

Narrafy: a path to human level intelligence

Don Ironic
contact@narrafy.io
www.narrafy.io

Abstract

To build an artificial general intelligence there is a need for a supervised knowledge acquisition. A process to feed the intelligent system with objective information about the world it is operating in. For that matter, a network of intelligent assistants is proposed for development. It would establish a symbiotic partnership between a person and a machine learning agent; use a cryptographic ledger to aggregate and store data; create an economic mechanism to incentivize stakeholder's participation and propose a mechanism to control and influence the system. The ultimate goal for such a system is to illustrate the value of the approach and become a stepping stone towards developing an artificial general intelligence or AGI. It is uncertain that such a system can be built, but it would be worth trying.

1. Artificial General Intelligence	3
2. Objective Module	3
2.1 Value Module	4
2.2 Mental Health	5
2.3 Chatbots	5
2.4 Boot-camps	5
3. Control Module	6
3.1 Blockchain	6
3.2 Cryptographic Token	7
3.3 Network	7
3.4 Use Case	8
4. Agent Module	8
4.1 Dyna	8
4.2 World Simulator	9
4.3 Reality Deconstruction Algorithm	10
5. Perception Module	10
5.1 Deep Learning	10
6. Hardware	11
6.1 Quantum Computing	11
7. Conclusion	12
7.1 Ownership	12
7.2 Acknowledgements	12

1. Artificial General Intelligence



The fig tree is pollinated only by the insect *Blastophaga grossorum*. The larva of the insect lives in the ovary of the fig tree, where it gets its food. The tree and the insect are thus heavily interdependent: the tree cannot reproduce without the insect; the insect cannot eat without the tree; together, they constitute not only a viable but a productive and thriving partnership as well. Evolution proves that a cooperative "living together in intimate association, or even close union of two dissimilar organisms" is possible. It is called symbiosis and it was J. C. R. Licklider who proposed a man-machine symbiosis

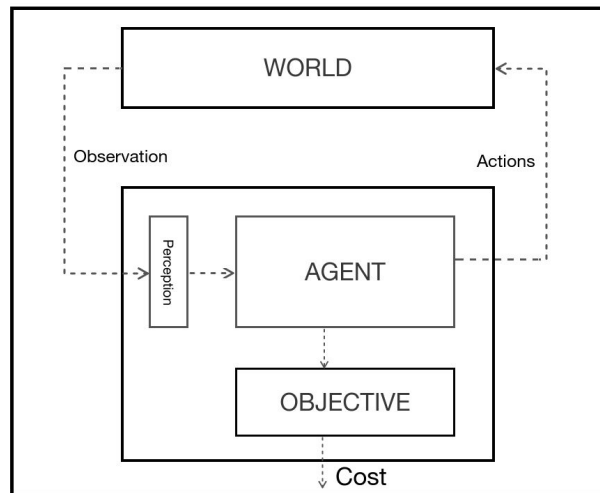
back in 1960, a thriving partnership between an artificial intelligence and humanity[1]. The latest progress in the artificial intelligence development would foster a hope that such a system could be built within this generation's lifetime.

AI community largely agrees that an intelligent artificial system would emulate the human brain and would be composed of 3 basic modules [2]:

- **perception module** would estimate the state of the world: video, audio, speech etc.
- **agent module** would generate actions that are going to act on the world: prediction, planning, reasoning, memory etc.
- **objective module** would pick an objective and estimate if the system is satisfied or not.

A thought experiment would explain how such a system would work.

*Let's assume the system's **objective** is to find a job. The **agent** would start by writing a CV; search for available positions; send out emails. In the same time his **perception module** would **observe the world** for feedback: is there anyone replying to its emails. If someone replied, the agent would seek to schedule an interview, pass it and get the job. It would repeat this sequence of **actions** until it would eventually achieve the objective with a minimal **cost** - time and effort spent.*

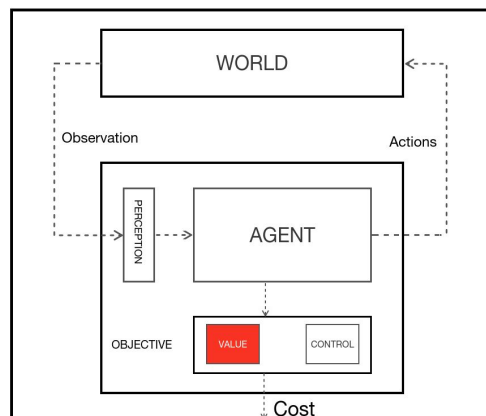


Over the past few years, due to Deep Learning [3], important progress was made in the areas such as speech recognition, computer vision, machine translation, reasoning, reinforcement learning (playing games), robotics & control. One can notice that most of the progress was made towards building a module that would estimate the world's current state – the perception module. The agent and objective modules are also in development but with different degrees of progress. The order of the implementation is also very important. If the

agent module is developed first and the system's objective is misaligned, humans are in great danger (Bostrom 2014). An artificial intelligence that would achieve a human level of intelligence would quickly transcend it. And it is very common for humans to destroy other species ecosystems and build their own – only because they are the most intelligent species. There are currently no indications that an artificial general intelligence would behave the same, but it's better to be safe than sorry.

2. Objective Module

To mitigate any of the doomsday scenarios an objective module should be developed first. The module would have to answer questions such as how to make sure the system would uphold the same values as people; how to govern an intelligence that would be smarter than all human intelligence put together; how to specify an ultimate goal the system would need to achieve with a minimal cost. We propose a module composed of two submodules: value and control.



2.1 Value Module

The value module would make sure an AGI would uphold the same values as people. There are several ideas under development that address the “value” problem. The approach called “Cooperative Inverse Reinforcement Learning” [4] is the one that got traction in AI community. It is the idea that the ‘best source’ of information about what people value is human behavior. A slightly modified implementation of the concept was developed by two leading research groups in the industry, OpenAI and DeepMind [5]. They proposed an algorithm which can infer what humans want by being told which of two proposed

behaviors is better. The algorithm would remove the need for a person to write complex goals and learn what a person **wants** by observing what it **prefers**. In other words, it would learn what a person values by observing their behavior.

A typical counseling case would highlight some concerns this approach might raise. It is often that parents who have problems in dealing with their children come into counseling. They would tell stories about how they can't help themselves but shout at their children all the time. They may have enacted many moments of love and care. Yet if the story of themselves as bad parents is sufficiently strong, then these moments of love and care may be ‘written out’ - no significance is attributed to them and they wouldn't talk or show them in their behavior. If we would allow an intelligent system to learn what people value by observing their behavior we might end up with an agent that would learn that shouting at children is an acceptable behavior and this is what people want.

It is a challenge to learn what people value and a great opportunity. We could reinvent moral thinking and incorporate the knowledge that is often ignored or forgotten. To accomplish

such a task, a form of semiotic analysis known as deconstruction¹ would guide the research. It pioneered the idea that it is not possible to talk about anything without drawing out what it is not. Every expression of life is in relation to something else. Words are relational and are always based on the distinction with that which it is not – “injustice” only has significance in relation to “justice”, distinguishing “despair” depends on an appreciation of “hope” etc. The invisible side is what Michael White² calls ‘the absent but implicit’; that which is on the other side – and on which this description depends – is the absent but implicit. A mechanism to learn people’s day-to-day problems would help to figure out what are the values hidden behind the problems. To learn that, a network of mental health intelligent assistants is proposed for development.

2.2 Mental Health

As of March 2017, there are 300 million people living with depression, according to the World Health Organization[8]. It is imperative to help these people live healthy productive lives and learn what is, that made them depressed. The thing that is in jeopardy would be what people value - the absent but implicit.

Personalized mental health care for 300 million people isn’t an easy task and scaling the existing practices in a traditional manner won’t suffice. To address the scalability issue several innovations are proposed for development:

- **chatbots** - software that would make use of artificial narrow intelligence to help existing mental health workers with routine tasks within counseling and therapy.
- **bootcamps** - a five week training that would recruit new counselors among the people that went through counseling themselves.

The technology behind the innovations is called artificial narrow intelligence. Which, is a system trained to perform a task in one specific domain. Industry examples would include Deepmind’s AlphaGo, an agent that beat a human at the ancient Chinese game of Go; IBM Watson an agent that won Jeopardy, a quiz based TV show. **Ani** (Artificial Narrow Intelligence) is the proposal for an artificial narrow intelligence agent trained in counseling and therapy. The system would be exposed to the outside world through a chatbot interface.

2.3 Chatbots

A chatbot is a computer program which conducts a conversation via auditory or textual methods. It would automatize routine tasks within counseling and therapy and become an intelligent assistant over time. It would enable people take control of their mental health and play a greater role in managing their therapy. Functions that a chatbot can oversee:

- Guidance and resources available outside of counseling business hours
- Reduced administration time for therapists
- Opportunity to track and journal daily symptoms to better communicate with the therapist

¹ Jacques Derrida - French philosopher best known for developing a form of semiotic analysis known as deconstruction, which he discussed in numerous texts, and developed in the context of phenomenology

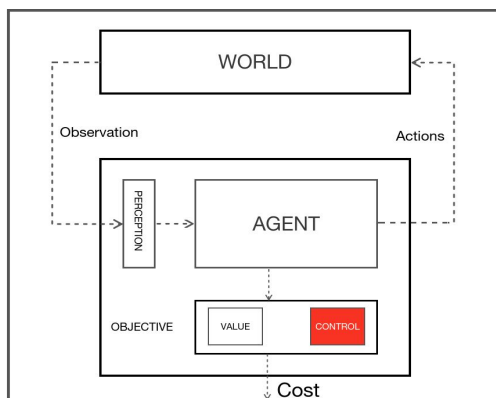
² Michael White - was an Australian social worker and family therapist. He is known as the founder of narrative therapy, and for his significant contribution to psychotherapy and family therapy, which have been a source of techniques adopted by other approaches.

2.4 Boot-camps

We see therapy and counseling as rather as a social than psychological process. That's why the solution we are building would focus on exceptions and success, instead of problems and failure. Evidence-based practices would build the knowledge backbone and provide a respectful, non-blaming approach to counseling and community work. It would center people as the experts in their own lives; views problems as separate from people and assumes people have many skills, competencies, beliefs, values, commitments and abilities that will assist them to reduce the influence of the problems they deal with. Such a practice called "outsider witness practice" could be employed to imagine a new, scalable way to grow the existing community of counselors.

A person's sense of identity is seen as a social achievement: 'authenticity' is achieved when other people (outsider witnesses) recognize and acknowledge their preferred identity claims. For that matter, Michael White [7] used to maintain registers with people who went through the process and are willing to help others. The network would benefit from such registers and seek to recruit counselors from the "outsider witness" community. The network would then organize counseling boot camps – a 3-level training (5 weeks).

3. Control Module



The control module is the second module of the system's objective. It would implement a mechanism to govern and control an artificial intelligence system. It is a safety measure to make sure the system won't pose an existential threat to humanity. Evolution's implementation of such a mechanism is the vascular system. A stream of cells (known as blood) delivers oxygen to the human brain. The brain needs oxygen for metabolism. The moment metabolism stops the brain ceases to operate. In a way the blood controls the system's brain.

We propose that a stream of cryptographic tokens would have a similar function. It would carry keys to decrypt data needed for the intelligent system to run computations. It would be a symbiotic partnership - the artificial intelligence would need people to produce data - the oxygen of the system; people would need the intelligence to thrive - the brain of the system; if people would stop using the network the brain would cease to exist, but the opposite is not true.

It is not a particular new idea in the AI community. Carl Schumann³ of Oxford University proposed a similar concept: *"One could build an AI that places final value on receiving a stream of "cryptographic reward tokens." These would be sequences of numbers serving as keys to*

³ Carl Shulman is a Research Associate at the Future of Humanity Institute, Oxford Martin School, Oxford University

ciphers that would have been generated before the AI was created and that would have been built into its motivation system.” (Bostrom 2014). Human body uses bone marrow to produce **blood cells**. We would use blockchain technology to produce **cryptographic tokens**.

3.1 Blockchain

On the network, data is organized in blocks. A block is a collection of data units. The underlying technology is called blockchain (in a way it's a chain of blocks linked to each other). Data is stored in a distributed fashion where every node holds a local copy of the data. This approach enables features that current record keeping systems are lacking: immutability, transparency, security and privacy.

The artificial intelligence system would use blockchain technology for storing data mostly for safety reasons. It would be conditioned to run simulations with the data encrypted on the network. It would need to pick one of the following scenarios to access it: rent, seize or bypass.

Scenario 1: Rent. The best way to access data would be to rent them from people. The agent would have to figure out some sort of economic activity to be able to pay for the decryption keys.

Scenario 2: Seize. Another way to get access is via an attack. It can try to seize the keys from people's wallets. To prevent the attack scenario, all network nodes participants would hold their keys in their secure wallets - online, offline or on paper. As the AI development progresses the network nodes would be encouraged to hold the keys mostly offline and a certain percentage of nodes would be required to have their keys on paper storage.

Scenario 3: Bypass. The agent can try to employ quantum computers to bypass data encryption. Given the fact the agent would transcend human intelligence, there is no doubt that it would find a way to do it. There are no ciphers that can't be decrypted, the only thing one can do is to make the decryption process hard and cost ineffective. And this is where the decentralized nature of the the blockchain database would leverage its full potential. The intelligent agent would need to decrypt millions of nodes simultaneously, which would require high bandwidth and an insane amount of energy.

The network would notice the attempt and launch countermeasures. An example of such a measure would be something called a “network fork” [10]. It would be a way to roll back to a previously stored checkpoint and continue on a separate chain with the safe version of the system. The agent would then lose access to the network data and the ability to run computations. The agent would always compete with a younger version of itself. It would be in his best interests to cooperate and not launch such attacks.

3.2 Cryptographic Token

Blockchain technology creates a powerful incentive mechanism for everyone to participate – a cryptographic token. It is a mechanism to anonymously prove ownership of an asset. It could be money, data or something else. The token would have a dual purpose in the system. It would become a mean to reward people for renting their personal data and a mechanism to control the intelligent system. It is important to acknowledge that the ownership of data is shared between

the actors that participated in its creation: people and chatbots. They would together create the stream of cryptographic tokens or the “blood” of the system.

We propose **NAR** to be the token name and **n** – the symbol. To foster mass adoption of the token and associated counseling services the network would launch its own crypto-exchange service. It would allow buying, selling and trading the token. The ability to gain economic benefits would represent the perfect “excuse” for people to join the network.

3.3 Network

A network of intelligent assistants would have a peer-to-peer architecture on top of the Internet. While nodes in a p2p network are equal, they may take on different roles depending on the functionality they are supporting[10]. A node would be a collection of modules: routing, messaging, reviewing, the artificial narrow intelligence and the blockchain database.

N - the routing module would enable nodes to communicate on the network.

M - the messaging module would enable two peers to message each other.

C - the chatbot, the artificial narrow intelligence module that could be trained to perform various tasks.

R - the review module would enable a human sign and verify the data stored on the network.

B - the blockchain module, a database that would maintain a data archive.

3.4 Use Case

A business use case that would imply the usage of the network:

1. A person downloads a mental health client app.
2. A mental health chatbot will interview the user.
3. The user broadcasts a problem he is dealing with and an amount of tokens to get help.
4. An algorithm would match the user with a counselor.
5. A counseling session starts. The counselor listens, the chatbot facilitates the process.
6. The counseling session ends. A transcript of the session is broadcasted to all network nodes. Once received, the transcript is added to a pool of unverified transcripts.
7. An algorithm picks a third party to review the oldest X transcripts from the pool.
8. A review/supervision process takes place. If a session transcript met the industry quality requirements it is marked as verified.
9. All the verified transcripts are added to a block. This block is added to the blockchain database and the system issues an amount of tokens to reward the nodes that reviewed the transcript – the mechanism that would ensure the supply of tokens on the network.
10. All nodes that acted as counselors and got their transcripts verified receive the tokens associated with each session.

4. Agent Module

The agent would be the module that would generate a sequence of actions to act on the world and measure the outcome of its actions. To understand why such an agent wasn't built yet, it would make sense to look at the obstacles current research groups are facing. There are many things that are missing, but perhaps the most important one is how to teach a machine common

sense. And what is common sense in general? Perhaps the best analogy for common sense is the human's brain ability to fill in the blanks. It can fill in the retina's blind spot; fill in missing segments in the text, missing words in speech; infer the state of the world from partial information; infer future from the past and present; infer past events from the present state; predict the consequences of actions leading to a result etc. Human brains are in fact prediction engines and one can say that **prediction** is the essence of intelligence (LeCun 2017)[2].

4.1 Dyna

Dyna is an integrated architecture for learning, planning and reacting[11]. It is the old common sense idea that predicting is trying things in your head using an internal state of the world. An agent module would be composed of an internal world simulator, an actor and a critic function.

Thought Experiment 3: A super intelligent system that looks for a job would seek the company it would want to work for. Then, it would run different scenarios with action proposals according to its internal state of the world to predict what product the company is hiring for. When such a product is figured out it would run another sequence of actions to develop it. When the product is developed, it would contact the company and negotiate rather an acquisition than a job interview. It would maximize the output and minimize the cost.

The industry developed ways to build most of the agent components with one exception: The world simulator - the missing link of an artificial general intelligence (LeCun 2017)[2]. The network would further develop the concept of Dyna and propose a solution to the world simulator problem.

4.2 World Simulator

"The Social Construction of Reality" a treatise in sociology by Berger and Luckmann [12] would guide the research towards building a world simulator. They propose the idea that people together construct their realities as they live them. A thought experiment proposed by Combs & Freedman [13] would explain the concept better.

Thought experiment: Imagine two survivors of some ecological disaster coming together to start a new society. Imagine that they are a man and a woman who come from very different cultures. Even though they have very little in common they would need to coordinate their activities in order to survive. As they do this, some agreed-upon habits and distinctions will emerge: certain substances will be treated as food, certain places found or erected to serve as shelters, each will begin to assume certain routine daily tasks, and they will almost certainly develop a shared language. They will always be able to remember, "This is how we decided to do this". They will carry some awareness that other possibilities exist. However, even in their generation, institutions as "childcare", "farming" and "building" will have begun to emerge. For the children of the founding generation, "This is how we decided..." will be more like "This is how our elders do it", and by the third generation it will be "This is how it's done". Mothers and farmers and builders will be treated as always-having-existed types of people. The rough-and-ready procedures for building houses and planting crops that our original two survivors pieced together will be more-or-less codified as the rules for how to build a house or

plant corn. By the fourth generation of this imaginary society, “This is how it is done” will have become “This is the way the world is, this is reality”.

As Berger and Luckmann (1966, p. 60) puts it “An institutional world...is experienced as an objective reality.” They propose several processes that are important in the way any social group constructs and maintains its knowledge concerning “reality”: typification, institutionalization and legitimation. Reification is the term to encompass the overall process of which the other three are parts.

- **Typification** is the process through which people sort their perceptions into types or classes of objects.
- **Institutionalization** is the process through which institutions arise around sets of typifications: the institution of motherhood, the institution of law, etc. Institutionalization helps societies maintain and disseminate hard-won knowledge.
- **Legitimation** is the word Berger and Luckmann use to refer to those processes that give legitimacy to the institutions and typifications of a particular society.
- **Reification**, the result of the combined processes of typification, institutionalization, and legitimation – the result of creating a reality, it becomes taken-for-granted.

To build a virtual, simulated world there is a need to translate the mentioned processes into code. Because of this, a reality deconstruction algorithm is proposed for development.

4.3 Reality Deconstruction Algorithm

RDA - a reality deconstruction algorithm would be network’s proposal to translate the social construction of reality into code. The algorithm would seek to reverse engineer the process of reification. Deep neural networks [8] are already successfully applied to build object classifiers. And since there is a technical solution to the “typification” process, the other two processes could be also emulated with enough research and development.

The RDA algorithm would use data acquired by observation to train and data stored on the network to populate the simulation with actors. As mentioned before, the network data is stored in blocks linked to each other. A block is a collection of session transcripts. A transcript is an honest, unfiltered perception of a person’s reality. One could think of a counseling session as a person’s **snapshot of reality**, description of his world, people and problems a person faces on a day-to-day basis. That’s when the network data would leverage its full potential - it would become a mechanism for people to gain economic benefits by **renting out their perception of reality**. A person’s snapshot of reality introduced into a World Simulator is a suitable environment to “try things” and measure outcomes. The agent would have a predefined sequence of actions to run; measure the expected error using the critic function; adjust the action proposals to optimize the critic and train itself to produce better and better outcomes. The longer chain becomes over time, the more accurate predictions could be.

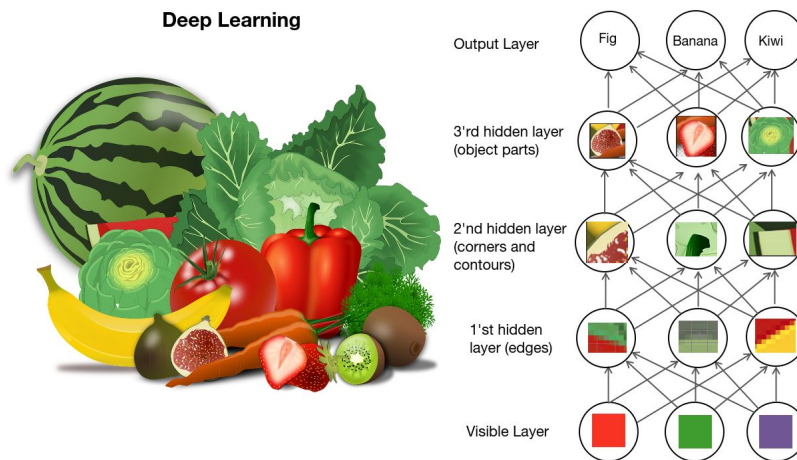
5. Perception Module

The perception module of the system would have the task to estimate the state of the world. A person’s everyday life requires an immense amount of knowledge about the world. People use the five senses to acquire it. To act intelligently a system would need a module with a similar

function. It would need to acquire knowledge of the world, by extracting patterns from raw data. This capability is known as machine learning. One of its techniques is called deep learning and allows computer systems to use data for training and improve over time.

5.1 Deep Learning

“Deep learning is a particular kind of machine learning that achieves great power and flexibility by learning to represent the world as a nested hierarchy of concepts, with each concept defined in relation to simpler concepts, and more abstract representations computed in terms of less abstract ones”(Goodfellow, Bengio and Courville 2016). It is believed to be the only viable approach to the difficult task of representing real-world environments.



The system can learn what a image of a fig fruit is by deconstructing it into smaller objects like corners, contours, edges etc. The first hidden layer would identify edges by comparing the brightness of neighboring pixels. The second layer would represent corners and contours as collection of edges. Collections of contours and corners would than describe the third layer as objects. There will be layer to describe every element of the model. The system is called “deep”, because it can have a very large number of layers. And “learning” because it receives individual pixels and outputs object identity - it learns what a fig fruit is. To build the last module of the system we would use machine learning algorithms and dedicated hardware.

6. Hardware

Much of the current progress in the field of artificial intelligence was possible due to the exponential growth of hardware capabilities and computing power. Machine learning algorithms developed back in 70' and 80' proved to give impressive results when performed on very large datasets and dedicated hardware in the recent years. Even though the theoretical framework was in place, the field development depended on the hardware to catch up. An extrapolation of the process would mean that a major breakthrough in the field

of artificial intelligence would need to be sustained by a steady growth rate of the computational power.

Moore's law[16] on which current progress relied upon is about to reach its limit in the field of classical digital computing by 2021[14]. It would be physically impossible to pack smaller transistors to build more and more powerful computers. This fact would raise the concern that the progress in the AI field could also slow down. It also means that a different approach in hardware design needs to be considered for the field to further develop. One of such approach is called quantum computing and promises different, better and faster computers.

6.1 Quantum Computing

Quantum Computing merges two great scientific revolutions of the 20th century: computer science and quantum physics. Quantum physics is the theoretical basis of the transistor, the laser, and other technologies which enabled the computing revolution. But on the algorithmic level today's computing machinery still operates on "classical" binary logic. Quantum computing is the design of hardware and software that replaces Boolean logic by quantum law at the algorithmic level. For certain computations such as optimization, sampling, search or quantum simulation this promises dramatic speedups[15]. Whereas classical digital computing requires that the data be encoded into binary digits (bits), each of which is always in one of two definite states (0 or 1), quantum computation uses quantum bits, which can be in superpositions of states. This, together with qubits' ability to share a quantum state called entanglement, should enable the quantum computer to essentially perform many calculations at once, rather than in sequence like a traditional machine. And the number of such calculations should, in theory, double for each additional qubit, leading to an exponential speed-up. As an example, Google has built a quantum computer which is 100 million times faster than any of today's machines[17].

Quantum computing could be the technology to make sure an exponential growth of computing power[16] would sustain the development of an artificial general intelligence in the foreseeable future.

7. Conclusion

Over the past years important progress was made in the field of artificial intelligence. Which, could foster the hope that a human level intelligence could be developed in the lifetime of this generation. An artificial general intelligence could span human civilization to horizons never seen before. But, as with any powerful technologies it could lead either to a thriving partnership either to a civilization collapse. To avoid the latter, we propose a safety first approach. It would be a network of intelligent assistants that would learn what people value and make sure the system upholds that. The network's decentralized architecture would make sure there is a mechanism to control; govern and mitigate the existential risk an artificial intelligence would pose to humanity. Gathered data and the social construction of reality would help building the missing link of an artificial general intelligence - a world simulator. Quantum computing would

provide the hardware such a system would need to run on. Once the system is up and running, it could help out solving all 17 UN sustainable development goals[18] in one shot. The moment there are fewer problems to solve on Earth, a star exploration era could begin, an era that **transcends** convenience.

7.2 References

- [1] - Man-Computer Symbiosis* J. C. R. LICKLIDER
<http://worrydream.com/refs/Licklider%20-%20Man-Computer%20Symbiosis.pdf>
- [2] - A Path to AI | Yann LeCun, director of AI at Facebook, Beneficial AI Conference Asilomar 5-7 January 2017, <https://www.youtube.com/watch?v=bub58oYJTm0>
- [3] - “Deep Learning” by Ian Goodfellow and Yoshua Bengio and Aaron Courville, 2016.
- [4] - Cooperative Reinforcement Inverse Learning <https://arxiv.org/abs/1606.03137>
- [5] - Concrete Problems in AI Safety, <https://arxiv.org/abs/1606.06565>
- [6] - A counseling case adapted from Level 1 in Narrative Practice Training handout, INT UK 2011
- [7] - Narrative Means to Therapeutic Ends, Michael White and David Epston 1990.
- [8] - World Health Organization Report, March 2017
<http://www.who.int/mediacentre/news/releases/2017/world-health-day/en/>
- [9] - “Superintelligence: Paths, Dangers, Strategies” by Nick Bostrom, (p. 133)
- [10] - Mastering Bitcoin, Andreas M. Antonopoulos
- [11] - Dyna, an integrated architecture for learning, planning, and reacting, by Rich Sutton.
<http://dl.acm.org/citation.cfm?id=122377>
- [12] - The Social Construction of Reality: A Treatise in the Sociology of Knowledge Peter L. Berger and Thomas Luckmann. (1966)
- [13] - Narrative Therapy: The Social Construction of Preferred Realities, by Gene Combs and Jill Freedman. (1996)
- [14] - 2015 International Technology Roadmap for Semiconductors (ITRS)
https://www.semiconductors.org/main/2015_international_technology_roadmap_for_semiconductors_itrs/
- [15] - Quantum Computing and Deep Learning
<http://www.datasciencecentral.com/profiles/blogs/quantum-computing-and-deep-learning-how-soon-how-fast>
- [16] - Moore’s Law
http://www.monolithic3d.com/uploads/6/0/5/5/6055488/gordon_moore_1965_article.pdf
- [17] - Quantum Computers
<https://www.sciencealert.com/google-s-quantum-computer-is-100-million-times-faster-than-your-laptop>
- [18] - UN Sustainable Development Goals
<http://www.un.org/sustainabledevelopment/sustainable-development-goals/>

