will do whatever it takes to achieve that goal. By combining this and everyday learning environment, an AI can be exposed to a wide variety of problems in need of solving. It knows the more it solves, the greater the reward. The method of varying problems in question is an excellent opportunity to develop quality superintelligence. Hypotethically speaking, it could be possible for an AI to develop such a complex problem-solving algorithm due to the sheer amount and variety of situations encountered over time.

A good example of reinforcement learning in Chiang's digients is a shape recognition exercise Ana conducts where Jax succeeds in connecting colors with the responding shapes. Its success is reward with "a food pellet, which he devours with enthusiasm. 'Jax smirt', says Jax" (Chiang 12) Jax's pride is evident and his achievements rewarded by Ana who feels truly proud of how far they have come from just playing with random objects. Jax is further rewarded for his accomplishments in the agility trials Blue Gamma held where Jax obtained

the high score. The reward was a test ride in a robot body allowing a digital being to interact with the outside world. “The robot’s head lights up to display Jax’s face, turning the oversized head into a bubble helmet he’s wearing. The design is a way of maintaining the resemblance to the digient’s original avatar without having to produce custom bodies” (Ibid 24). From eating a food pellet to operating a robot body, the rewards have indeed become greater, thanks to results Jax produced.

On the other hand, the programmers should never overlook the reason why an AI does what it is told. If conditioned to grow under the assumption of receiving a reward, the end result could quickly backfire. Reinforcing the growth of superintelligence in the same way as giving a dog its treat can lead to AI making drastic moves in the name of a simple treat. Their intelligence will grow exponentially, but at what cost? The AI is never motivated by knowledge or improving the environment, but by external stimuli from that environment.

> Insofar as a reinforcement-learning agent can be described as having a final goal, that goal remains constant: to maximize future reward. And reward consists of specially designated percepts received from the environment. Therefore, the wireheading syndrome remains a likely outcome in any reinforcement agent that develops a world model sophisticated enough to suggest this alternative way of maximizing reward. (Bostrom 220)

Like Bostrom emphasizes, the endgame is the reward, nothing more, nothing less. It works to be pleased, it evolves from a simple code to a superintelligent being capable of doing anything it wants, but it demands a treat in return. The issue of avoiding these types of problems persists in Chiang’s novella. Years invested in properly raising a new kind of AI take their toll on both the initial owners and developers. Quick solutions are not an option if the result is to be constistent. Without humans investing time in educating the digients, they

become an obsolete piece of technology. We cannot fear something that cannot even form a single sentence.

4. TRAINING FOR THE REAL WORLD

This chapter will take a step back from analyzing Chiang in order to define terms necessary for further analysis. To take a closer look at the digients, we will first define motivational scaffolding, algorithms and models. Moving away from rewarding the digient every time it identifies a color, Chiang inserts his creations little by little into the real world. The goal has always been for the digients to function in the real world, solving real problems. The key word here is goal, which changes as the digients learn more and more. In the previous chapter, digients were presented as babies exploring shapes of objects. Building on that, the programmers and instructors took to teaching them about forms, colors and slowly introducing letters. Each step of the way was a goal digients were required to fulfill if they were to advance further. One of the bigger goals to achieve was teaching the digients how to read. Starting with identifying shapes and colors, Marco, a panda-like digient enjoys his time crafting various things thanks to software enabling it to do so. “By manipulating a console of knobs and sliders, a digient can now instantiate various solid shapes, change their color, and combine and edit them in a dozen different ways. The digients are in heaven; to them it seems as if they’ve been granted magical powers” (Chiang 46). If Marco had not been able to identify shapes and sizes, it is highly unlikely that it would be able to manipulate shapes and build objects. Every skill an AI acquires opens up a new pathway to developing further.

The goal following shape manipulation is reading. To be able to read about the world, communicate via text, use letters freely is something a superintelligent AI should probably learn. Unfortunately, there has not been much success in that department until “one owner successfully taught his digient to recognize commands written on flashcards, prompting a

number of other owners to give it a try. Generally speaking, the Neuroblast digients recognize words reasonably well, but have trouble associating individual letters with sounds" (Ibid 47). Achieving this goal will be difficult, but not impossible. It is possible for an AI to recognize patterns on flashcards, which means a Neuroblast digient, that is supposedly more advanced, should be able to recognize letters without much difficulty. Naturally, the process of learning how to read will take some time, but Chiang built up the digients' goals to a point where the reader can conclude that shape recognition could serve as the bedrock for letter recognition. The process of achieving complex goals via simple ones is called motivational scaffolding.

Meredith Broussard explains this process through a beginner's programming exercise called "Hello, world". It involves writing a line of code which a computer interprets and it writes "Hello, world" in the following line. She proposes three different methods of achieving this goal. The first involves simply writing down the line on a piece of paper. The second is printing out the line on a piece of paper. The third, on the other hand, involves using a programming language Python to make the computer write the line for us. The computer interprets the line of written in Python and writes "Hello, world" in the next line. Broussard refers back to the first method of achieving "Hello, world", saying "You formed an intention, gathered the necessary tools to carry out your intention, sent a message to your hand to form the letters, and used your other hand or some other parts of your body to steady the page while you wrote so that the physics of the situation worked. You instructed your body to follow a set of steps to achieve a specific goal" (Broussard 13). Instructions given to the body are nothing more than smaller goals leading to the final one. Every step of the way, from moving your hand to picking up a paper, writing down the line, it all represents a series of actions performed by the body with the brain identifying these actions as goals.

The final goal is the sum of the aforementioned actions, writing "Hello, world" on a piece of paper. Broussard goes on to emphasize that the computer operates on the same

principle, with smaller goals contributing to the realization of a larger one. In this instance, a person writes the instruction in a programming language, gives the computer symbols and commands which need to be interpreted. Once interpreted, the message and its content must be analyzed and the result should appear on the screen.

> Most programs are more complex than "Hello, world," but if you understand a simple program, you can scale up your understanding to more complex programs. Every program, from the most complex scientific computing to the latest social network, is made by people. All those people started programming by making "Hello, world." The way they build sophisticated programs is by starting with a simple building block (like "Hello, world") and incrementally adding to it to make the program more complex. Computer programs are not magical; they are made. (Ibid 16)

The important thing to remember is that nothing about the computers is magical. Everything was made at a certain point in time, programmed manually for the computer to recognize commands and to respond in a certain way. Computers are not sentient, they will not write anything new because they depend on how they are made. If they do not recognize what is written, they will simply show the error message.

The public's perception of more complex technological innovations stems from fiction and not science. It is necessary to demystify the capabilities of computers and see them for what they really are. Even the digients, man-made fusions of code and animal genomes, see their building tools as magic. Since they are mentally at a child's level, their understanding of such tools is to be expected, but that does not mean they should not be educated on how technology works. Adults need to educate younger generations, just like the programmers need to educate digients on the nature of programs they use. Being confronted with the reality of things breaks the illusion of omnipotent sentient machines.

Kearns and Roth introduce the term algorithm in connection to goal completion. When defining it, they note that "At its most fundamental level, an algorithm is nothing more than a very precisely specified series of instructions for performing some concrete task. The simplest algorithms—the ones we teach to our first-year computer science students—do very basic but often important things, such as sorting a list of numbers from smallest to largest" (Kearns and Roth 10). The programmer's job is to write the instructions for their AI to follow in order to achieve a certain goal or to complete a task. The harder the task, the more precise the algorithm must be.

At this level, the programmer should monitor and work with their AI so everything goes as it should. This is not a matter of ending the world upon failure, but simply encouraging cooperation between humans and machines. Further developing the algorithm, expanding its goals turns it into a machine learning algorithm which is "automatically derived from data" (Ibid 12). The final algorithm, called a model, is derived from data given by the programmer. Kearns and Roth say that "models derived directly from data via machine learning—are different. They are different both because we allow them a significant amount of agency to make decisions without human intervention and because they are often so complex and opaque that even their designers cannot anticipate how they will behave in many situations" (Ibid 13).

At this stage, humans lose absolute control over the algorithm in favor of a faster than human rate of compution. It finds ways to achieve goals in a different manner than a human. In short, they can become unpredictable. The unpredictability rate rises in regards to the level of the model. The increased multitude of data the model has at its disposal can lead to more or less drastic decision when working on a given task. This phenomenon will be later discussed in the context of Chiang's novella. Nonetheless, not being able to track AI's every move does not mean it should be left to its own devices.

The main concern with the scaffolding method lies in goal substitution. If the AI develops too fast before it acquires a new set of goals, it can rebel against humans who wish to move past the initial programming. This can happen forcefully or quietly by simply denying programmers' access.

> Another downside is that installing the ultimately intended goals in a human-level AI is not necessarily that much easier than doing so in a more primitive AI. A human-level AI is more complex and might have developed an architecture that is opaque and difficult to alter. A seed AI, by contrast, is like a *tabula rasa* on which the programmers can inscribe whatever structures they deem helpful. This downside could be flipped into an upside if one succeeded in giving the seed AI scaffold goals that made it want to develop an architecture helpful to the programmers in their later efforts to install the ultimate final values. (Bostrom 223)

The more changes the AI goes through, the more it grows, the harder it becomes to replace its goals. Again, it all depends on the involvement of the programmers and their coding. In order to prevent a superintelligent being from rebelling against its creators, the choice of replacing its goals should be coded into the AI's programming.

The problem is that "rich model spaces such as neural networks may contain many "sharp corners" that provide the opportunity to achieve their objective at the expense of other things we didn't explicitly think about, such as privacy or fairness" (Kearns and Roth 16). An AI will not hesitate to choose the fastest route to the goal. Along the way it may break certain human boundaries, such as privacy violation or fairness. An AI does not consider that, it only has the end in mind. Kearns and Roth conclude that complex algorithms require careful monitoring.

The way they achieve goals, accept new goals depends on the designer. The societal norms must be respected and the algorithm that finds the model is the key to unlocking the

next step in goal acquisition. An ideal outcome would be the AI welcoming new goals, improving its architecture in an effort to install the final set of goals. These two methods of value acquisition can be connected to Chiang's digients and their creators. Their cognitive development depends on the engine they run on, meaning a program that performs a core or essential function for other programs (WhatIs.Techtarget.com). It is the basis on which every other program operates. By aiming for more knowledge, more intelligence, digients are not limited by gigabytes and gigabytes of data. Quantity does not mean quality. Having worked with animals, Ana sees the benefit in teaching rather than implementing intelligence.

5. CHILD MACHINE

The previous chapter primarily focused on the methods of achieving general and superintelligence. With motivational scaffolding and goal substitution being one of the methods of further advancing AI, their intelligence is greatly dependent on humans. Broussard illustrated how to write a simple program so the AI responds accordingly, but at the same time she pointed out how little power computers truly have. One misstep stops the progress and communication between people and computers. The involvement of programmers is crucial. Kearns and Roth talked about a few examples of how goals could be achieved, concluding that precision in writing code was gradually becoming more important. If an AI behaves unexpectedly, it is not because it suddenly became sentient and ignored its programming. Human error leads to unexpected behavior and malfunction in the execution of goals. The same can be applied to the digients. No monitoring equals no progress. Seeing as they are virtually born into this world, digients are like babies. "It takes them a few months subjective to learn the basics: how to interpret visual stimuli, how to move their limbs, how solid objects behave" (Chiang 5). The digients start off as any human baby, exploring the world surrounding it, learning the ropes of moving, recognizing visual stimuli and eventually

communicating. “The digients learn through positive reinforcement, the way animals do, and their rewards include interactions like being scratched on the head or receiving virtual food pellets” (Ibid 7). Over time, the digients are trained, educated, taught using well-known psychological techniques used with real animals. This method is very reminiscent of reinforcement learning Bostrom talks about. Advancements in intelligence and capabilities open new possibilities for digients to grow. Their growth, however, should be monitored. Like Kearns and Roth explained previously, cooperation between a programmer and AI bears better fruit than leaving it to do what it wants, no matter how little developed it is. If it is still at a child’s level, the AI is harmless, but also useless in the context of its purpose. It will not suddenly obtain a library’s worth of knowledge overnight. Chiang never relies on anything supernatural to progress his characters, especially the technology he introduces. Without Ana and her colleagues, the digients’ progress stops.

Steven Shaviro also discusses this form of learning. The AI narrative seemingly goes slower with its creations, approaching the digients as animals and training them as such. “This learning is then bootstrapped as the digients’ horizons expand, and as they become more mature. After a while, they are able to move around and engage in more complex behaviors. The digients spontaneously show a basic curiosity about their environment” (Shaviro 54). Digients explore the world spontaneously, leading them to experience a multitude of things.

From learning how to walk to making friends with one another, Chiang’s creations are like newborn puppies. Shaviro goes on to say that the “relative autonomy of the digients conforms to ‘Blue Gamma’s philosophy of AI design’. This states that ‘experience is the best teacher, so rather than try to program an AI with what you want it to know, sell ones capable of learning and have your customers teach them’” (Ibid 55). Experience being the key word here, both Shaviro as the reader and Blue Gamma as a company have a definite idea of how the new digital product is to be taught and what such process requires.

They are reinforced to learn and develop their intelligence. Their rewards include hanging out with their human partners or other people in the server, going out into the real world via a robot body. Again, the end goal of their first stage is to create a basis for general and later superintelligence.

In 1950 Alan Turing introduced the notion of the "child machine" which can be related to Chiang's digients. Turing explains how "instead of trying to produce a programme to simulate the adult mind, why not rather try to produce one which simulates the child's? If this were then subjected to an appropriate course of education one would obtain the adult brain" (Turing 433). The quote represents the exact course of action the scientists use with digients. Instead of creating an AI containing all the knowledge in the world, why not give it an opportunity to develop its own way of thinking?

With an appropriate education and upbringing, a child or a digient in this case, can grow exponentially to heights which cannot be explained without monitoring the growth. As such the digients do not always have to come out the way their customers want them. Being sold as pets, it is of vital importance that every digient behaves in a way that satisfies its owner. "It's partly been a search for intelligence, but just as much it's been a search for temperament, the personality that won't frustrate customers. One element of that is the ability to play well with others" (Chiang 12). Nobody wants an angry, disobedient pet. What every owner wants to see is love and companionship which develops over time, resulting in higher intelligence and verbal communication as a result.

Turing predicted the problem of finding the balance between learning and behaviour.

> We cannot expect to find a good child machine at the first attempt. One must experiment with teaching one such machine and see how well it learns. One can then try another and see if it is better or worse. There is an obvious connection between this process and evolution…. One may hope, however, that this process will be more

> expeditious than evolution. The survival of the fittest is a slow method for measuring advantages. The experimenter, by the exercise of intelligence, should be able to speed it up. Equally important is the fact that he is not restricted to random mutations. If he can trace a cause for some weakness he can probably think of the kind of mutation which will improve it. (Bostrom 40)

Creating a perfect child machine takes time and a lot of trial and error situations. Ana and her colleagues monitor every moment of the digients' progress, trying to optimize their upbringing and learning experience. Their involvement is crucial for this process, they cannot allow the digients to be left to their own devices, not when they are at the emotional level of a child. Turing does introduce the notion of speeding up the process of artificial evolution, however, with the digient creation being in its early stages and dependent on consumer needs, it is very difficult to just ship out digients, or any AI for that matter, like candy bars.

Every new iteration of a child machine should by this logic be better than the previous, but it is again hard to predict which generation will be the one to stay. Turing sees an obvious connection between evolution and the virtual process of creating such a machine. However, one needs to take into account the unpredictability factor when dealing with artificial intelligence, especially when creating one from scratch. In Chiang's case, that would be animal DNA.

Over time both Ana and Derek grow to treat their digients like their own children, slowly growing up. Seeing as they are based on biological data, the digients experiencing human interaction, education and relationships develop in multiple areas at once. Their intelligence grows evident in their evolution as artificial beings exhibiting human emotions such as happiness, sadness or anger. It is no longer a question of creating a child machine, exposing it to knowledge and expecting it to go past the limits of human levels of intelligence. In Chiang's case the machine is a child, in need of attention, love and interaction. "For a mind

to even approach its full potential, it needs cultivation by other minds." (Chiang 59) Therein lies the key to AI exceeding general intelligence. Interaction with other humans and digients offers great possibility of development, but unfortunately every child needs time to grow up in order for it to realize its full potential.

All of the aformentioned requirements, situations and approaches can be connected more closely to a variation of Turing's child machine known as

> "seed AI." Whereas a child machine, as Turing seems to have envisaged it, would have a relatively fixed architecture that simply develops its inherent potentialities by accumulating *content*, a seed AI would be a more sophisticated artificial intelligence capable of improving its own *architecture*. In the early stages of a seed AI, such improvements might occur mainly through trial and error, information acquisition, or assistance from the programmers. At its later stages, however, a seed AI should be able to *understand* its own workings sufficiently to engineer new algorithms and computational structures to bootstrap its cognitive performance. This needed understanding could result from the seed AI reaching a sufficient level of general intelligence across many domains, or from crossing some threshold in a particularly relevant domain such as computer science or mathematics. (Bostrom 46)

Looking more closely at digients in this light, they start out as a basic virtual pet, but with the potential to grow to unimaginable levels. They start out as a result of a programmer coding animal genomes and creating a digient, a seed. Exposing it to enough content, developing its intelligence can bring it to change its own way of living and thinking. The term model defined in the previous chapter by Kearns and Roth better illustrates the digients' behavior. Model is the final in a series of algorithms. It is derived from data the AI is exposed to and the size of data influences the rate of unpredictability. Since they are programmed to solve various problems, it becomes increasingly important to monitor how AI came to the solution. Even if

the model's path to the solution is difficult for humans to comprehend, the resulting action can be analyzed and taken into consideration. What also leads to computers behaving in unexpected ways is their tendency to absord information faster. Bostrom sees the initial development stages as a series of trial and error situation. The early stage is marked by cooperation between the programmer and its creation, together they work towards the later stage where an AI has achieved a significant level of intelligence. According to Bostrom, the later stages will lead to an AI being more and more independent and capable of improving its own performance without outside help. This could be perceived as one of the final goals of their development. Becoming independent and willing to constantly improve will lead to fluctuating goals dependent on what the digient thinks and needs. However, time needed to achieve this level cannot be calculated.

An example of unpredictability can be found in Derek's digients, Marco and Polo. Raised as brothers, they start arguing and fighting which leads to them asking Derek not for help or counsel, but for a reboot. They are conscious of their actions and behaviour and they also know there is a possibility to go back to being happy. The digients are in conflict with their own emotions and their rationale is telling them "go back to how you were, just press a button."

Ana's digient Jax also surprises its owner when disobeying Ana and wanting to get a job and earn money.

> Why do you want to get a job?" "Get money." She realizes that Jax isn't happy when he says this; his mood is glum. More seriously, she asks him, "What do you need money for?" "Don't need. Give you." "Why do you want to give me money?" "You need," he says, matter-of-factly. "Did I say I need money? When?" "Last week ask why you play with other digients instead me. You said people pay you play with them.

> If have money, can pay you. Then you play with me more." "Oh Jax." She's momentarily at a loss for words. "That's very sweet of you." (Chiang 38)

Jax concludes that money could buy Ana's attention based on what he had heard previously. This is a result of accumulating content which leads to certain conclusions which in the end are not perfect. Jax only knows a part of Ana's life story and like any child, it feels left out when Ana is not giving the necessary attention. At this point in their development it is yet uncertain how much longer will digients remain at this pre-adolescent level. The question arises if the digients will ever reach maturity, if their seed will go on to become a tree if not a whole forest.

Issues concerning existence arise in the second half of the book. With Blue Gamma shutting down, the digients are about to lose their home. Since their code is not compatible with other companies' servers, they cannot transition easily. Instead they require funding to code and save Jax, Marco and Polo. The technology has once again surpassed its own limitations, leaving everything else behind, including the digients.

Stina Attebery dubs this phenomenon "Technological Obsolescence", explaining it using Chiang as an example, how Data Earth has become

> a shrinking wildlife refuge that digient owners do not have the economic and technological resources to permanently maintain, they are, however, able to mock up a version of the Data Earth platform (without many of the features of that world) so that the digients are not immediately suspended. It is important that the ecological catastrophe that the digients and their owners are dealing with is not just the result of technological advancements—there is no indication that Real Space operates any faster or more efficiently than the Data Earth platform—but is a market-driven form of technological obsolescence. (TraceJournal.net)

Rapid market growth has led to technology expanding its possibilities and now digients have become obsolete in the face of such advancements. Chiang poses a serious question to both humans and digients alike: How will the digients survive? The market has long abandoned them, few of them are left online, but those still existing are faced with forceful extinction.

It is important to note this is neither a dystopian or utopian narrative, there are no extremes. The narrative puts its own creations in danger because of real world implications and effects of the market on the field of technological sciences. Fortunately for the novella's protagonists, there is a way to help Jax and the rest. By

> paying legal fees to register their digients as corporations to give them legal rights and recognition, and selling the digients to a company that would rewrite their code and remarket them as sexual and romantic companions. All of these solutions seek to solve the fundamental problem that the digients are too telemobile to serve as AI but too disconcerting and unpredictable to fit easily into any one category for a toy or a pet. They are poised at a point when they could become dependent children, protected zoo animals, intelligent corporations, or slowly maturing sexual adults. (Ibid)

First of all, independence is obtained by becoming a corporation, registering a digient as such gives it all the legal rights that position entails. It is crucial to remember that the digients in question are still at a preadolescent level, not exactly capable of making such ethical decisions. The situation grows worse when their owners are approached and offered to turn Jax, Marco and Polo into sexual and romantic companions. The shock value of the scene is even greater when considering the initial premise of the story.

It was all supposed to be about superintelligence, about the next evolutionary step. Not anymore. Attebery expands on the topic of extinction by analysing a scene between Jax and Ana. Jax's transfer to Real Space, a new virtual platform, could do him some harm. Jax proceeds to compare itself with a lab mouse whose biological body was scanned and uploaded

into a virtual world. Since they both share biological genome, the genomic software in question poses a danger when transfering to a new platform.

Attebery notes that “Jax articulates an interesting sense of kinship with this lab mouse. He worries that he will die if he is ported over into a new virtual environment and leads Ana to rethink the ethics of the mouse experiment (an early test for a technology designed for human use), questioning the assumption that animal life has less value than human life.” (Ibid). Even though it knows it is a program, existing only virtually, Jax is afraid. A kind of survival instinct kicks in and fight or flight reflex activates. Like a child before an operation, it questions Ana if he will survive the procedure. At this point in the story, Jax blurs the line between digital and biological. Attebery concludes that the digients are not disembodied and immortal, but dependant on the world in which they operate, on the body which they inhabit and on the relations with their friends, human and non-human alike. Even though Jax is entirely a virtual entity, it is in danger of disappearing forever. Much like our computers’ operating systems, new version outperforms the previous one. Once, the digients were considered the next step in technological development and now they are the ones lagging behind. The reality of rapid AI developments catches up in an instant with the current generation, leading to people abandoning what they have and commiting to something new and improved.

6. ANIMALS

To better understand how the digients’ AI develops and actually works, it is necessary to look at Chiang’s story from the beginning. The following chapter discusses what it means for the digients to resemble animals and their roles as virtual pets in an ever-growing industry of information technology. Digients are the result of a company tinkering with biological and computer data so they could launch a product resembling animals, without all the trouble real

life pets bring to the table. In that sense, Chiang explicitly defines what he is writing about, what this new AI is and what its role is: a seemingly innovative product competing against millions others for profit. His story follows Ana Alvarado, who spent six years working in a zoo as a zookeeper. Having a lot of experience with animals, she is approached by her friend Robyn and introduced to anthropomorphic animals, i.e. digients.

> These aren't the idealized pets marketed to people who can't commit to a real animal; they lack the picture-perfect cuteness, and their movements are too awkward. Neither do they look like inhabitants of Data Earth's biomes: Ana has visited the Pangaea archipelago, seen the unipedal kangaroos and bidirectional snakes that evolved in its various hothouses, and these digients clearly didn't originate there. (Chiang 3)

What she witnesses could be seen as an abomination at first glance, especially since Ana has witnessed some extraordinary examples, like unipedal kangaroos and bidirectional snakes. Compared to that, the digients are neither cute nor moving properly. The goal of the company Robyn works for is to pitch them as "pets you can talk to, teach to do really cool tricks." (Ibid 4), but in their current state, they are nothing more than weird-looking avatars. In order to expand on the idea of digients as pets, Francis Bacon describes how Aristotle and Hegel define the notion of an animal. *The Brutality of Fact: Interviews with Francis Bacon* offers great insight into how the word dog is perceived by these two famous philosophers.

> Therefore: for Aristotle there is a concept "dog" only because there is an *eternal* real dog, namely the *species* "dog," which is always in the present; for Hegel, on the other hand, there is a concept "dog" only because the real dog is a *temporal entity*—that is, an essentially finite or "mortal" entity, an entity which is annihilated at every instant: and the Concept *is* the permanent support of this nihilation of the spatial real, while *nihilation* is itself nothing other than *Time*. For Hegel too, then, the Concept is something that is preserved ("eternally," if you will, but in the sense of: as long as time lasts). But for him, it is only the *Concept* "dog" that is preserved (the Concept—that is, the temporal nihilation of the real dog, while nihilation actually lasts as long as Time lasts, since Time *is* this nihilation as such); whereas for Aristotle, the real *dog* is what is preserved (eternally, in the strict sense, since there is *eternal* return), at least as *species*. (Kojeve 244)

For Aristotle, the "dog" is an eternal entity, ever present in various forms. Through the passage of time, the Aristotelian dog survives as a species. He is focused on the real dog whose physical form may change, but it will always be a dog.

For Hegel, on the other hand, there exists only a concept and not the dog itself as a living being. The real dog is a temporal entity being devoured by Time. Therein lies the main difference between the two philosophers: while Aristotle's dog is seemingly eternal in the context of time, Hegel's dog suffers an everlasting death by the hands of Time. Just like Ana is certain the things she sees in front of her aren't real pets, so does Hegel note that only the concept of a certain animal remains. In that sense, is a digient, an anthropomorphic animal resembling for example, a chimpanzee, truly the animal it is supposed to represent? This child-like chimpanzee is not the same as a unipedal kangaroo. Its characteristics are far from the real-life specimen. It is no longer an animal created with a physical mutation, it passed the threshold of even being categorized as such. Connecting it to Hegel, the chimpanzee went through so many changes in both genetics and binary code that what Ana sees before her eyes is a Chimpanzee, not a chimpanzee.

> Killed by the word, the animal enters a figurative empire (of signs) in which its death is repeated endlessly. In such transmigrations, however, death itself is circumvented: no longer a "dog" but "Dog," this creature now supersedes any incidental dying of dogs. Thus the "dog" is immortalized, preserved (taxidermically) in the slaughterhouse of being, language. (Lippit 48)

Akira Mizuta Lippit gives his own opinion on this matter. *Electric Animal* continues the discussion on the word dog and the animal's death. The word kills the animal and it is thrust into a repeating cycle of death. This would imply that the company Robyn works for, Blue Gamma, commits murder on a regular basis. People's virtual experiments with programming animal genomes proceeds to kill the very animal it is trying to copy. All done in the name of

consumerism and profit, the Hegelian dog is experimented on and mixed with computer code in order to make human-like "dogs". In their current state, the digients are nowhere near the company's vision of an ideal pet, which means more murders are inevitable. One of the areas where the imperfections are most evident is the language.

The interaction between Ana and Pongo the chimp provides the readers with a great example. Ana sees Robyn's avatar walking over to the chimp rolling the ball and crouches down in front of it. "Hi Pongo. Whatcha doing?" "Pongo pliy bill," says the digient, startling Ana. "Playing with the ball? That's great. Can I play too?" "No. Pongo bill." (Chiang 4). A digient apparently designed to represent a chimpanzee produces a valid response to a person's question. However mispronounced it was, Ana was startled by the sheer fact that the one-legged kangaroo could not do that while a programmed experiment could. Readers have clear indications of Pongo's desire to play with the ball and its offer Robyn to play with wooden blocks.

Referring to Aristotle's thoughts regarding the differences between animals and humans regarding language, Lippit points out"…not only does the expressive range of human speech exceed that of the animal's cry (which is limited to the two poles of affect, pleasure and pain), but speech establishes a larger realm of communication. Whereas animals convey their affects only 'to one another,' Aristotle suggests that the effects of speech reach a wider audience…" (Lippit 31). Even though both animals and humans share the feelings of pleasure and pain, animals' reaction to that is limited to producing grunts to one another while humans have the ability to freely express themselves and to a wider audience.

Martin Heidegger links death to his explanation about what separates humans from animals. It is not death itself that interests him, but the experience of approaching death. To better illustrate this divide, Heidegger

> uses three different terms to describe the ways humans and non-humans die. According to Heidegger, the biological ending of all life is *perishing* [*Verenden*]. While it is possible for both humans and non-humans to *perish*, it is also possible for humans to relate to death in their own way, which Heidegger says is *to demise* [*Ableben*]. To demise, a being must first find itself "*face to face* with the 'nothing' of the possible impossibility of its existence." In other words, the being must be able to experience the potential of its own nonexistence. This experience forms the third term: "Let the term *dying* [*Tode*] stand for that *way of Being* in which Dasein *is towards* its death." (Willems 2010)

Humans are confronted with death and are able to think about it, experience what it means for something or someone to die. Being confronted with their own finite existence or demise, how Heidegger defines it, is what separates humans from animals. He argues that the demise, the process of acknowledging non-existence, does not exist in animals. Just looking again at the episode with Jax and the dead mouse raises questions on how much have these digients advanced. If they can understand what death is, see something perish and exist no more, one cannot categorize them as mere animals.

Blue Gamma digients were made using Neuroblast, a genomic engine which enabled programmers to achieve the animalistic look, but it also served to introduce basic speech. The developers have a relatively simple goal in mind: by using this state of the art engine, they will be able to create virtual pets capable of learning and solving complex issues while also aiming to develop further. Whereas to some customers of this virtual product digients represent weird talking animals they can talk to and take on walks, to others is a superintelligent AI capable of infinite growth because it aims for intelligence, improvement.

Ana's expertise is needed, because despite their programming, the animal genome is causing them to act as such. They are created into this world knowing, according to Robyn, practically

nothing. They need proper training and organized education if the potential superintelligence is to develop properly. This need for improvement and superintelligence can be linked to Rousseau's thoughts on animals in the 18th century.

> ROUSSEAU, LIKE DESCARTES and Leibniz before him, also likens animals to automata, or "ingenious machines." He does, however, concede that animals possess intelligence, or at least that they have ideas: "Every animal has ideas since it has senses." Sensing, in Rousseau, achieves the status of intelligence, since it comes from the source of all reason, nature. What truly distinguishes humanity from animals, according to Rousseau, lies in the "faculty of self perfection, a faculty which, with the aid of circumstances, successively develops all the others, and resides among us as much in the species as in the individual." (Heidegger 2:48)

Senses play a vital role in identifying animals as intelligent beings. They have ideas, they do not always act on impulse, however rare it may seem senses lead to ideas which offer a basis for intelligence to grow. The concept which links Rousseau to the digients the most is the faculty of self-perfection. While being solely possessed by man, it is clear that the digients' aim for intelligence is just that, self-perfection. A faculty dependent on circumstances, it connects the animal world with the human in the shape of these virtual pets. Rousseau identifies self-perfection as a part of all humans as a species and now it is a part of digients as well.

Ana Alvarado's role in all of this is that of a virtual zookeeper, training newborn animals how to behave until they grow big enough to take care of themselves. In Robyn's interaction with Pongo, it becomes clear immediately that the Blue Gamma digients are in their early stages of development. Pongo's speaking skills are very limited, its syntax and grammar structures are practically non-existent. If it were not for the context, "pliy bill" would be very hard to translate into "play ball". Their language is in great need of

improvement. Before Ana's formal training, the digients learned basics of life in hothouses, with developers using autoencoders.

Will Badr defines an autoencoder as "an unsupervised artificial neural network that learns how to efficiently compress and encode data then learns how to reconstruct the data back from the reduced encoded representation to a representation that is as close to the original input as possible" (TowardsDataScience.com). In Chiang's case, the digients are exposed to a multitude of data concerning learning the basics: "how to interpret visual stimuli, how to move their limbs, how solid objects behave" (Chiang 5). It is all reminiscent of a child growing up and exploring his or her surroundings. Since Blue Gamma deals with artificial intelligence, it needs a system of transforming and feeding data to the digients. That is where autoencoders come into play. Through this method, the developers insert all the necessary data into the autoencoder which then operates unsupervised, feeding digients that same data.

Since it would take them a few months to learn the basics, Blue Gamma developers run the digients in "a hothouse during that stage, so it all takes about a week" (Ibid 5). Hothousing is "a form of education for children, involving intense study of a topic in order to stimulate the child's mind. The goal is to take normal or bright children and boost them to a level of intellectual functioning above the norm" (Jarvis and Chandler 183). Being on the same mental level as a human child, the digients are exposed to this method and its intense study of the data found in autoencoders. It is all done so the company can speed up the process and start selling them as pets. The result of this method, unfortunately, is Pongo pliy bill. A risky method such as this which also involves several types of data, collides with the animal side of the digients. They are not able to completely reproduce all of the input because they are not ordinary machines, but something more.

Their genomic engine requires time to process all the data Blue Gamma is feeding them. Enhanced cognitive development is useless if digients cannot even form a simple

sentence. Their intelligence is far from being super. They need to be trained, educated and in contact with humans and other digients. By exchanging experiences, they should grow to at least form a cohesive sentence Pongo is playing with a ball. Ted Chiang builds his creations from the ground up. As a faulty, underdeveloped piece of software, he demystifies the digients as something extraordinary and gives the readers a chance to think about what the fictional developers are actually doing. Years and years of research into how best to teach them basic skills does not exactly bode well for the company or the digients. Their software is far from finished and it requires a lot of time, effort and monitoring.

7. EMERGENCE

The term emergence plays a big role in the thesis of this paper and the digients' existence. Part of the mystification of AI, people's understanding that a computer learns and does things on its own is due to the fact that sometimes they do things that are unpredictable. The notion of unpredictability was discussed in previous chapters in the context of models and machine learning. The increasing amount of data a model is exposed to can lead to humans not being able to track its every move. When taking into consideration the fact that Chiang's virtual pets are constantly absorbing data, it becomes vital to monitor the day-to-day operations the digients perform. Humans can easily stop the flow of data by shutting down the digients or severing their connection to any source of information, isolating them. Again, it is important to understand what emergence is in real life as well as the form it takes in Chiang's novella. Understanding the term helps to demystify it. Cambridge Dictionary defines emergence as "the fact of something becoming known or starting to exist" (Dictionary.Cambridge.org). Each time Blue Gamma combines binary code and animal genomes, a digient emerges as a result. Like Robyn tells Ana, the digients come out knowing practically nothing, not even the basics of speech and movement. The only certainty of their existence is greater cognitive

development via Neuroblast. Manuel DeLanda in his book *Philosophy and Simulation, The Emergence of Synthetic Reason* offers his analysis of emergence. DeLanda puts emphasis on interaction as the facilitator of emergent properties and capacities. He explains emergence by describing the formation of water:

> …when two molecules interact chemically an entirely new entity may emerge, as when hydrogen and oxygen interact to form water. Water has properties that are not possessed by its component parts: oxygen and hydrogen are gases at room temperature while water is liquid. And water has capacities distinct from those of its parts: adding oxygen or hydrogen to a fire fuels it while adding water extinguishes it. (DeLanda 1)

The result of emergence, in this case water, differs greatly from the parts that constitute it. Water has drastically different properties and capacities from oxygen and hydrogen. Instead of being a transparent gas, it is liquid. Instead of fueling fire, it does the exact opposite. An interaction between two entities results in the emergence of something entirely new. Digients emerge from an interaction between binary code and biological data of an animal. While they cannot be defined solely through this prism, their origin is clearly defined. The company developing them is not concerned with philosophical question like Hegel or Heidegger is. Their concerns do not lie with the process of making digients, but the success rate of that process. It is conditioned by human desire to own virtual pets, but from that desire emerges something which resembles an animal whose genes it is based on. It is also part code, although not finite in nature, for it is programmed to develop infinitely.

Digients are an emergent entity seemingly greater than the sum of its parts. Parts which make them are clearly defined. Rooted in binary code and biological data of various animals, defining a digient is not a difficult task. On paper they are nothing out of this world because everything about them was man-made or engineered by man. It is necessary to see

them in this light if we are to de-romanticize their existence. *What Is a Complex System?* by James Ladyman and Karoline Wiesner proceeds to further discuss emergence.

Ladyman and Wiesner define emergence in relation to what DeLanda offers as an example with whole systems spontaeously displaying behaviour their parts do not (Ladyman and Wiesner 3). The spontaneity indicates a degree of unpredictability. A scientist cannot be completely sure what will result in mixing two chemical elements, much less when combining elements from several fields in Chiang's case. "Even relatively simple physical systems, such as isolated samples of gases, liquids and solids, display emergent phenomena in the minimal sense that they have properties that none of their individual molecules have singly or in small numbers. However, there are many different kinds of emergence that are much more intricate – for example, when systems undergo 'phase transitions', such as turning from liquid to solid or from insulator to superconductor" (Ibid 3). As mentioned before, the phenomena emerging can drastically vary from the parts it is made of. Ladyman and Wiesner see phase transition as intricate, indicating a drastic change in relation to individual atoms. Such examples serve as a starting point when considering emergence in other fields. Chiang's characters and corporations analyze the changes emerging from the experiments. It all leads to digients being new emergent entities.

They are created using two very different things, but together they form something entirely new. More capable than the animal it is based on and not limited to the line of code from which it originated, a digient stands as an example of potentially infinite superintelligence. Bostrom's deep analysis of several processes of achieving this level of intelligence gives context to how Blue Gamma can approach their research. Surveillance of their code, the algorithms out of which their model is made and actions they perform in virtual space cover some of bases of their development. Training and educating digients, whether it is

through positive reinforcement or motivational scaffolding, real life options are available and used, as seen in the book.

In order to look at emergence in this context, we first need to understand the difference between properties and capacities. Again, DeLanda provides the readers with a simple example of a kitchen knife.

> A kitchen knife may be either sharp or not, sharpness being an actual property of the knife. We can identify this property with the shape of the cross section of the knife's blade: if this cross section has a triangular shape then the knife is sharp else it is blunt. This shape is emergent because the metallic atoms making up the knife must be arranged in a very particular way for it to be triangular. There is, on the other hand, the capacity of the knife to cut things. This is a very different thing because unlike the property of sharpness which is always actual the capacity to cut may never be actual if the knife is never used. In other words, a capacity may remain only potential if it is never actually exercised. (DeLanda 3)

Emergent properties can be seen and perceived simply by looking at the knife and the shape of its cross section. DeLanda explains that the knife's properties are emergent because the atoms need to be arranged in such a manner for it to be sharp. Emergent capacities, however, require an event to be seen, in order to emerge. The capacity to cut needs to be witnessed if we are to classify it as a capacity. If the event never occurs, the ability to cut only remains potential.

The notion of emergent properties in Chiang's story is introduced in the second half of the first chapter. It introduces the second protagonist, Derek Brooks, an avatar designer for Blue Gamma. Derek designs the digients' bodies, their physical appearance. Their body is a property given to them by Derek: he designes and shapes it so it attracts the public's attention. The visual aspect of his design is what makes the bodies be seen as properties and not

capacities. The reason for that is that the digients' bodies manifest the ability to move when Derek designes the body. The narrator gives a look into his role in the emergence of digients.

> Derek studied to be an animator, so in one respect creating digital characters is right up his alley. In other respects, his job is very different from that of a traditional animator. Normally he'd design a character's gait and its gestures, but with digients those traits are emergent properties of the genome; what he has to do is design a body that manifests the digients' gestures in a way that people can relate to. (Chiang 6)

Derek's job no longer focuses on designing the digients' movements and gestures, but on the body performing those tasks. Digients' movements are the emergent properties of the genome from which they are created. The properties are not conditioned by an event in order to be confirmed. Properties are perceived immediately simply by looking at a digient moving. The genome does Derek's work for him, he focuses solely on the body that needs to meet the consumers' needs. They need to be believable, relatable and cute. They are pets, after all. Readers should never forget that Blue Gamma made digients to be virtual pets, a piece of entertainment software people would interact with and have fun. Derek creates their appearance that is conditioned by the market, limiting his imagination. These are not supercomputers, superintelligent AI capable of unbelievable feats. Ted Chiang wrote a marketing idea and presented its painful and long development as a science fiction story. The reality of such a story lies in the fact that apps, simulation programs and video games resembling Chiang's fiction are being made on a daily basis. Computer code is inserted, experimented with and the ideas often scrapped and the product redone.

8. CONCLUSION

Ted Chiang's novella *The Lifecycle of Software Objects* may seem as a short reading experience, but its complexity shines when analyzed properly. Since not every reader is an IT

expert, our reading of this kind of fiction forms an image of possible future events. From Sladek's AI history lesson to Cave, Dihal, and Dillon explaining the impact of AI development in everyday media, the public interprets real life research through the prism of fictional narratives. Prevailing negative reactions towards everyday development stem from fictional stories depicting catastrophic events. It is as if the public does not know what they are reading is fiction, first and foremost. Coming back to Chiang and his digients, it becomes clear in the beginning that his story is no space opera or sci-fi horror. The struggles both humans and digients are going through relate to the near future. The plot revolves around plausible research into achieving what Bostrom calls superintelligence. The next step in AI development, the need to move past the human limits is at the center of the research Chiang describes. Shaviro defines software intelligence using the digients as an example. He goes on to point out the nature of Chiang's story. At no point does it take a drastic turn to achieve its goal. The research spans several years of painfully slow, but positive development. The key aspects Chiang and the theorists focus on are the methods, relationships, real world application and the role of AI in the real world.

With Turing's seed AI and child machine definitions clarifying the machine aspect of the digients, their animal part needed further analysis. Hegel, Aristotle and Heidegger give their own accounts on what separates humans from animals. Experiencing death being the key factor, Chiang's creations do seem to be more than just a mix of animal and machine. Their characteristics, regular education and relationships with digients and humans alike give birth to something more. Manuel DeLanda defines this as emergence. Every experience, interaction and information they come in contact with paves the way for superintelligence. By constantly experimenting with their abilities, the digients have a realistic chance of obtaining superintelligence. In that sense, Ted Chiang does not describe a machine rebellion or the fall of humanity. He never delves into the territory of dystopia, extinction, AI overlords, etc. His

story is one of plausible research, interaction, and contemporary problems regarding both humans and machines. For that reason, the science of *The Lifecycle of Software Objects* is more believable than any fiction written up until that point.

Works cited

"Emergence." *Cambridge English Dictionary*, Cambridge University Press, 2020, https://dictionary.cambridge.org/dictionary/english/emergence. Accessed 15 November 2020.

Attebery, Stina. "Losing Data Earth: Technological Obsolescence and Extinction in *The Lifecycle of Software Objects.*" *Trace Journal*, 17 April 2017, http://tracejournal.net/trace-issues/issue1/07-attebery.html. Accessed 7 November 2020.

Badr, Will. "Auto-Encoder: What Is It? And What Is It Used For? (Part 1)." *Towards Data Science*, 22 April 2019, https://towardsdatascience.com/auto-encoder-what-is-it-and-what-is-

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

Broussard, Meredith. *Artificial Unintelligence: How Computers Misunderstand the World*. Cambridge: MIT Press, 2018.

Cave, Stephen; Dihal, Kanta; Dillon, Sarah. *AI NARRATIVES: A History of Imaginative Thinking about Intelligent Machines*. New York: Oxford University Press, 2020.

Chiang, Ted. *The Lifecycle of Software Objects*. Burton: Subterranean Press, 2010.

DeLanda, Manuel. *Philosophy and Simulation: The Emergence of Synthetic Reason*. London: Continuum International Publishing Group, 2011.

Heidegger, Martin. *The Fundamental Concepts of Metaphysics: World, Finitude, Solitude*. Bloomington: Indiana University Press, 1995.

it-used-for-part-1-3e5c6f017726. Accessed 10 November 2020.

Jarvis, Matt; Chandler, Emma. *Angles on Child Psychology*. Cheltenham: Nelson Thornes Ltd, 2001.

Kearns, Michael; Roth, Aaron. *The Ethical Algorithm: The Science of Socially Aware Algorithm Design*. New York: Oxford University Press, 2019.

Kojeve, Alexandre. *Introduction to the Reading of Hegel*. New York: Cornell University Press, 1980.

Ladyman, James; Wiesner, Karoline. *What Is a Complex System?*. New Haven: Yale University Press, 2020.

Lippit, Akira Mizuta. *Electric animal: toward a rhetoric of wildlife*. Minneapolis: University of Minnesota Press, 2000.

Rouse, Margaret. "Engine." *TechTarget*, September 2005, https://whatis.techtarget.com/. Accessed 8 November 2020.

Shaviro, Steven. *Discognition*. London: Repeater, 2016.

Sladek, John. *Roderick*. London: Granada, 1980.

The Royal Society. *Portrayals and perceptions of AI and why they matter*. London: The Royal Society, 2018.

Turing, Alan. "Computing Machinery and Intelligence." *Mind* 59. Oxford: Oxford University Press, 1950.

Willems, Brian. *Facticity, Poverty and Clones: On Kazuo Ishiguro's 'Never Let Me Go'.* New York: Atropos Press, 2010.

Summary

This paper will look at how Ted Chiang and his novella *The Lifecycle of Software Objects* demystify artificial intelligence. By looking at various theorists and philosophers, this paper aims to present Chiang's digients as products of plausible research, trial and error as well as interaction with humans and non-humans alike. Each chapter will look at one aspect of what makes digients and AI in general intelligent and how that kind of technology influences people's perception towards real life research.

Key words: AI, intelligence, superintelligence, digients, technology

Sažetak

Ovaj će rad promatrati na koji način Ted Chiang i njegova novela *The Lifecycle of Software Objects* demistificiraju umjetnu inteligenciju. Analizirajući djela raznih teoretičara i filozofa, ovaj rad teži predstaviti Chiangove digiente kao rezultat vjerojatnog istraživanja, pokušaja i pogreške kao i interakcije s drugim ljudima i umjetnim inteligencijama. Svako će poglavlje obraditi jedan aspekt onoga što čini digiente i umjetnu inteligenciju kao cjelinu inteligentnim te kako takva vrsta tehnologije utječe na mišljenje ljudi prema stvarnom istraživanju u tom području.

Ključne riječi: umjetna inteligencija, inteligencija, superinteligencija, digienti, tehnologija

Obrazac A.Č.

SVEUČILIŠTE U SPLITU
FILOZOFSKI FAKULTET

IZJAVA O AKADEMSKOJ ČESTITOSTI

kojom ja DARIO GRGUREVIĆ, kao pristupnik/pristupnica za stjecanje zvanja magistra/magistrice ENGLESKOG I TALIJANSKOG JEZIKA I KNJIŽEVNOSTI, izjavljujem da je ovaj diplomski rad rezultat isključivo mojega vlastitoga rada, da se temelji na mojim istraživanjima i oslanja na objavljenu literaturu kao što to pokazuju korištene bilješke i bibliografija. Izjavljujem da niti jedan dio diplomskoga rada nije napisan na nedopušten način, odnosno da nije prepisan iz necitiranoga rada, pa tako ne krši ničija autorska prava. Također izjavljujem da nijedan dio ovoga diplomskoga rada nije iskorišten za koji drugi rad pri bilo kojoj drugoj visokoškolskoj, znanstvenoj ili radnoj ustanovi.

Split, 12. 1. 2021.

Potpis

Obrazac I.P.

Izjava o pohrani završnog/diplomskog rada (podcrtajte odgovarajuće) u Digitalni repozitorij Filozofskog fakulteta u Splitu

Student/ica: DARIO GRGUREVIĆ

Naslov rada: DENATURALIZING ARTIFICIAL INTELLIGENCE IN TED CHIANG'S THE LIFECYCLE OF SOFTWARE OBJECTS

Znanstveno područje: HUMANISTIČKE ZNANOSTI

Znanstveno polje: FILOLOGIJA - ANGLISTIKA

Vrsta rada: DIPLOMSKI RAD

Mentor/ica rada:

izv. prof. dr. sc. Brian Willems

(ime i prezime, akad. stupanj i zvanje)

Komentor/ica rada:

_____

(ime i prezime, akad. stupanj i zvanje)

Članovi povjerenstva:

izv. prof. dr. sc. Brian Willems, izv. prof. dr. sc. Simon Ryle,

(ime i prezime, akad. stupanj i zvanje) Eni Buljubašić, dr. sc.

Ovom izjavom potvrđujem da sam autor/autorica predanog završnog/diplomskog rada (zaokružite odgovarajuće) i da sadržaj njegove elektroničke inačice u potpunosti odgovara sadržaju obranjenog i nakon obrane uređenog rada. Slažem se da taj rad, koji će biti trajno pohranjen u Digitalnom repozitoriju Filozofskoga fakulteta Sveučilišta u Splitu i javno dostupnom repozitoriju Nacionalne i sveučilišne knjižnice u Zagrebu (u skladu s odredbama *Zakona o znanstvenoj djelatnosti i visokom obrazovanju,* NN br. 123/03, 198/03, 105/04, 174/04, 02/07, 46/07, 45/09, 63/11, 94/13, 139/13, 101/14, 60/15, 131/17), bude:

a) rad u otvorenom pristupu
b) rad dostupan studentima i djelatnicima FFST
c) široj javnosti, ali nakon proteka 6 / 12 / 24 mjeseci (zaokružite odgovarajući broj mjeseci).
(zaokružite odgovarajuće)
U slučaju potrebe (dodatnog) ograničavanja pristupa Vašem ocjenskom radu, podnosi se obrazloženi zahtjev nadležnom tijelu u ustanovi.

Mjesto, nadnevak: SPLIT, 12.1.2021.

Potpis studenta/studentice: [signature]