An Overview of the GUT-AI Foundation:

Vision for an Ecosystem of Concepts and Implementations

Ioannis Kourouklides*, Kleitos Alexandrou*

* equal contribution

Whitepaper

GUT-AI Initiative

July 2023

“In a dynamic, decentralized system of individual choice and responsibility, people do not have to trust any authority but their own.” — Virginia Postrel

Abstract

This whitepaper presents the vision, mission, and overarching role of the GUT-AI Foundation. The Foundation aims to promote the research, development and eventually the deployment of user-friendly, human-centred and developer-friendly Artificial Intelligence (AI) systems for the betterment of humanity through an Ecosystem of Concepts and Implementations (ECI). The Foundation recognizes the potential of AI to revolutionize numerous industries, ranging from Healthcare and Education to Financial Services and Self-Driving Cars. However, the development of such AI systems poses significant challenges and impediments, such as multiple single points of failure, lack of interoperability, and lack of user adoption. Therefore, this whitepaper proposes a multidimensional approach to promote a whole ecosystem that has the ability to overcome such challenges. Primarily, the Foundation will encourage research into AI systems that are accessible, intuitive, and ready-to-use. The research will focus on proposing AI system architectures that meet the needs of the users, while they address their pain points in order to enhance both the User Experience (UX) and Developer Experience (DX). Furthermore, the Foundation will focus on promoting the adoption of best practices and Optional Open Standards for AI development and deployment that is automated, cost-effective, and scalable. For instance, these best practices will include the use of modular architectures, microservices, and containerization. By adopting these practices, the Foundation aims to enable interoperability and reuse of AI components. In addition, the Foundation envisions creating a marketplace, which will enable buyers to discover and adopt AI Solutions that meet their needs and preferences. The marketplace will also provide sellers with opportunities to showcase their AI Solutions, reach out to potential buyers and receive feedback from them. Finally, the Foundation will leverage emerging technologies such as Blockchain and Decentralized Autonomous Organizations (DAOs) to enhance the transparency, security, and trustworthiness of AI systems, while incentivizing innovation and collaboration among humans.

Table of contents

1 Introduction

1.1 Motivation

The Ecosystem is itself a concept that refers to the interconnected network of exchanging *ideas, information, and knowledge* that shapes and informs the way we think, act, and interact with each other and the world around us. It includes everything from scientific discoveries and technological advancements to cultural beliefs and artistic expressions, while it is envisioned to be constantly evolving and growing in response to new insights, experiences, and interactions.

At its core, the Ecosystem is a reflection of the complexity and variety of **human experience**, both individually and as a whole. It encompasses a vast array of disciplines, perspectives, and traditions, and it is shaped by a multitude of factors, including social, political, economic, and environmental conditions. As such, it represents a rich and dynamic resource for human innovation, creativity, and progress.

One of the most important implications of the Ecosystem is that it will enable us to build on the achievements of others, to learn from their successes and failures, and to collaborate across boundaries and disciplines. By sharing information and insights, we can create new knowledge and generate innovative solutions to complex problems. For example, the development of medical equipment, water purification systems, emergency communication systems, airbags, seat belts and other life-saving technologies has been made possible by the *collaborative* efforts of Researchers and Inventors from around the world.

Moreover, the Ecosystem will encourage us to challenge our assumptions, biases and belief systems in order to further expand our horizons. It exposes us to new perspectives and ways of thinking, and it fosters an appreciation for challenging the status quo. By engaging with new ideas and experiences, we can broaden our understanding of the world and become more empathetic, compassionate, and open-minded individuals.

However, the Ecosystem also presents some challenges and risks. For instance, it poses a significant risk to **privacy**, as the proliferation of digital technologies and the widespread sharing of information has made it easier than ever for personal data to be collected, analysed, and exploited without individuals' consent. This can result in a range of negative consequences, including identity theft, financial fraud, and reputational damage, as well as potential abuses of power by *untrustworthy* governments and corporations.

Another risk of the Ecosystem is the potential for **unintended consequences**. New ideas and technologies can have both positive and negative impacts, and it is often difficult to predict or control their effects. For instance, the widespread use of social media has brought many benefits, such as increased connectivity and access to information, but it has also been linked to issues such as dullness, censorship, the amplification of echo chambers and ultimately, leading to an oligopoly or even a monopoly in the marketplace of ideas.

In spite of these challenges, the Ecosystem will remain a powerful force for human progress and innovation. Further to the future, it is likely that the Ecosystem will continue to evolve and expand, driven by new discoveries, technological advancements and social changes. However, the responsibility will be up to us as individuals, communities and societies to (a) address the associated challenges and risks, and (b) to ensure that reaping the benefits of such an Ecosystem is implemented in a rational, equitable and responsible manner.

1.2 Whitepaper structure

This document is *not* intended as a deep technical whitepaper, but as an introduction to the vision of the GUT-AI Foundation targeted towards a general audience. In order to put things into perspective, a brief structure for this whitepaper is outlined in this Section. In Chapter 2, a comprehensive picture of the current situation is presented. In this way, readers from a diverse background are in a position to adequately comprehend the subsequent Chapters. Furthermore, Chapter 3 focuses on an extended problem specification suitable for background reading and for identifying the market gaps that the proposed Ecosystem intends to fill in. Chapter 4 highlights the proposed function, design and implementation of the Foundation and the Ecosystem. Chapter 5 presents how this work fits in the overall *initiative*, and finally, conclusions are drawn, including the importance of this work.

1.3 Legal note

Unless otherwise stated, the authors make the content of this document available under **Creative Commons CC0 1.0 (Public Domain)**. Please refer to the legal text of CC0 1.0[1] for details. The authors have intentionally and deliberately used this way to release this document because of their beliefs and ideology on how research content should be released. The authors also invite others following their example and welcome feedback on this move. In case the reader wishes to do so, he can *cite* this document similar to how OSF[2] preprints are cited. It should be noted that this whitepaper is part of the **GUT-AI Initiative**[3] (see Chapter 5). This document might also contain copyrighted work by others, in which case, the copyrights belong to their respective copyright owners.

[1] https://creativecommons.org/publicdomain/zero/1.0/
[2] https://osf.io/bxw4h
[3] https://gut-ai.org/

The views, thoughts, opinions and beliefs expressed in this document are solely those of the authors and do not necessarily reflect the official policy or position of the GUT-AI Foundation.

Nothing in this whitepaper is an offer to sell, or the solicitation of an offer to buy, any tokens. The GUT-AI Foundation is publishing this whitepaper solely to receive feedback and comments from the public. If and when the Foundation offers for sale any tokens (or a Simple Agreement for Future Tokens), it will do so through definitive offering documents, including a disclosure document and risk factors. Those definitive documents also are expected to include an updated version of this whitepaper, which may differ significantly from the current version.

Nothing in this whitepaper should be treated or read as a guarantee or promise of how the Foundation's business or the tokens will develop or of the utility or value of the tokens. This whitepaper outlines current plans, which could change at its sole discretion, and the success of which will depend on many factors outside the Foundation's control, including market-based factors and factors within the data and cryptocurrency industries, among others. Any statements about future events are based solely on the Foundation's analysis of the issues described in this whitepaper. That analysis may prove to be incorrect.

2 Situation

2.1 Users and their needs

The users that would benefit the most from the use of AI, even though they might not fully realize its potential yet, typically fall into three categories:

(1) **Conventional Small and Medium Enterprises (CSEs)**

(2) **Conventional Large Enterprises (CLEs)**

(3) **Innovation-Driven Enterprises (IDEs)**, irrespective of their size (i.e. startup to mature corporation), as defined in [2]

(4) **Individual Researchers**, including Citizen Scientists

(5) **Individual Developers**

As explained in [2], a major trait of both the CSEs and the CLEs is their focus on specific *niche* markets (e.g., typically serving a specific *location* or specific *demographics*). Typical examples of CSEs include local e-shops selling pieces of clothes, shoes or accessories, which do make use of technology and do collect data, but they do *not* have the internal capacity or the intention to innovate. They typically outsource the innovation part to an IDE and focus on for the most part on “non-tradable” functions that cannot be outsourced to someone somewhere else, since that would entail *decrease* in the UX and the customer satisfaction, which is where the **expertise** of the CSEs lies (i.e. what they do best). Similarly, the same applies for CLEs, but with a much larger headcount and orders of magnitude higher revenue. A typical example of CLE could be a company in the Energy industry (e.g., Oil and Gas, Renewables, Nuclear power) or the Hospitality industry (e.g., hotel chain, restaurant franchise). When an organization does not have the maximum time to focus on its own expertise, then this increases the chances of failure (see also Section 2.2).

As a side note, for reasons explained in [2], the IDEs are intentionally *not* termed 'technology-driven' since the innovation might not necessarily lie on the specific technology used. However, typical examples could include Tech startups (see Section 4.10 for specific use cases) and arguably, companies similar to Big Tech.

2.1.1 The need for AI solutions

The need for AI solutions regarding enterprise users (i.e. CSE, SLE, IDE) is pretty much self-evident, since there is a vast array of enterprise companies that incorporate AI in their product offering, as shown in Fig. 1. The existence of all those AI companies proves that a large market for AI solutions already exists. Each of these product offerings consists of one or more AI systems packaged as AI solutions that solve practical, real-life business problems in order to address the specific needs and demands of the enterprise user.

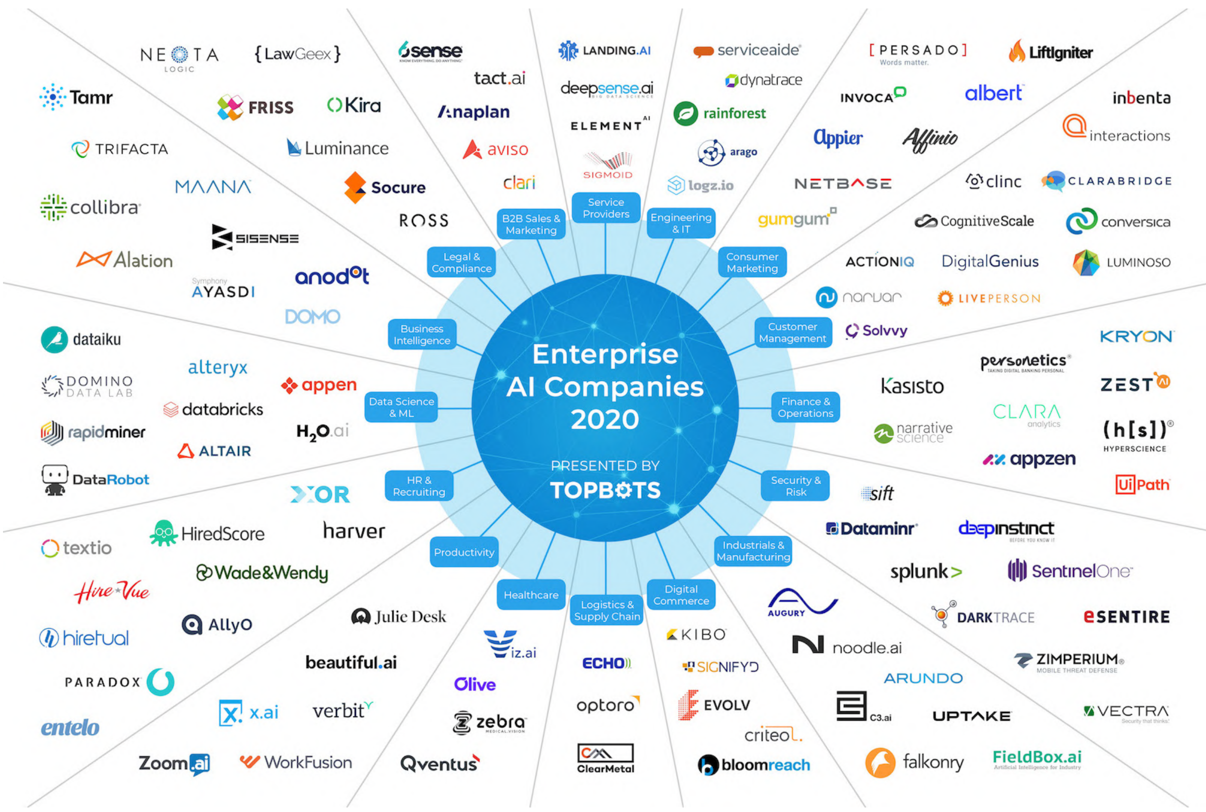


Figure 1 - Top Enterprise AI companies [5].

2.1.2 Data: AI Solutions in not all you need

All of the aforementioned examples of enterprise users have something in common: they collect **data** (or could collect data) through a piece of technology, either hardware or software or both. These data can be used in a variety of ways, ranging from personalizing the experience and improving the satisfaction of their customer to extracting insights for the general market and using them to launch new products and services. At the risk of stating the obvious, this is where AI solutions (Section 2.1.1) comes in, so that it can meet the particular *needs* of the users, while addressing their specific *pain points*. Currently, various AI solutions exist that attempt to make use of AI for tackling the aforementioned needs, as described above.

Even though some of those AI solutions might be using *fancy* cutting-edge technologies and *super-duper* AI algorithms, the real (business) value does not necessarily lie there. The real value lies in the data collected, how clean they are, how up-to-date they are, how relevant they are, how useful they are, how standardized they are, which data are missing, what new data need to be acquired and how, etc. This is arguably a main shortcoming (or “failure”) of a company ironically called OpenAI when it released ChatGPT (or other versions of GPT), in contrast to Google, for example, which does collect a gazillion of our data on a daily basis, which obviously raises other concerns, such as lack of *privacy*, *transparency* and the likes (as described in Section 3.1). This has actually given rise to companies that specialize in collecting such data (e.g., image data for autonomous vehicles [6 - 8]) by themselves or creating a whole **marketplace** for Self-Driving Car data (also called a data exchange or data aggregator). Therefore, AI Solutions is not all you need as an enterprise user. You also need data that are specific to your company, but sometimes data that cover your whole industry, not just one company [6].

2.1.3 AI infrastructure: Data is not all you need

Even if an enterprise user has ensured the necessary data, this is still not enough. Since there is a multitude of independently developed AI Solutions, plus a multitude of non-AI software products (as shown in Fig. 2), then this gives rise to the need for **integrating** all of the aforementioned products (AI and non-AI software). Naïvely, one might think that simply connecting Application Programming Interfaces (APIs) will resolve this issue, but it is not all sunshine and roses. These are products that were *not* designed or implemented with the assumption that they should be easily integrated with each other.

As an illustration, consider a **Customer Relationship Management (CRM)** software that is closed-source. Since such a CRM restricts access to its source code, limiting the ability of a third-party provider to integrate its AI solutions effectively, since it has to rely on the CRM vendor for any modifications or customizations, which can be time-consuming, costly and may not even align perfectly with the interests of the vendor. Therefore, a closed-source CRM often lacks flexibility and extensibility, making it challenging to integrate cutting-edge AI solutions. The enterprise client using the CRM may face limitations in harnessing the full potential of AI due to vendor-imposed restrictions, thus hampering the ability to stay competitive and to cater the specific needs of the enterprise client. Additionally, closed-source CRM vendors tend to enforce strict control over their software, which can lead to restrictions on external modifications or extensions. This inhibits the third party's ability to fully optimize their AI solutions within the CRM environment, thereby limiting the potential benefits for the enterprise client. This results in a lose-lose situation for both the third party offering the enterprise AI solutions and the enterprise client, who might actually have to stop using the closed-source CRM in case the lack of effective integration of AI solutions potentially jeopardizes the competitive edge of the enterprise client.

As a result, the need for integrated solutions gives rise to the impediment of *multiple single points of failure* (as described in Section 3.1), which itself gives rise to the need for a common (or at least interoperable) **AI infrastructure**. Currently, various centralized (Section 2.1.4) and decentralized (Section 2.1.7) platforms exist that attempt to create some sort of AI infrastructure, as described below. However, the issue of *interoperability* (as described in Section 3.1) remains.

2.1.4 DX: AI infrastructure is not all you need

Even if the problem of AI infrastructure is addressed through open-source, this is still not enough. The lack of good **Developer Experience (DX)** can have a detrimental impact on the adoption and effectiveness of open-source AI solutions. DX refers to the ease and efficiency with which Developers can understand, use, and contribute to a software project. In the context of AI solutions, a poor DX can discourage Developers from engaging with open-source projects, resulting in limited community involvement, slower innovation, and fewer contributions. Without an intuitive and well-documented DX, Developers may struggle to integrate, customize, or extend AI solutions, thus leading to suboptimal implementation, wasted time, and plenty of frustration.

A strong DX, on the other hand, fosters collaboration, empowers Developers to create and share enhancements, and encourages widespread adoption. Ultimately, by prioritizing DX, open-source AI solutions can attract a vibrant community, drive rapid advancements, while delivering more valuable and accessible solutions to organizations. However, a further issue that arises is that of censorship and bureaucratic hegemony due to potential centralization (as described in Section 3.1), which itself gives rise to the need for controlling the physical infrastructure.

2.1.5 Physical infrastructure: DX is not all you need

As extensively explained in Section 4.3, the lack of control over the physical infrastructure can have significant drawbacks when it comes to utilizing open-source AI solutions. Such solutions typically require substantial computational resources to run efficiently, such as powerful servers or specialized compute resources. Without control over the physical infrastructure, enterprise clients may face limitations in terms of scalability, performance, and customization options. They may rely on third-party providers or cloud platforms, which can introduce potential bottlenecks, security concerns, and dependency risks. Moreover, lack of physical control hampers the ability to optimize hardware configurations for specific AI workloads, thereby hindering overall system performance and cost efficiency. Therefore, not having control over the physical infrastructure can impede the seamless implementation and optimal utilization of open-source AI solutions by the enterprise client, thus limiting their potential impact and effectiveness.

So, is physical infrastructure all you need? The answer is still no. Any of the following can still contribute to the failure of DEI organizations:

- Lack of separation of concerns (Section 2.2.1)
- Unnecessary replication of work (Section 2.2.2)
- Misaligned interests (Section 2.2.3)
- Delays on the product roadmaps (Section 2.2.4)
- Lack of user-friendliness and holistic approaches (Section 2.2.5)
- Lack of plug-and-play solutions (Section 2.2.6)
- Lack of self-correcting mechanisms (Section 2.2.7)
- Lack of a collaboration platform among organizations (Section 2.2.8)

2.1.6 Centralized platforms for AI

Regarding a user who is an Individual Developer (Section 2.1), various open-source centralized platforms for AI infrastructure exist, each with its own merits and demerits, including:

(1) TensorFlow Hub [11]

(2) PyTorch Hub [12]

(3) HuggingFace [13]

(4) Kubeflow [14]

(5) MLflow [15]

(6) Airflow [16]

(7) Seldon Core [17]

(8) BentoML [18]

(9) KServe [19]

It should be highlighted that it does not necessarily mean that the one is a (direct or indirect) competitor to the other. For instance, some of them are repositories of pretrained models that are ready to then be used, retrained and deployed anywhere. In contrast, other ones are platforms for Data Workflow Management and High-Performance Model Serving.

2.1.7 Decentralized platforms for AI

Regarding a user who is an Individual Developer (Section 2.1), various open-source *decentralized platforms* (dPlats) for AI infrastructure exist (not to be confused with AI for blockchain, which the reverse), each with its own merits and demerits, including:

(1) SingularityNET [21]

(2) DeepBrain Chain [22]

(3) OpenMined [23]

(4) Ocean Protocol [24]

(5) AICHAIN [25]

(6) Aigang [26]

(7) NeuroChain [27]

(8) AIWORK [28]

It should be highlighted that it does not necessarily mean that the one is a (direct or indirect) competitor to the other. For instance, some of them follow a bottom-up approach, i.e. mesh disparate elements into a collective system. In contrast, other ones follow the opposite, i.e. a top-down approach.

2.2 Why Tech startups and other organizations fail

Apart from not having the maximum time to focus on its own expertise (as explained in Section 2.1), an IDE in particular has some *additional* reasons why it might fail. Therefore, one of the main issues that this whitepaper addresses is the situation of **high failure rates** among IDE organizations. The rationale behind this focus will become more apparent in Chapter 4, which describes the solution.

In order to address 'success', one has to firstly define some evaluation metrics (business, technical or otherwise), since the concept is completely meaningless without such metrics. Unfortunately, when it comes to organizations (either a startup or a particular division within a large, mature enterprise), it is very difficult to derive a specific formula for success.

Typically, various metrics would need to be defined within the organization. In this Section, the authors define success using three summarized, generic and possibly vague metrics:

- M1: Success in managing **innovation** in an IDE
- M2: Success in a **sustainable growth** (not only in terms of revenue)
- M3: Success in mass adoption, Product-Market Fit and **Product Management**

There are various studies and data that can support the use of the aforementioned metrics, such as the CB Insights [3], as shown in Fig. 2. For instance, the number one reason, along with seven more reasons in the Top 10, are either directly or indirectly related to metric M3. The number two reason (i.e. financials) is related to metric M2, while the number three (i.e. the team itself) is related to metric M1. Therefore, all reasons why a startup fails can arguably be clustered into the three metrics listed above.

Figure 2 - Top 10 reasons why startups fail [3].

Undeniably, all startups that no longer exist have failed. Given the three metrics mentioned above, one can also argue that all startups that have been acquired before Series A faced a “soft” failure, from the perspective of the startup itself (but not necessarily from the perspective of the Investors and the rest of the Shareholders). Similarly, a lot of divisions or organization within large, mature enterprises (e.g., Big Tech) have also arguably failed. For instance, one can claim that all the AI divisions of Alphabet (such as Google DeepMind) and Amazon Science (such as Amazon Alexa) have failed, especially when it comes to competing against Generative AI startups (such as OpenAI, which released ChatGPT and other GPT versions). The root cause for each specific division can vary. For instance, the organization can be too *bureaucratic* to innovate (thus affecting metric M1), the interests of the individuals and the organization as a whole might be misaligned due to the *lack of incentives* when it comes to allocating financial compensation (thus affecting metric M2), the *lack of user-friendliness* when it comes to the products they release (thus affecting metric M3) or any combination of them.

2.2.1 Lack of separation of concerns

One of the main claims made by the authors of this whitepaper is that there needs to be a **'separation of concerns'** when it comes to an ecosystem that is designed to deliver Tech products (either software, hardware or a combination). In Software Engineering, separation of concerns is an Engineering principle that advocates separating a system (or application) into multiple, distinct parts (or units) with each one addressing a separate concern or aspect of the system, while there is minimal overlapping between the functions of the individual parts. The goal of separation of concerns is to make the system more *modular* and easier to understand, maintain, scale and modify over time, while enabling Developers, Designers and Testers to focus on modifying specific aspects of the system *without* having to understand the entire system at once and *without* affecting other parts of the system. The authors argue that this Engineering principle should also be applied to the *management* and *implementation* of the whole Ecosystem and not just when it comes to writing code. The argument made in this whitepaper is that this principle directly affects metric M1.

An easy-to-understand example of the aforementioned argument is to consider a football team (read ‘soccer’ in American English). Hiring talent is a necessary, but not sufficient condition for the football team to succeed. Consider the scenario of hiring only talented goal-keepers and nothing else. This scenario is doomed to fail. Also, consider the scenario that all specialities are hired, but one defender is placed as a goal-keeper and vice versa. Moreover, consider the scenario that all specialties are hired, they play at their right positions, but they don’t get along with each other. In other words, eleven good players do not necessarily make a good football team. The same goes for all those organizations that hire, manage and train “talent” in Research & Development (R&D). One would be surprised how often in an IDE, people with expertise in one field end up with roles and responsibilities in a field that they do not have any or how a Decision-Making process can end up being both bureaucratic and counter-intuitive.

2.2.2 Unnecessary replication of work

In addition to lack of separation within departments of an organization, there is often an **unnecessary replication of work** among organizations. For example, a lot of IDEs, both startups and non-startup, replicate a lot of their work (e.g., AI infrastructure, internal tools, pretrained models), which is typically closed-source and proprietary. In the case of startups, this work goes wasted once it burns and crashes. In the case of downsizing and elimination of company divisions, the same is also true. Alternatively, all this replication of work creates extreme inefficiencies in IDEs, which is often a hindrance for the organization to achieve its end goals in serving their customers by focusing on the actual **expertise** of the organization, instead of all the intermediate steps in order to achieve that (Section 2.1). For instance, just like there are *consortiums* for open-source Apache projects [4], there should have been something similar for AI solutions as well, instead of the aforementioned replication of work.

2.2.3 Misaligned interests and incentives

After an organization addresses the lack of separation of concerns (Section 2.2.1) and the unnecessary replication of work (Section 2.2.2), it has probably created **silos** of independent departments (and independent individuals on those departments), which operate in isolation and whose interests do not necessarily align with the interests of other departments or the organization as a whole. So it is advantageous that concerns have been separated, but those silos need something like an internal ‘glue’ in order to make them work towards a common goal. Therefore, an additional claim made by the authors is that there needs to be a *cohesion mechanism* in order to address the potentially **misaligned interests** that might occur in large corporations or startups, thus affecting metric M2.

For instance, when financial compensation becomes the primary focus, employees may prioritize short-term gains over long-term innovation. This can hinder creativity, risk-taking, and the pursuit of breakthrough ideas. Misaligned interests can also lead to a lack of collaboration and knowledge sharing among team members, as individuals may be driven to protect their own interests rather than work towards common goals. Additionally, if incentives are not properly aligned with the organization's innovation strategy, employees may be motivated to pursue incremental improvements rather than disruptive innovations. This can result in missed opportunities for transformative advancements. To foster a truly innovative culture, it is crucial to align interests and incentives with long-term goals, while providing a supportive environment for creativity and experimentation. A further example is the so-called “principal-agent problem”. In Economics, the principal-agent problem occurs when the interests of individuals (i.e. principals) and representatives deciding on their behalf (i.e. agents) are misaligned. For example, Directors may chase short-term profits, which benefit them, at the expense of long-term growth that benefits Investors.

2.2.4 Delays on the product roadmaps

Even if interests and incentives are aligned, **delays on product roadmaps** are a very common problem for IDE companies, especially those active in the Web3 space (as defined in Section 3.1.1). Such delays can be detrimental to a Web3 foundation for several reasons, which directly affect metric M3. For instance, these technologies often operate in a highly competitive and rapidly evolving landscape. Delays can result in missed opportunities, allowing competitors to gain an edge, attract users or secure partnerships.

Additionally, Web3 projects heavily rely on community engagement and investor confidence. Delays erode trust, leading to disillusionment among stakeholders, potential loss of funding and a decline in community participation. Furthermore, blockchain platforms or solutions often require coordination and interoperability with other platforms or solutions. Delays can disrupt integration efforts, hinder ecosystem development and impede the delivery of promised functionalities. Ultimately, delays on product roadmaps can undermine the foundation's growth, reputation, and long-term viability in the Web3 space.

2.2.5 Lack of user-friendliness and holistic approaches

Another common situation that directly affects metric M3 is the **lack of user-friendliness and holistic approaches**, when it comes to a specific product offering, especially in IDE companies whose only product offering is merely a Software as a Service (SaaS) dashboard. For instance, user-friendliness is crucial for ensuring a smooth user experience, thus promoting adoption and increasing customer satisfaction. If products are difficult to use or navigate, then users may become frustrated and disengaged, thereby resulting in decreased usage and potential churn.

Additionally, a lack of holistic approaches can limit the effectiveness and usefulness of a software product. Focusing only on one aspect or functionality, such as a dashboard, while neglecting other ones, can lead to compatibility issues, limited integration capabilities and ultimately, a failure to address the broader needs and paint points of the users. Also, by focusing solely on the dashboard as the primary offering, it can result in a deteriorated and fragmented user experience, thus significantly decreasing the (business) value of the software product. Furthermore, a lack of holistic thinking can stifle innovation and hinder scalability, since companies limit their potential for growth by solely focusing on the dashboard, while risking being outpaced by competitors who offer more holistic solutions.

Therefore, the lack of user-friendliness and holistic approaches can result in a fragmented user experience, decreased efficiency, and missed opportunities for innovation and growth. Adopting a *user-centric* and *holistic* mindset is essential for IDE companies to succeed in delivering valuable, intuitive and comprehensive solutions to their customers.

2.2.6 Lack of plug-and-play solutions

Even if holistic approaches are implemented, the **lack of plug-and-play solutions** can lead to significant problems and challenges, which also affect metric M3. Integrations play a crucial role in enabling seamless communication and data exchange between different information systems and platforms. Without plug-and-play integrations, companies often face time-consuming and complex manual processes to connect disparate systems, leading to avoidable inefficiencies and increased costs. For instance, the average enterprise company uses 91 cloud services or software applications for Marketing alone (!), while using another 70 for collaboration, 43 for CRM and Sales and 37 for productivity [76].

Additionally, customers may find themselves having to manually bridge the gaps between the core software product and other essential tools they rely on, which undermines the (business) value of such a product. Also, the lack of integrations can hinder collaboration among teams and limit scalability, as companies may struggle to adopt new technologies or expand their product offering. This can result in missed opportunities, delayed projects and reduced competitiveness. Therefore, the absence of plug-and-play solutions can impede smooth operations and hinder technological advancements within IDE companies.

2.2.7 Lack of self-correcting mechanisms

One of the most promising fields of AI is the one of Reinforcement Learning (RL). One of the most impactful methodologies of Product Development is the Agile Methodology. One of the best pedagogical approaches is feedback. What all of them have in common is that they are based on the principles of **Control Theory**, a discipline that originated in Engineering, but has spread among many other disciplines. The block diagram of a *control system* appears in Fig. 3 and it is an abstract representation of many (but not all) real-life systems, ranging from an electric circuit to a nuclear power plant.

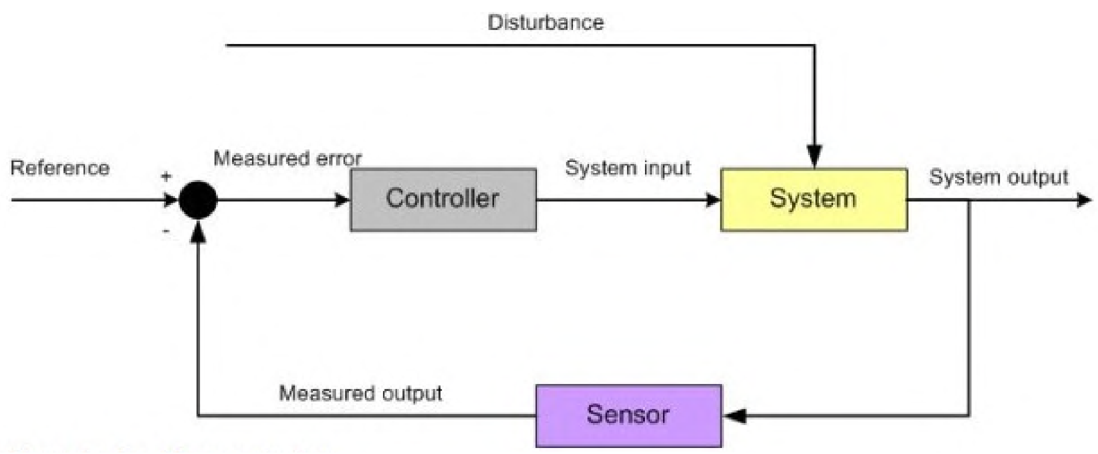


Figure 3 - Block diagram of a closed-loop control system.

In the context of organization governance, the implementation of a control system can introduce **self-correcting mechanisms**. Alternatively, the lack of such mechanisms can give rise to a wide variety of problems and inefficiencies, which negatively affect metric M2. For instance, without self-correction, organizations become susceptible to a range of errors, mistakes and inefficiencies. In the absence of feedback loops and corrective measures, these errors can go unnoticed and propagate throughout the system, thus leading to suboptimal performance, poor Decision-Making and an avoidable waste of resources.

Additionally, the lack of self-correction typically inhibits the ability of an organization to adapt and evolve in response to changing circumstances or market dynamics, thereby hampering innovation, agility and competitiveness. This rigidity can hinder growth, responsiveness to customer needs and the ability to seize new opportunities. Furthermore, the lack of self-correction undermines the development of a learning culture inside the organization. Without mechanisms in place to identify and address mistakes, organizations may not foster an environment that encourages experimentation, risk-taking and continuous improvement. The absence of self-correction can stifle creativity, inhibit employee growth and ultimately, impede the organization's ability to adapt and evolve.

2.2.8 Lack of a collaboration platform among organizations

The **lack of a collaboration platform** among IDE organizations can lead to numerous problems that affect metrics M2 and M3. For instance, the absence of a collaboration platform inhibits the sharing of resources and expertise. Companies often possess unique skills and specialized resources that can benefit others in the industry. However, without a platform to connect and collaborate, these valuable assets remain untapped, resulting in missed opportunities for innovation and growth.

Additionally, without a decentralized *common point of reference*, it becomes difficult to foster a culture of *collective learning* and continuous improvement. A collaboration platform enables companies to share best practices, lessons learned, and emerging trends, thus fostering a collaborative environment that accelerates individual and collective growth. Also, the absence of a collaboration platform can hinder cross-company innovation and *idea generation*. Collaborative platforms provide a space for brainstorming, ideation, and open innovation, thereby allowing tech companies to leverage diverse perspectives and co-create solutions that can drive industry advancements.

Furthermore, the absence of a collaboration platform can hinder *standardization* and compatibility. Companies may develop products or solutions that could be mutually beneficial, but without a platform to align their efforts, interoperability issues may arise, thus leading to a waste of resources and limited market adoption.

2.3 Why scientific discovery is problematic

In order for an organization to succeed in terms of metric M1, i.e. *managing* innovation (as defined in Section 2.2), it is a minimum prerequisite for **innovation** itself to already exist from the first place or be created beforehand. This affects both Individual Researchers and DEI organizations. Also, innovation is not possible without scientific discoveries and technological breakthroughs. However, there are a series of problems that negatively impact or delay scientific discovery, as described below. Acknowledging those problems and then attempting to address them is particularly important. The rationale behind this focus will become more apparent in Chapter 4, which describes the solution.

2.3.1 The ten dogmata of Scientism

An important distinction should be made between Science (which stems from Scientific Method, as shown in Fig. 4) and 'Scientism', which is something in-between a *belief system* and a religious *cult* (as shown in Fig 5). The fundamental difference between real Science and Scientism lies in their approach to knowledge. While Science is a disciplined, systematic and evidence-based approach, Scientism is a *narrow* and *dogmatic* world-view that denies the value of other forms of knowledge and human experience. Science is open to revision and refinement as new evidence and data become available. As such, Science recognizes the *limitations* of its methods, while it seeks to expand our understanding of the natural world, whereas Scientism typically *reduces* a phenomenon to a set of measurable and independent components (Section 2.3.6), ignoring the complexities and richness of human life or nature. In other words, Science is based on: (a) methodological inquiry, (b) challenging the status quo, and (c) disputing the 'popular opinion' (also known as 'common belief' or 'traditional wisdom'), which is rarely correct and highly subjective. However, Scientism stands for the opposite of that since it inhibits inquiry, and this causes some serious consequences, as described below.

Although most of the power of Science comes from its *practical* applications (i.e. applied research), it also has a strong intellectual appeal (i.e. fundamental research). For instance, it offers new ways of understanding the world, including the mathematical order at the heart of atoms, molecules, genes and the universe. Religious adherence to the ten dogmata of Scientism (as described below) is troublesome for both kinds of research, but it also has a real-life, practical impact (sometimes positive and sometimes negative) in our everyday lives. This kind of impact can often have enormous *social*, *political*, *economic* and *environmental* consequences. Without trying to instil irrational fear, one might even argue that they have an impact on the existence of the human species as a whole. Eventually, a person or a group of people are going to make some very important decisions that at some point might affect the whole of humanity. So it is absolutely crucial for those decisions *not* to be made by the wrong assumptions (or assumptions that might be *partially* wrong or of *uncertain* nature).

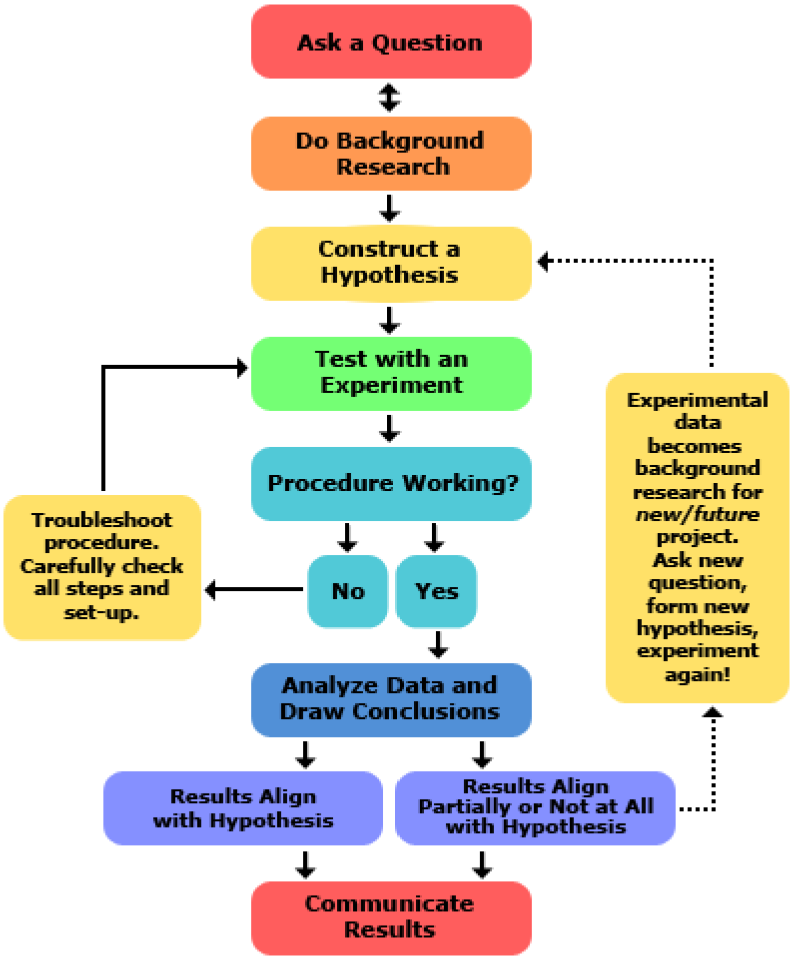


Figure 4 - The Scientific Method, i.e. the scientific research process.

As codified by Sheldrake [31], the ten core beliefs that most Researchers take for granted are the following:

(1) **Everything in nature is mechanical:** All living organisms are machines with no goals of their own.

Figure 5 - Screenshot from the video ‘Exposing Scientific Dogmas’ by Sheldrake [20].

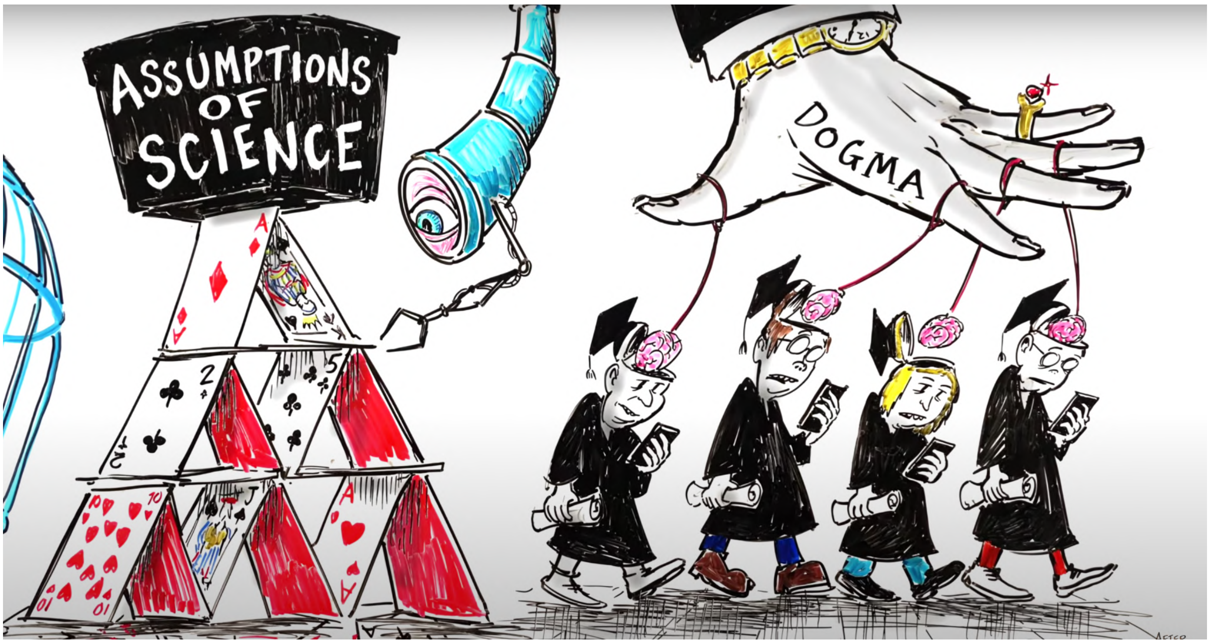


Figure 6 - Screenshot from the video ‘Exposing Scientific Dogmas’ by Sheldrake [20].

(2) **All matter is unconscious:** Stars, galaxies, planets, animals and plants have no inner life or subjectivity or point of view. Even human consciousness is an illusion produced by the material activities of brains, i.e. consciousness does not even exist at all.

(3) **The total amount of matter and energy are always the same** (with the exception of the Big Bang, when all the matter and energy of the universe suddenly appeared).

(4) **The laws of nature are all fixed:** They are the same today as they were at the beginning (i.e. Big Bang), and they will stay the same forever. In other words, the laws are completely static like constants. So, even though the universe evolves, no laws of nature do.

(5) **Nature is purposeless:** Biological evolution has no goal or direction.

(6) **Biological hereditary is material:** It is all carried in the genetic material, DNA and in other material structures.

(7) **Memory is stored in the brain:** Even though nobody really understands how memory works, memories are stored as material traces in brains and are wiped out at death.

(8) **Minds are inside the heads:** The minds are nothing but the activities of brains. When you look at a tree, the image of the tree you are seeing is not 'out there', where it seems to be, but inside your brain.

(9) **Psychic phenomena are impossible:** Unexplained phenomena like telepathy are completely illusory.

(10) **Mechanistic Medicine is the only kind that really works:** All rival medical systems are outright rejected.

The main questions about all of the aforementioned beliefs are that:

- "Are they metaphors or are they testable hypotheses?"
- "If they are scientific hypotheses, how could they be tested or refuted?"
- "If they are assumptions, how could we know that they are valid ones, and what are the specific cases in which they would no longer be valid?"
- "If an AI machine or an extraterrestrial creature conducted scientific discovery without any previous knowledge by humans, could it independently deduce all these beliefs and reach the same conclusions?"

It turns out that, for the vast majority of time, Researchers, research councils, educational systems, governments and others religiously adhere to these beliefs without ever questioning them (Fig. 6). In other words, it is their default world-view, constituting all of the above the "scientific" creed or simply, **dogmatism**.

2.3.2 Obscurantism, ostracization and heretics

Even though dogmatic denial is particularly problematic for scientific discovery (as explained in Section 2.3.1), there are a series of other problems that also negatively affect it. An immediate consequence of this dogmatism is the widespread **obscurantism** in various scientific fields and disciplines, which is what reinforces the *status quo*, which is the equivalent of inhibiting self-correction mechanisms (Section 2.2.7) in organizations. In general, obscurantism in scientific research refers to the rejection or suppression of new and unconventional ideas, theories, or findings that challenge established beliefs or dogmas. This can result in Researchers who explore these new ideas being treated as **heretics** or outcasts in the scientific community, thereby dismissing their work without proper examination or debate, and eventually leading to their **ostracization**.

The main problem with obscurantism is that it hinders scientific advancement by limiting exploration of new ideas and concepts about our knowledge and understanding of the world. Science is based on the continuous questioning and testing of established theories and hypotheses. Hence, when Researchers are discouraged from pursuing innovative approaches and sharing their ideas or findings, scientific progress can be impeded and important discoveries may be delayed or overlooked. Ultimately, the rejection of new and innovative ideas can significantly stifle creativity and innovation.

Furthermore, obscurantism can have serious implications for scientific integrity and objectivity, as Researchers may feel pressure to conform to established norms and beliefs, rather than pursuing evidence-based inquiry. Additionally, treating Researchers as heretics can lead to a *toxic work environment*, in which scientific inquiry is discouraged, and Researchers may be afraid to speak out or share their ideas. Researchers may also feel discouraged from pursuing new ideas or questioning established theories, as they fear backlash or ostracization from their peers. This undermines the fundamental principles of scientific inquiry, such as the importance of evidence-based analysis and the need for open debate and dialogue, while also further limiting creativity, innovation, and collaboration in the scientific community, leading to a loss of pluralistic perspectives and expertise.

To combat obscurantism, it is crucial to foster an open and inclusive scientific community that welcomes and encourages different ideas and perspectives. Researchers should be willing to consider and test new ideas, while institutions should support and protect Researchers who explore unconventional theories and findings. By doing so, it can be ensured that scientific research remains a dynamic and evolving field that continues to expand our knowledge and understanding of the world.

In the 20th century, the most famous person that was attacked as science heretic and then vindicated was arguably Albert Einstein (Fig. 7). Another example from history is Galileo Galilei who said *“Eppur si muove”* (“And yet it moves”) after literally facing charges of heresy by the Roman Inquisition of the Papist Church, which was the personification of obscurantism in Western Europe during the late Middle Ages. An example from more recent history is Rupert Sheldrake who has been censored by TED talks [20] and ostracized by the research community for daring to propose the hypothesis of *morphic resonance* [32 - 34]. Weather Sheldrake will eventually be vindicated or not is irrelevant, but what is particularly problematic is the *reaction* towards his “heretical” hypothesis, which was enough to make him face persecution by the modern-day version of the Roman Inquisition. Also, his status as a *persona non grata* has hampered the ability for anyone to obtain funding [35] in order to conduct any of the experiments proposed by Sheldrake in his book [34]. Even if funding was to be obtained to hire Researchers and buy lab equipment, the odds of successfully going through the gatekeepers (Section 2.3.3) and the peer-review process (Section 2.3.4) would be close to zero.

2.3.3 Gatekeepers, biases and political interference

For those who idealize (and idolize) ‘Science’, Researchers are the epitome of objectivity, rising above the sectarian divisions and illusions that afflict the rest of humanity. This belief is sustained by the ideal of disembodied knowledge, unaffected by ambitions, hopes, fears and other emotions. However, this *materialist* philosophy or the “scientific” world-view is extremely *flawed*, because as human beings, we are all flawed in one way or another. Therefore, a problem related to dogmatism (Section 2.3.1) and obscurantism (Section 2.3.2) are the various types of **biases** that humans might have. Furthermore, humans are subject to (a) the limitations of personality, (b) politics, (c) peer pressure, (d) what is fashionable or trendy, and (e) the need for funding (Section 2.3.8), as explained below.

Plato's allegory of the cave is a relatively famous story in which he describes a group of people who had been living their whole lives in a dark cave, chained to face a wall. They could only see shadows on the wall cast by objects passing by behind them, and they believe these shadows to be reality. When one of them was freed and shown the outside world, he initially struggled to comprehend what he saw. But as he became accustomed to the light and the reality outside the cave, he realizes that the shadows were only a poor representation of reality (i.e. an illusion). In the allegory of the cave, Researchers are supposed to venture forth into the light of objective truth and bring back their discoveries for the benefit of ordinary people, trapped in a world of opinion, self-interest and illusion.

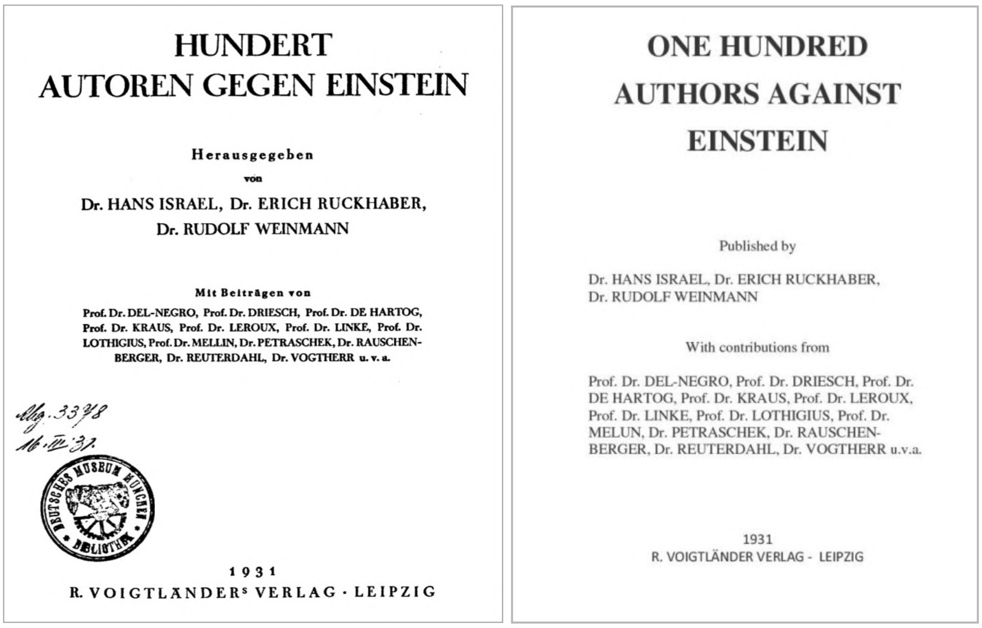
Nonetheless, to state the obvious, humans are humans, so whatever we produce is bound to be **anthropocentric** up to a certain extent. For instance, why do we call them "laws" of nature? More importantly, no matter what it is called, how can we be sure that the concept of laws exists and is real? This might touch the sphere of Philosophy, but it also has plenty of practical implications. For example, how can we be sure that the speed of light, *c*, or the gravitational constant, *G*, are actually "constant"? Even relatively small variations (or oscillations) could create immeasurable problems for measuring devices, critical infrastructure, life-saving equipment, etc. For example, Sheldrake [34] proposes 'habits of nature' (which can be inherited by species), which is compatible with and could be a hypothesis for the emergence of **'learnable rules'** as proposed by the GUT-AI theory [1], thereby hinting that the mental connection between biological evolution and AI could be based on **self-learning systems** (as explained in Section 3.4.2) through nonlinear *cyclic* advancements or breakthroughs (Fig. 11-12), instead of a linear pattern.

In general, published data have to pass through three selective **filters**, namely:

- **Cherry-picking:** The first filtration of the data occurs when experimenters decide to publish some results rather than others.
- **Publication bias:** The second one is when editors of journals consider only certain kinds of results eligible for publication, which are deemed as "positive" or "significant", meaning that "negative" or "inconclusive" findings do not get published, thus leading to a phenomenon that is particularly problematic for scientific discovery.
- **Reviewers bias:** The third one happens in the peer-review process (Section 2.3.4), which ensures that expected or conventional results are more likely to be approved for publication than unexpected results.

Therefore, most Researchers publish only a small proportion of their results, which is likely to introduce serious biases into the scientific literature. As an illustration, if businesses were required to publish only 10% of their accounts, they would probably publish those that made their business look as profitable and as well managed as possible. Conversely, if they needed to submit only 10% of their accounts to the tax authorities, they would tend to show their least profitable activities. Suppressing 90% of the data gives a lot of scope for selective reporting. How much does this practice affect scientific research? No one really knows.

In general, **bias** can affect scientific discovery in many ways, such as in the choice of research questions, the design of experiments, and the interpretation of results. Bias can lead to incorrect conclusions or an incomplete understanding of phenomena. However, bias can also be political. More specifically, politics can interfere with scientific discovery, particularly in fields such as "climate change", where policymakers may choose to ignore or discredit scientific findings. Hence, **political interference** in scientific research can limit the ability of Researchers to conduct meaningful research or limit access to research findings.



HUNDERT
AUTOREN GEGEN EINSTEIN

Herausgegeben
von
Dr. HANS ISRAEL, Dr. ERICH RUCKHABER,
Dr. RUDOLF WEINMANN

Mit Beiträgen von
Prof. Dr. DEL-NEGRO, Prof. Dr. DRIESCH, Prof. Dr. DE HARTOG, Prof. Dr. KRAUS, Prof. Dr. LEROUX, Prof. Dr. LINKE, Prof. Dr. LOTHIGIUS, Prof. Dr. MELLIN, Dr. PETRASCHEK, Dr. RAUSCHENBERGER, Dr. REUTERDAHL, Dr. VOGTHERR u. v. a.

1931
R. VOIGTLÄNDERs VERLAG · LEIPZIG

ONE HUNDRED
AUTHORS AGAINST
EINSTEIN

Published by

Dr. HANS ISRAEL, Dr. ERICH RUCKHABER,
Dr. RUDOLF WEINMANN

With contributions from

Prof. Dr. DEL-NEGRO, Prof. Dr. DRIESCH, Prof. Dr. DE HARTOG, Prof. Dr. KRAUS, Prof. Dr. LEROUX, Prof. Dr. LINKE, Prof. Dr. LOTHIGIUS, Prof. Dr. MELUN, Dr. PETRASCHEK, Dr. RAUSCHENBERGER, Dr. REUTERDAHL, Dr. VOGTHERR u.v.a.

1931
R. VOIGTLÄNDER VERLAG - LEIPZIG

Figure 7 - Book entitled 'One Hundred Authors against Einstein' by Israel et al. [30].

A prominent example of political interference in scientific research from the modern history is that of one hundred authors against Einstein (Fig. 7), in which their political beliefs motivated them in a writing a heavily biassed critique, which did not actually provide constructive feedback, but was mostly targeted in discrediting the personality of Einstein due to his opposing political beliefs.

Additionally, **peer pressure** is a form of bias that can influence the research agenda, leading to a lack of innovation and originality in scientific research. When Researchers feel pressured to conform to prevailing ideas or follow the trends set by their peers, they may be less likely to explore unconventional or risky avenues of inquiry. This conformity can stifle creativity and hinder breakthrough discoveries. This is typically a direct consequence of obscurantism and ostracization (Section 2.3.2). Hence, the fear of being criticized or ostracized by the scientific community often leads Researchers to play it safe and stick to established methodologies and theories.

Moreover, Researchers who are **militant atheists** can pose a significant threat to scientific research and discovery since their biases can interfere with scientific inquiry by imposing their views on the minority that might have different beliefs. While it is important for Researchers to maintain objectivity and remain open-minded, militant atheists may exhibit a tendency to dismiss or overlook evidence that contradicts their world-view. This kind of *dogmatic approach* can stifle meaningful scientific debates, while hindering the exploration of alternative perspectives. Scientific research thrives on the pluralism of ideas and the freedom to challenge existing theories. When Researchers allow their atheistic beliefs to dominate their work, not only does it undermine the integrity of the scientific process, but it also limits the potential for groundbreaking discoveries.

When it comes to biases, a particularly crucial problem that arises is that of **gatekeepers**. The problem of gatekeepers in scientific research refers to the individuals or institutions that control access to research funding (Section 2.3.8), publication opportunities or other resources needed to conduct and disseminate research. These gatekeepers can include academic departments, editorial boards of scientific journals, funding agencies and other small cliques or inner circles that hold the power to approve or reject research proposals, manuscripts, or grant applications. The existence of gatekeepers becomes problematic when they use their power to block or restrict access to these resources based on factors *unrelated* to the quality or merit of the research itself. This can result in biases against certain Researchers or research topics, thereby perpetuating a lack of pluralism in scientific thinking and a lack of innovation in scientific inquiry. Furthermore, gatekeepers can have the power to influence the direction of scientific research by prioritizing certain areas of study or research questions over others, which can limit the scope of inquiry and potentially stifle progress in certain fields.

It is worth mentioning that not only do some Academics act as gatekeepers of stupidity and obscurantism in their dark ivory towers, but they often take particular pride in doing so by wearing their black academic gowns, as if it was some kind of high moral duty. They sometimes go as far as to publish papers in which they literally self-identify as "'gatekeepers", while also explaining why they consider it crucial to "take seriously their role as primary gatekeepers" [29]. To make matters worse, they even reach to the point of using religious terminology (!) in these papers, such as "sacrosanct" [29] in order to justify their self-appointed role as gatekeepers and to also impose some sort of Mosaic Law on what and how other Researchers should publish, thereby acting as modern-day scribes and Pharisees who criticize acts they deem as "sacrilegious" with an unfounded sense of self-righteousness and sacrosanctity. It is highly probable, that this phenomenon could be a by-product of the religious character of the mediaeval "universities" in Western Europe in direct contrast to the secular character of the 'Pandidacterium' (i.e. the Imperial University of Constantinople), which was probably the one they were trying to imitate. As a result, this persona of "Researchers" might have actually taken scientific progress, but potentially humanity as a whole, decades back.

2.3.4 Flawed culture and processes

A shortcoming that immediately follows dogmatism, obscurantism and biases is the existence of a **flawed culture** and **flawed processes** in scientific research, since Individual Researchers exist and function as part of research *communities* inside a social network. This situation is essentially equivalent to the existence of misaligned interests (Section 2.2.3) in an organization. For instance, **peer review** (Fig. 8) is a critical component of the scientific process, but it can also be flawed, leading to inaccuracies or biases in the review process. However, because it is *not* an incentivized task, reviewers have been known to delay their work or provide unhelpful reviews. Moreover, authors regularly report facing “reviewer bullying” wherein reviewers force authors to conduct additional experiments, cite certain papers, make unnecessary changes, and so on.

Figure 8 - List of AI papers that were rejected *at least once* during the peer-review process (Read ‘Kalman Filters’ instead of ‘Kalyan Filters’).

Most journals opt for single-blind peer review, which leaves room for biases and professional jealousy to creep in. Apart from this, the *excessive dependence* on the peer review system has led authors, editors, and third-party services to take advantage of it leading to peer-review scams. As a result, the peer review system in its present form is questioned by many, mainly due to the fact that it acts as a **single point of failure** (Section 3.1).

Additionally, the current academic "reward system" typically prioritizes **quantity** over quality, thus leading to a focus on publishing as many papers as possible (Fig. 9) rather than conducting high-quality research. Besides, the culture of competition in academia can lead to a lack of collaboration, replication of research, and an emphasis on publishing rather than the quality of research.

Figure 9 - Viral meme (by an anonymous online user) criticizing the current state of AI research.

Furthermore, the emphasis on **short-term results** and immediate impact can also lead to a neglect of long-term research questions and exploration and also, to a lack of investment in long-term scientific research in general. This kind of flawed culture often results in resistance to change and to adopting new technologies or research methods, which can significantly hinder real progress and innovation in scientific inquiry.

2.3.5 Reproducibility crisis

In addition to the aforementioned problems, a major problem in scientific research is that of **reproducibility crisis** (also called replicability crisis) [9, 10]. In recent years, there has been a growing concern about the reproducibility of scientific research (Fig. 10). This means that other Researchers are often unable to replicate the findings of previous studies, which can be particularly problematic for scientific discovery. Similar to most aspects of human life, scientific research is not immune to misconduct or fraud. This can undermine the integrity of scientific discovery and damage the reputation of scientific research in the eyes of the public and the taxpayers funding such research.

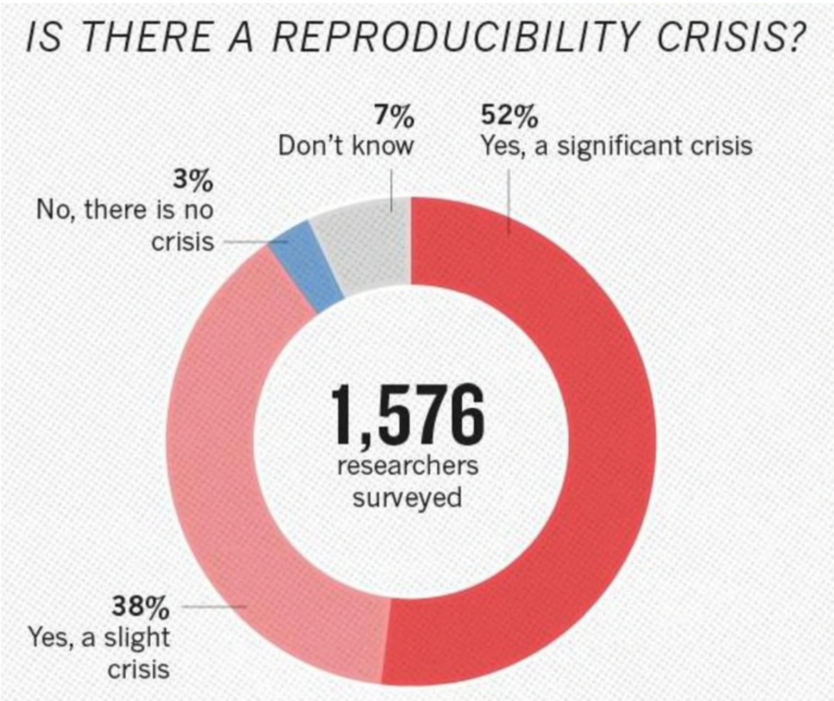


Figure 10 - The vast majority of Researchers believe that there is a reproducibility crisis [10].

For instance, poorly designed studies have become a major concern for academia. One of the primary reasons behind this problem is that statistical flaws in published research often go undetected. Since breakthrough results are valued the most, Researchers feel compelled to hype their results in order to get published. Moreover, they tend to focus on particular patterns in data and manipulate their study designs to make the results more attractive for the journals. Instances of “p-hacking” in which Researchers report only those hypotheses that end in statistically significant results are also on a rise. Additionally, the lack of standardization in research methods and protocols can lead to inconsistencies in research findings. Also, complete reproducibility would be expected on the assumption that all the laws of nature are eternal, the same at all times and in all places.

2.3.6 Multidisciplinary and interdisciplinary research

Similar to enterprises that might consist of *silos*, Researchers have separated by design their respective research areas into ‘fields’, which are remarkably very similar to agricultural fields that are well-marked, well-fenced and well-guarded. If an Independent Researcher dares to overstep to a field next to him, then he should prepare himself for a battle! As also explained in [1], this is in direct contrast to Ancient Greeks who were *not* confined to a particular specialization or interest group, but were *polymaths*. This kind of enmity creates all sorts of problems, since phenomena around us can rarely be interpreted as a mere collection of independent components (i.e. extreme reductionism [32]), but it usually requires a **multidisciplinary** and **interdisciplinary** approach instead, equivalent to the *holistic* approach (Section 2.2.5) required in an organization and also, equivalent to the situation of unnecessary replication of work (Section 2.2.2).

More specifically, many of the research problems today are extremely complex, requiring the use of advanced methodologies and expertise from multiple fields. Therefore, many of the most pressing research problems today require collaboration between Researchers from different fields.

However, this makes the process of scientific discovery much more difficult and time-consuming, since there can be barriers to interdisciplinary collaboration, such as disciplinary silos or differences in *terminology* or *methodology*. Hence, the fragmentation of scientific research in silos can limit the scope and impact of research findings, particularly in interdisciplinary fields. Furthermore, interdisciplinary collaboration is essential for many scientific discoveries, but it can be difficult to find experts from different fields who can work together effectively.

As an illustration, a very good example is that of the **Two Queen hypothesis** or conjecture. This is a hypothesis which states that, given certain *assumptions*, an ecosystem consists of entities (e.g., organisms in Evolutionary Ecology, companies in Macroeconomics, superpowers in International Relations or agents in RL), whose behaviours, relationships and interactions can be described using three main patterns or types of dynamics: (a) *Red Queen dynamics* (e.g., prey-predator co-evolution, Free Markets or a multipolar world), (b) *White Queen dynamics* (i.e. biotic homogenization, oligopoly or unipolar world), and (c) purely random dynamics (e.g., "genetic accidents"). As shown in Fig. 13 and Table 1, Red Queen dynamics can be further categorized [83] into: (i) Fluctuating Red Queen dynamics, (ii) Escalatory Red Queen dynamics, and (iii) Chase Red Queen dynamics. In real-life ecosystems, all types of dynamics can be assigned a weight and be *combined*.

Due to the fact that this is a hypothesis that describes an ecosystem and the entities interacting with it and also among themselves (i.e. a Complex System [1, 36]), only a *multidisciplinary* approach can be used to both mathematically and empirically study this proposed phenomenon, due to the fact that it necessarily requires knowledge from various disciplines, including Game Theory, Network Science, Optimization, Control Theory, Dynamical Systems, Collective Intelligence, etc.

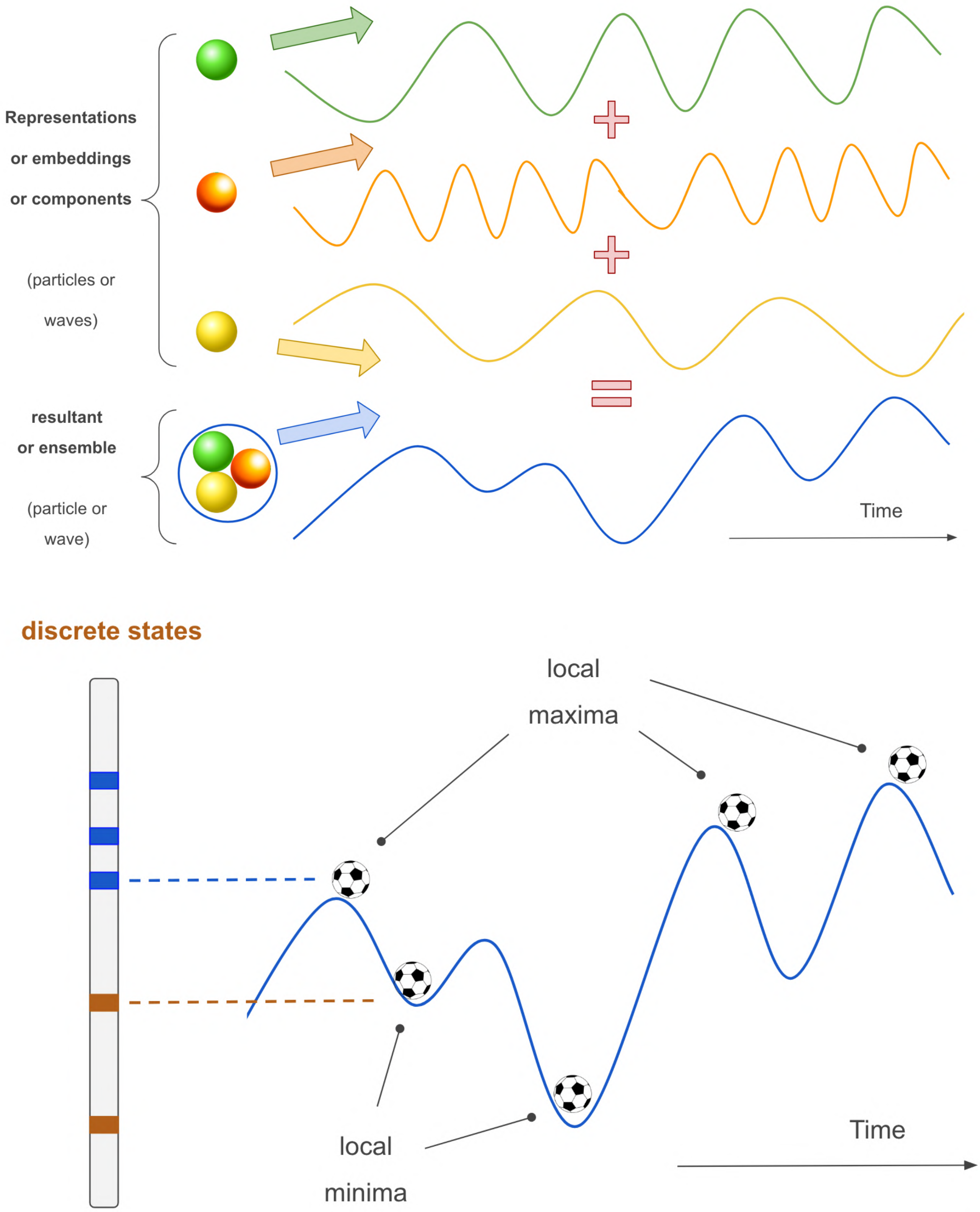


Figure 11 - The nonlinear cyclical advancement of a system due to stochasticity and determinism shown in only *two* dimensions due to simplicity (without loss of generality).

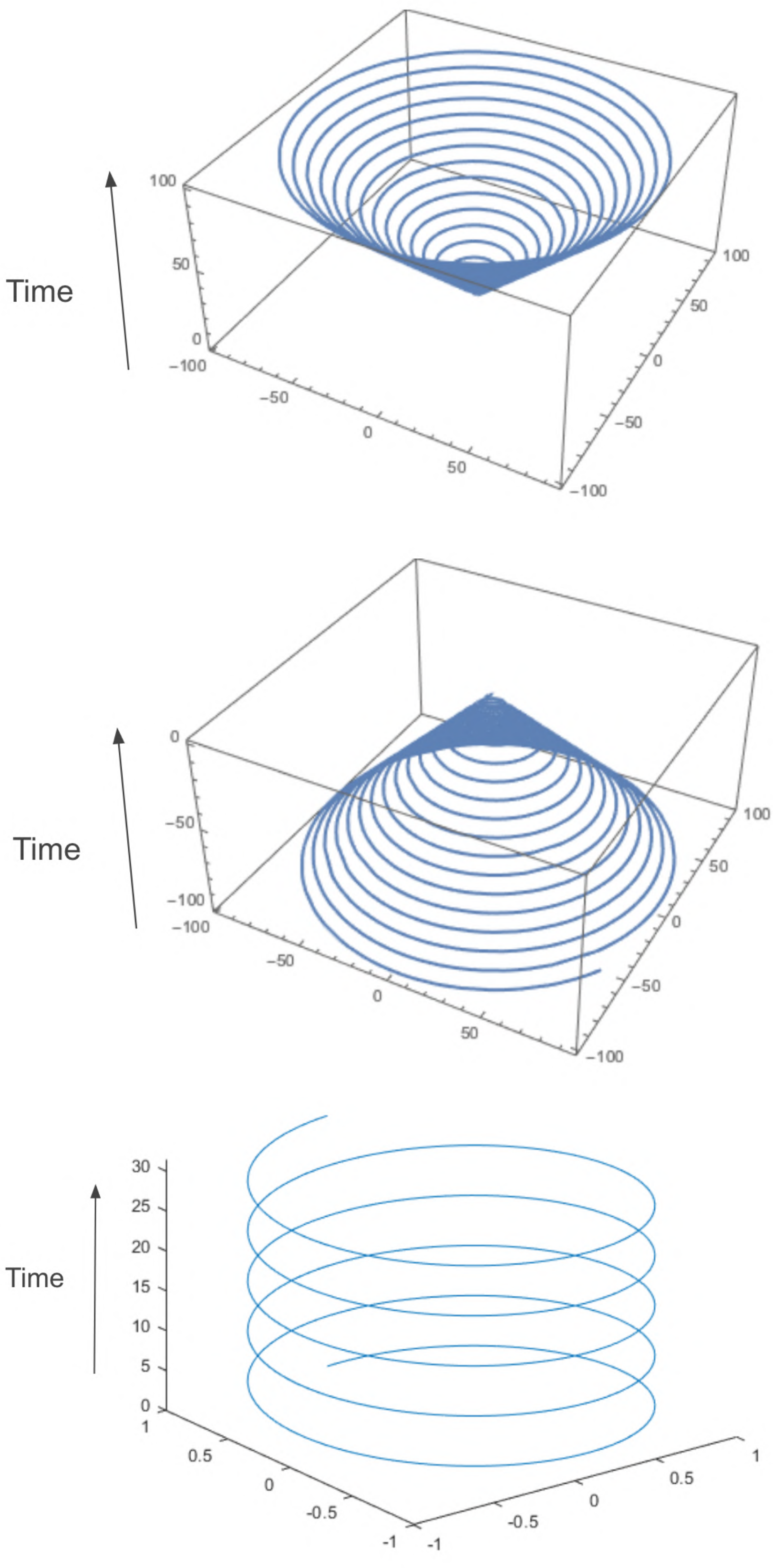


Figure 12 - Three distinct edge cases of nonlinear cyclical advancements, anti-fragile (top) [84], fragile (middle), and robust (bottom), shown in only *three* dimensions (without loss of generality).

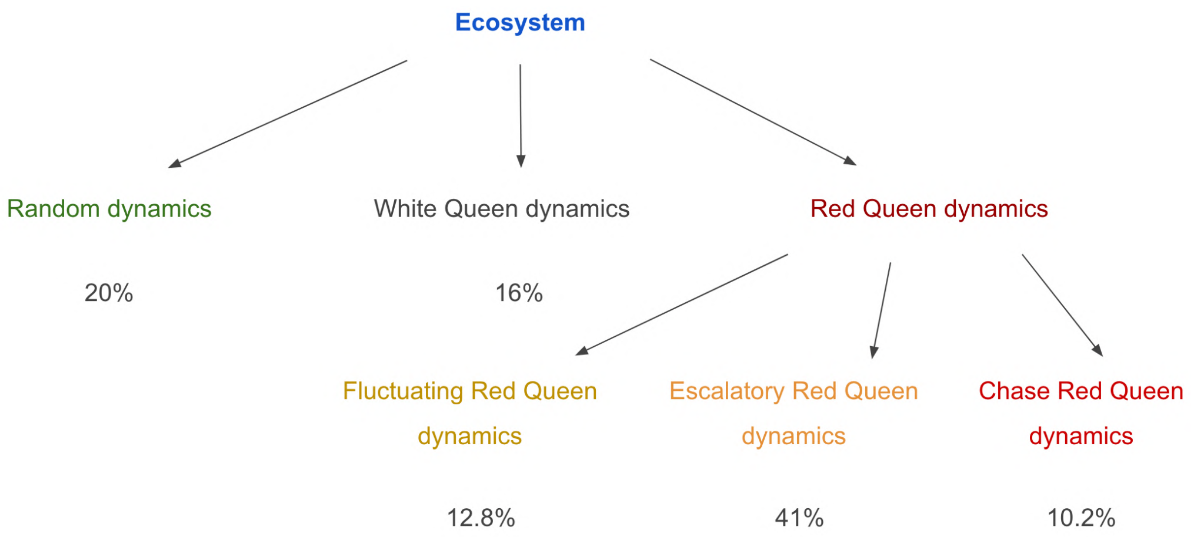


Figure 13 - Dynamics of an ecosystem according to Two Queen hypothesis.

	White Queen	Fluctuating Red Queen	Escalatory Red Queen	Chase Red Queen
Example from human history	Unipolar world	Byzantine Empire (under Justinian I)	Cold War (arms race)	World War II
Stability	Too stable	Very stable	Marginally stable	Unstable
Mode of selection	Fluctuating	Fluctuating	Directional (unidirectional)	Directional (multidirectional)
Divergence between entities	N/A	Low	High	Very high
Within-population diversity	Very low	High	Low	Very low
Competition	Negligible	Low	High	Very high
Innovation	Negligible	Incremental	Breakthroughs	Breakthroughs

Table 1 - Comparison of the White Queen dynamics and the three types of Red Queen dynamics.

Additionally, this hypothesis happens to be fully compatible with the *group dynamics* per **GUT-AI theory** [1], which incorporate both a percentage of *stochasticity* (e.g., 20%) and a percentage of *'oscillatory' determinism* (e.g., 80%), both of which need to be 'learnable'. In this case, determinism oscillates between the two states of Red Queen dynamics and White Queen dynamics, similar to a pendulum *oscillating* between two extrema. Even within a single state, such as the Red Queen dynamics, there are oscillations between entities (e.g., host-parasite co-evolution), which is compatible with the *pyramidal* **fractal pattern** [1], also proposed by GUT-AI theory. As a direct consequence of such a pattern, the theory also explains how **physical constraints** or **limits** (e.g., Planck length, Planck time, Shannon limit, speed of light) [1], such as the *physical size* of an entity can and will affect the *expression* of the two aforementioned states. For instance, *homogeneous entities* (e.g., unicellular organisms, generic AI solutions or homogeneous RL agents) exhibit more White Queen dynamics than Red Queen dynamics, whereas *heterogeneous entities* (e.g., multicellular organisms, industry-specific AI solutions or heterogeneous RL agents) exhibit relatively more Red Queen dynamics than White Queen dynamics.

Without excluding any alternative interpretations, GUT-AI theory also accommodates the theory that evolution is a kind of **self-learning system** (as explained in Section 3.4.2), through nonlinear *cyclic* advancements or breakthroughs (Fig. 11-12), instead of a linear pattern. The same might apply for human history too, since it is a pure subset of evolution. Additionally, at the lowest levels of the pyramid (i.e. microcosm), Fig. 11 provides a framework for studying *particles*, whereas at the higher levels of the pyramid (i.e. macrocosm), Fig. 12 provides a framework for studying *Big Bang, Big Crunch*, etc. Furthermore, at the intermediate levels of the pyramid, Fig. 11 provides a framework for studying *Audio Source Separation* in Signal Processing, *human brain* [36] in Computational Neuroscience, and ultimately, *Automated Scientific Discovery* (Section 2.3.10).

2.3.7 Data Management and accessibility

A further problem in scientific discovery is that of **Data Management and accessibility**. Data management in scientific research refers to the process of collecting, organizing, storing, and sharing data generated during the research process. With the vast amount of data being generated by scientific research today, there are often challenges with managing and analysing this data. This can make it difficult to identify patterns or draw meaningful conclusions from the data, which is one of the main aims of AI. The problem arises when Data Management practices are not properly implemented, which can lead to a range of issues including data loss, inconsistencies and difficulties in data interpretation. Effective Data Management is essential for ensuring the reproducibility (Section 2.3.5) and transparency of scientific research, as well as promoting collaboration and the sharing of knowledge among Researchers.

Additionally, the problem of Data Management in scientific research is exacerbated by the **lack of standardization** and automated ways to implement best practises for data collection, storage and sharing by eliminating any guesswork. This requires the use of a free and open-source software for using standardized data formats, documenting data collection procedures and ensuring that data is stored in a secure and easily accessible manner.

As a consequence, with the significant increase in available data, information and publications generated in scientific research on a daily basis, Researchers may struggle to manage, analyse and interpret large amounts of data, leading to potential errors, bias and inefficiencies. **Data overload** in scientific research is a real problem, since having too much data to effectively analyse and interpret can result in a lack of clarity in research findings and can hinder scientific progress. Therefore, while this abundance of data can provide valuable insights and opportunities for discovery, it can also overwhelm Researchers and make it difficult to identify relevant patterns and trends. The problem of data overload is compounded by the fact that many Researchers may not have the necessary skills, resources or even time to effectively analyse and interpret the vast of amount of new information that gets generated. This can result in data being underutilized, potentially leading to inefficiencies for scientific discovery and innovation, while also leading to missed opportunities for quick and effective commercialization of research findings.

2.3.8 Lack of funding and resources

Similar to misaligned interests (Section 2.2.3) in an organization, the **lack of funding and resources** or more precisely, the lack of *aligning* funding with tangible and practical impact of the research outcomes, renders scientific discovery considerably inefficient, ineffective and somewhat useless from a practical point of view.

For instance, the Human Genome Project was formally launched in 1990 with a projected budget of USD 3 billion, which also led to the creation and subsequent bursting of the biotechnology bubble. However, instead of closing the gap (or at least bridging the gap), there was a realization that there is a huge gap between gene sequences and the way living organisms grow and behave [32], mainly due to the relatively small number of genes in humans (approximately 25,000) as compared to plants or animals much simpler than humans. The issue is in this specific case was that the Researchers were fixated too much on a specific assumption or hypothesis (i.e., that the vast majority of human characteristics are mostly encoded inside our genes).

The root cause of this issue was the reckless waste of taxpayer money to a hypothesis that did not align with the interests of those funding this research (i.e. the taxpayers), hence there was a considerable *misalignment* of funding with tangible and practical impact of the research outcomes (e.g., the improvement of healthcare or tackling any other practical life that would improve human health or quality of life in general). Additionally, if this research was never funded, then a private company which was already active in sequencing the human genome would have made the same discoveries *anyway*, with the added benefit that government involvement (such as the speech by the then President Bill Clinton) would *not* have created such a huge setback to the biotechnology sector at the time.

Furthermore, all that taxpayer money could have (directly or indirectly) been allocated to other causes with more practical and tangible impact towards improving human health or quality of life in general. It should be highlighted that it was not the involvement of government *per se* that led to these huge inefficiencies, but the *way* the initiative was approached and eventually implemented, i.e. in a purely centralized way (Section 3.1).

In general, funding for scientific research can be limited due to a variety of reasons, thus leading to a lack of resources for Researchers to conduct experiments or investigations that are necessary for the improvement of humanity as a whole. Also, limited resources, such as access to expensive equipment (e.g., compute resources or data storage) or limited reading materials, can significantly hinder scientific discovery, particularly in low-income or developing countries, who are in desperate need for improvement of their quality of life.

2.3.9 Dissemination and exploitation of research findings

An additional problem is that even when scientific discoveries are made, there can be challenges with **disseminating** this information to the broader public or policy-makers or **exploiting the research findings** in order to subsequently fund further research, thus limiting the impact of scientific discovery on society. In general, it is a relatively well-known fact that a wide communication gap exists between the research and the non-research community. This has resulted in miscommunication of science, divided opinions about scientific matters, and lack of informed Decision-Making among the public. Researchers are partly responsible for this because they lack time or sometimes the inclination to engage with the public about their research work.

Therefore, the public is largely dependent on the legacy media, which is often blamed for misconstruing scientific facts. The competitive nature of the research is conducted in academia is also responsible for poor communication of research, whereas research in the industry seems to be communicated in a considerably improved way. For instance, in an attempt to grab attention, sometimes Academics, universities and even journals mislead the public by hyping the results or promoting only positive results, while the impact of these results might be close to zero (Section 2.3.8) or not even reproducible at all (Section 2.3.5).

Additionally, the **lack of commercialization** of research findings in academia is a significant problem that can prevent society from fully benefiting from the fruits of scientific progress, even though such research was fully or almost fully funded by the taxpayers. Research findings that are not commercialized may remain unused or not even become public knowledge, thus the potential benefits they could have provided will probably never be realized.

One of the reasons for this problem is the difficulty of transferring know-how from academia to the industry and also the difficulty of translating scientific discoveries into practical applications that can be marketed and sold. It often takes significant resources, skills and expertise to develop a product or service from a research finding, which many Researchers or institutions may not have. Furthermore, there can be a relatively large disconnect between Researchers and the business world. On the one hand, Researchers may lack knowledge or interest in commercialization, while on the other hand, businesspeople may not understand or even have access to the latest research findings, even though part of their taxes were used to fund such research. Even if those problems were to be solved, there is still a lack of an ecosystem in order to align the interests and bring every stakeholder under one *fully decentralized* roof, as explained in Sections 2.310 and 3.4.3.

2.3.10 The need for Automated Scientific Discovery

The authors would like to highlight they do *not* want the reader to paint a black picture about the current states of scientific discovery. Such a subject is *not* black and white, so the authors do not take a nihilist approach, but they take a more balanced and probabilistic (i.e. Bayesian [1]) approach to this matter and the whitepaper in general. A lot of interesting research results have come out so far, with a positive impact on the world. Therefore, the reader should not draw the wrong conclusions. However, all of the aforementioned weaknesses remain and an improved process is necessary, which will make use of the current strengths, while also addressing these weaknesses.

In addition, the Middle Ages were necessary in order for the Renaissance to arrive. Human history is full of lessons learnt from previous mistakes or shortcomings. And one might argue that the last one or two decades have been the modern-day Middle Ages in AI and scientific discovery. In other words, humanity did not experience the level of breakthroughs before that. One of the reasons might have to do with Geopolitics (e.g., unipolar world, globalization, globalism, financialization), but for sure there are also reasons that relate to the *centralized* system that we all live in (as explained in Section 3.1), since this is *not* a system that fosters *healthy competition* and hence, innovation.

Actually, **centralization** exists inside *every step* of the process, namely:

- **STEP 1:** The *generation* of scientific discoveries and technological breakthroughs (Section 2.3)
- **STEP 2:** The *commercialization* of the research findings (Section 2.3.9)
- **STEP 3:** The growth and success of commercialized research findings, whose purpose is to enable the *funding* of even more scientific breakthroughs (Section 2.3.8) in a cyclical pattern, as shown in Fig. 14.

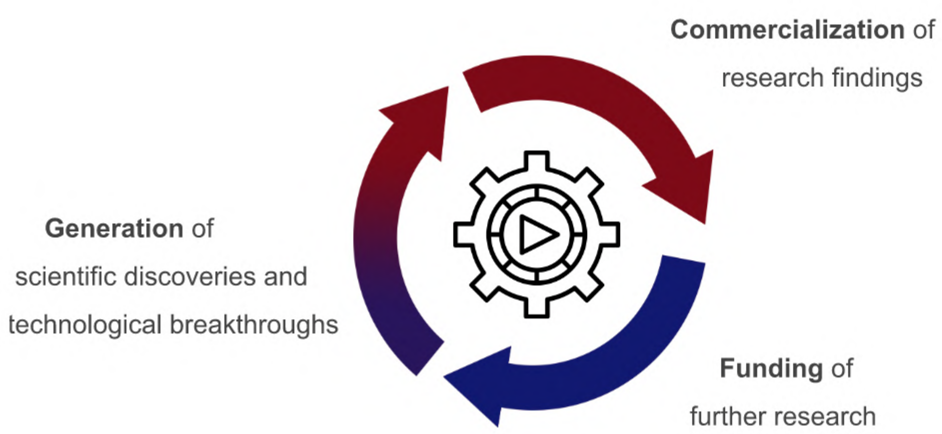


Figure 14 - The cycle of innovation.

Therefore, this gives birth to the need for a more *efficient*, more *automated* and more *decentralized* mechanism for scientific discovery, which would eliminate most of the guesswork. As a result, regarding STEP 1, the use of AI could have a tremendous impacted towards achieving **Automated Scientific Discovery**, which would remove the human "from the loop" as much as possible (as explained in [1]) and thus, freeing him to focus on more creative tasks in pursuit of his happiness. Meanwhile, STEP 2 includes multiple processes, including Product Development, Product Discovery (Section 4.6), Marketing, Sales, Customer Support, and Governance and Compliance (Section 4.7). STEP 3 requires the existence of a Marketplace (Section 4.4), for example. In other words, regarding all of the aforementioned steps, there is a need for an ecosystem, which maximizes: (i) Self-learning (Section 3.4.2), (ii) Self-correction (Section 2.2.7), and (iii) Self-funding (Section 2.3.8), all under **one fully decentralized roof** for the reasons further explained in Section 3.4.3. As a consequence, for reasons that will become clearer and more apparent in Chapter 4, the cycle of innovation serves as the *key operating mechanism* of the Ecosystem (Section 4.2) proposed in this whitepaper in order to cater the needs of the users (Section 2.1).

3 Problem

The problem statement in this Chapter is expressed in such a way that it *neither* identifies a solution *nor* does it create a bias towards any specific sub-problem.

3.1 Challenges and impediments

Currently, in order for the users to address their needs (as explained in Section 2.1) there are countless centralized "solutions" in the cyberspace (as mentioned in Sections 2.1.6 and 2.2.5), but all of them have some or all of the following challenges and impediments:

(1) multiple single points of failure

(2) no interoperability

(3) limited communication

(4) inefficient processes

(5) bureaucratic hegemony

(6) censorship

(7) no privacy

(8) no transparency

(9) no customization

(10) security vulnerabilities

Even the decentralized "solutions" (as mentioned in Section 2.1.7) face their own challenges for a series of reasons (as described in Section 3.2). However, they can be considered as *complementary* and synergistic towards the approach of the GUT-A Initiative (Chapter 5), and they can potentially be integrated into the Ecosystem (Section 4.2).

3.1.1 Web3 vs. Web2

The aforementioned challenges and impediments stem from the *centralized* nature of Web2 (or Web 2.0) in contrast to **Web3** (or Web 3.0). Various definitions for Web3 exist, while admittedly many of them are difficult to comprehend or outright nonsensical. Therefore, the authors believe that they should provide a definition that is both clear and easy-to-understand. Before providing such a definition, the authors would like to remind the reader of the well-established definition of the Web (short for World Wide Web or WWW) and its difference with the Internet. The Internet is a global network of networks *mutually* connected together, while the Web is a collection of *information* that can be accessed via the Internet and can then be read by humans. Without getting into technical details, the Internet operates on more than one protocol and more than one port. A good analogy would be to compare the Internet with the electrical grid (or power grid) and the Web with electrical sockets (or power sockets), which can be used to access electricity via the grid.

Typically, the best way to describe Web3 is by comparing and contrasting it to Web2 (and Web1), as shown in Fig. 15. Therefore, Web3 can be **defined** as a version of the Web that has all of the following characteristics:

(1) It is *dynamic* in contrast to Web1.

(2) The user to be both a consumer and a producer (i.e. *'prosumer'*) in contrast to Web1.

(3) It is based on *platforms* or *apps* in contrast to Web1.

(4) Ownership, power and control is vested upon each individual *user* over their online identities, content, data, and digital assets in contrast to Web2.

(5) Platforms, apps, but also Decision-Making are *decentralized* in contrast to Web2, thus eliminating the need for intermediaries or centralized gatekeepers.

(6) There is more openness, transparency and trustlessness compared to Web2.

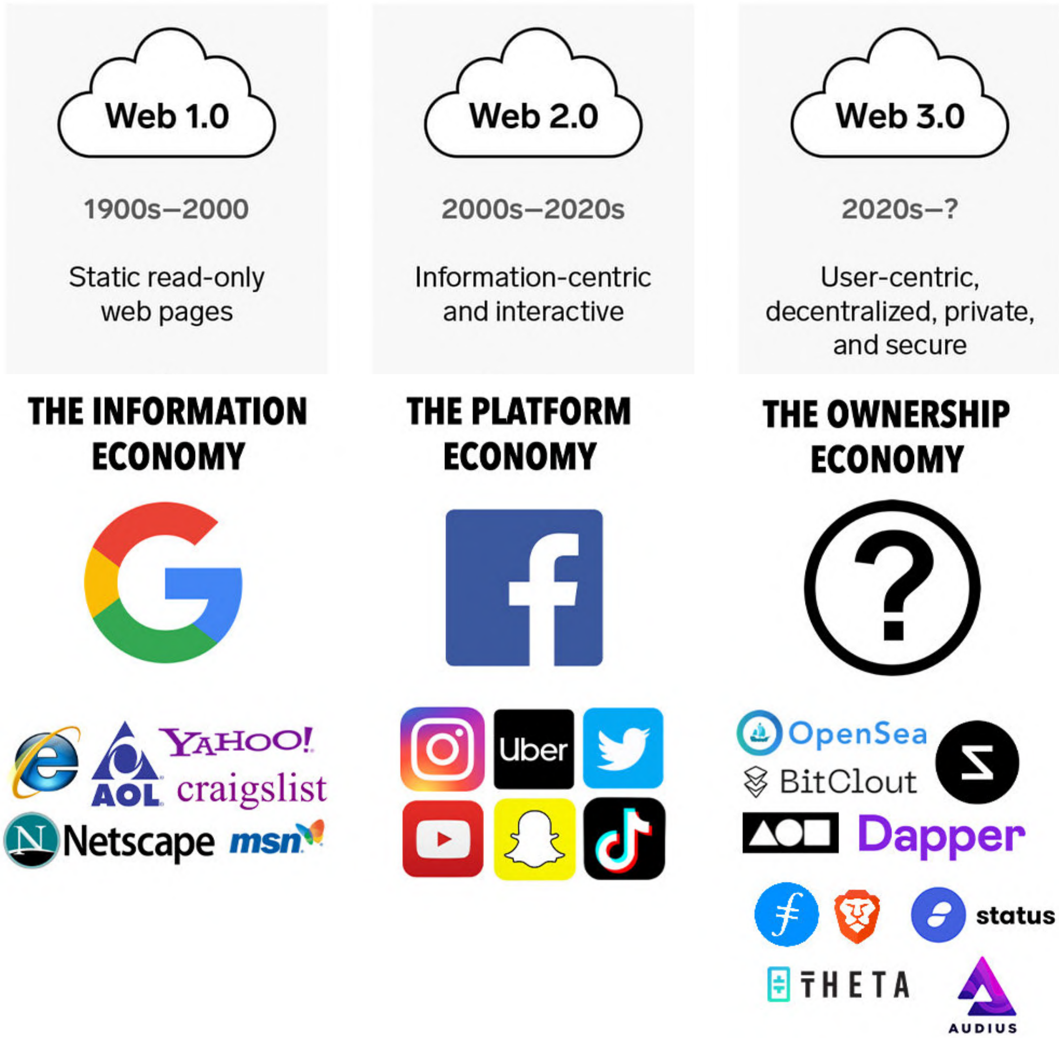


Figure 14 - Comparison and evolution of the Web from Web1 to Web3

(Trademarks belong to their respective owners).

The reader should notice that the definition of Web3 given by the authors of this whitepaper does not contain terms such as Blockchain, Decentralized Finance (DeFi), Non-Fungible Tokens (NFTs), fungible tokens, Peer-to-Peer (P2P) protocol or federated protocol. The reason is that the definition is focused on the *'what'* should be implemented in contrast to *'how'* should be implemented in order for allowing maximum creativity and not to limit or preclude alternative approaches that might occur in the future. The concept of **'maximizing for creativity'** (see also Section 3.3.3) is a *recurring* pattern and concept in this whitepaper, because it is central to the design of the whole Ecosystem as proposed by the authors.

3.1.2 No Free Lunch

In other words, the emergence of Web2 (Fig. 15) transformed the globe by providing a platform for users to produce, view and consume content, thus creating a new business *opportunity* for individuals (i.e. prosumers) to earn money. In this way, companies generate income for both content creators and website owners by advertising on such platforms (e.g., social media, news sites and search engines), with the whole hosting, infrastructure and distribution system implemented as a Platform as a Service (PaaS), thus removing most of the guesswork for the prosumer. However, as the saying goes, "There ain't no such thing as a free lunch", and therefore, these companies provide hosting, infrastructure and distribution services (Fig. 16) in exchange for personal information of the users. In the field of AI and algorithms in general, the **No Free Lunch** Theorem is relatively well-known, which follows the same principles, but for the design of specialized algorithms.

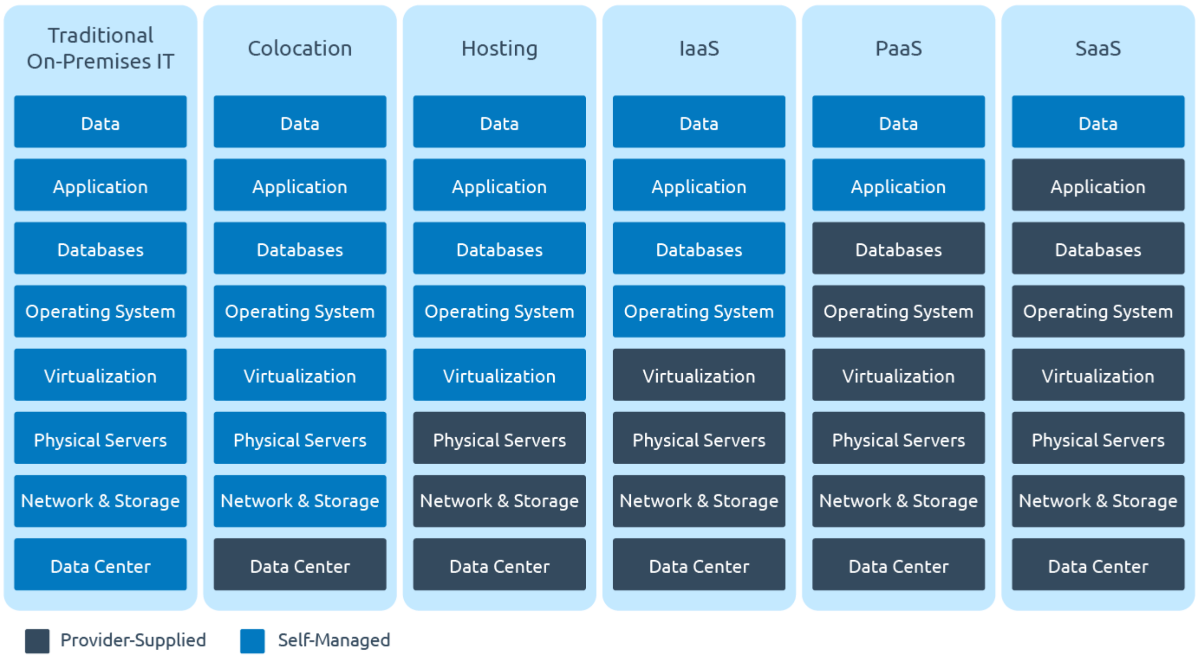


Figure 16 - Comparison of the levels of infrastructure currently provided through the Web.

These corporations act as both **rule-makers** and **rule-breakers** of the Web2 era, since they have full *power* and *control* over determining what content appears or is removed from the Web, i.e. what content is 'searchable', 'findable' or 'reachable' and what is not. For instance, Elon Musk had described this concept as "freedom of reach". Therefore, platforms such as Twitter, YouTube, and Twitch enforce their own guidelines and regulations to control content (i.e. *censorship*). Web2 is also vulnerable to internet blackouts, which could have serious consequences for individuals who depend on the internet for their livelihoods. Even the specific rules that they follow often *lack transparency*, since the platform is closed-source and these rules are arbitrarily being enforced by humans in contrast to being enforced in an automated way (such as 'smart contracts' in Web3). Furthermore, Facebook and Google have both experienced multiple blackouts during the history of Web2. All these are examples of *multiple single points of failure*, which are ubiquitous and inherent in Web2. Also, it should be clarified that while it is true that Web3 is *not* without its limitations, it still offers several advantages when compared to Web2, as mentioned above, but also in Section 3.1.3 below.

3.1.3 The lack of Fair Rewarding

One of the major drawbacks of Web2 is the lack of **Fair Rewarding** for content creators, as explained in this Section. Content creators on Web2 platforms such as YouTube, Instagram, and Facebook often struggle to monetize their content. While these platforms generate significant amounts of revenue from advertising and selling (anonymized) user data, the creators themselves receive only a small fraction of the profits, if any. In many cases, they rely on sponsorships, merchandise sales, or donations from their followers to make a living, even though they produced enough content that has benefited the platform and their followers already. This is a typical scenario for the lack of Fair Rewarding, as termed by the authors of this whitepaper.

One of the reasons for this lack of Fair Rewarding is the centralization of power and control in the hands of the platform owners. As explained in Section 3.1.2, these centralized authorities (i.e. companies owing the platform) set the terms and conditions for content creation and distribution at their sole discretion, while taking a large cut of any revenue generated from advertising, often at the expense of both the content creator, but also the follower (who interacts with the platform, but without being rewards for that). Again at their sole discretion, with almost zero transparency, they often use algorithms and data analysis to promote certain types of content over others, which can disadvantage certain creators, while damaging both the income of such a content creator and the experience of the follower. Furthermore, the ever-changing recommendation algorithms make it challenging for creators to maintain a relatively stable stream of income. It also becomes more and more usual to shadowban, demonetize and, eventually completely censor and deplatform certain content creators, sometimes without even providing a reason for doing so.

In addition, the ‘reward’ is not only about the financial compensation of the content creator and the follower interacting with the platform, but also a form of **recognition** that such users produced certain ideas, feedback, suggestions or that they have somehow contributed towards the betterment of the platform. For instance, creators invest significant time and effort in producing content targeting a specific niche audience, and they deserve recognition for their work. However, the metrics used by many Web2 platforms to measure success (such as likes, views, and followers) do not necessarily reflect the *quality* or *value* of the content. This can lead to a situation where creators feel undervalued and their work goes unrecognized. The mere fact that the vast majority of Web2 platforms opted for an **advertising** business model, does not necessarily mean that it is the only business model that is practical, workable and economically viable.

Moreover, it is also about the values of **creative control** and the **impact** of their work on society and the economy, both of which can be restricted by the arbitrary rules and their enforcement by the centralized Web2 platform, along with a level of self-censorship that creators imposed on themselves in order to protect their stream of income. The two aforementioned values used to be celebrated in a pre-Web2 era, especially by journalists and their publishers. The greatest ideas in the history of mankind went “viral” after the invention of the (mechanical) printing press. When it comes to spreading new ideas, a lot of creators prefer to focus towards the greater good and the benefit of the public, and less on their individual benefit, thus, often considering financial compensation and even recognition as either by-products of their service to the public or not important at all.

In particular, regarding creative control, creators often have limited control over how their content is presented and distributed on Web2 platforms, with platform policies around content moderation limiting *creative expression* and stifling *innovation*. Regarding the second value, content creators on Web2 have the potential to shape public discourse and influence societal norms, but the lack of diverse voices and perspectives (i.e. “outliers”) on these platforms often leads to a *homogenization* of content and an *echo chamber* effect. Additionally, the focus on sensational or clickbait content can prioritize entertainment over *educational* or *informative* content, which can have negative consequences for society as a whole. This effect usually stems from the business model chosen by the vast majority of Web2 platforms, that of advertising, as mentioned above. Web3, however, gives rise to new opportunities to further engage content creators and followers in order to reach a consensus about applying different business models on Web3 platforms (i.e. dPlats), reinforcing the concept of Fair Rewarding in terms of (a) financial compensation, (b) recognition, (c) creative control, and also (d) on the impact of the users’ work on society and the economy.

3.1.4 Web as a utility

In light of the above, the internet has become an indispensable part and a fundamental aspect of modern life for all the people of the planet, providing real-time access to information, communication, and services on a global scale. Therefore, in this sense, the Web can be considered a **utility**, due to the fact that it has become a ubiquitous and essential service that people rely on for daily tasks, much like other utilities, such as electricity and water. Hence, it should be treated as such. As explained above, the Web has already revolutionized the way people communicate, work, learn, and consume media. From Healthcare and Education to Financial Services and Self-Driving Cars, it also has the capacity and potential to further revolutionize the future of human life in both developed and developing countries.

However, in order for the Web to fulfil its full potential, it needs to be directed in the correct way. And the ones directing it in the correct way should be no other than the individual users of the Web. The Web needs to facilitate a marketplace where ideas, concepts and implementations flourish, while also ensuring Fair Rewarding (Section 3.1.3). In order to eventually facilitate such a marketplace, it needs a whole ecosystem to design, implement and run such a marketplace, while also implementing feedback from its users. Therefore, Web3 is a necessary, but not sufficient condition in order to achieve that. In other words, it is essential, but there is a series of reasons why it is not enough on its own, as explained in Section 3.2 below. It should also be highlighted that it is entirely possible for more than one idea, ecosystem and eventual implementation to exist in order for the Web to fulfil its full potential. However, the authors strongly claim that the one suggested in this whitepaper is the most optimal to be used as a starting point, for the reasons that will become clearer in Chapter 4.

3.2 Web3 is not all you need

It should be re-emphasized that Web3 is *not* a panacea for addressing all the challenges, as they were described in Section 3.1. For example, decentralization is not a black and white situation. A solution might be partially decentralized (but still not fully centralized as in Web2), as described in Section 3.2.2. Furthermore, all the current implementations of Web3 are constrained to the digital world (e.g., metaverse), thus lacking real-life experiences, as described in Section 3.2.3. In other words, in the context of this whitepaper, Web3 is a means to an end, and *not* the end itself. Hence, this is where AI (Section 3.2.4) come in, along with an ecosystem (Section 3.4.3) to generate AI Solutions (as defined in Section 2.1.1), ideally in a fully automated way (Section 3.4.2). However, such an ecosystem will have to somehow ensure that the generated AI Solutions will be trustworthy enough (Section 3.3) and account for any potential risks to humanity as a whole.

3.2.1 Dilemma of user-friendliness vs. security

In spite of being a common misconception, decentralization is *not* an inherent part of Blockchain. As a matter of fact, Blockchain is a distributed technology (Fig. 17) *not* a decentralized one, but with the provision of building decentralized (or centralized) networks on top of it. Without delving into too many technical details, the difference between a distributed and a decentralized network are briefly explained below.

On the one hand, a **distributed** network is a type of network where the *workload* is divided among multiple nodes or computers, and each node stores a copy of the data or information, so Blockchain technology is a typical example. Without any further constraints, nothing can be deduced about which user and how exercise control in a generic distributed network.

On the other hand, a **decentralized** network is a type of network where there is *no central authority* controlling the system, but rather, multiple nodes or participants work together to achieve consensus and maintain the system. Each node has a degree of autonomy and can communicate and interact with other nodes in the network to carry out various tasks. It is crucial to note that decentralized networks tend to be more resilient and resistant to censorship since there is *no single point of failure*.

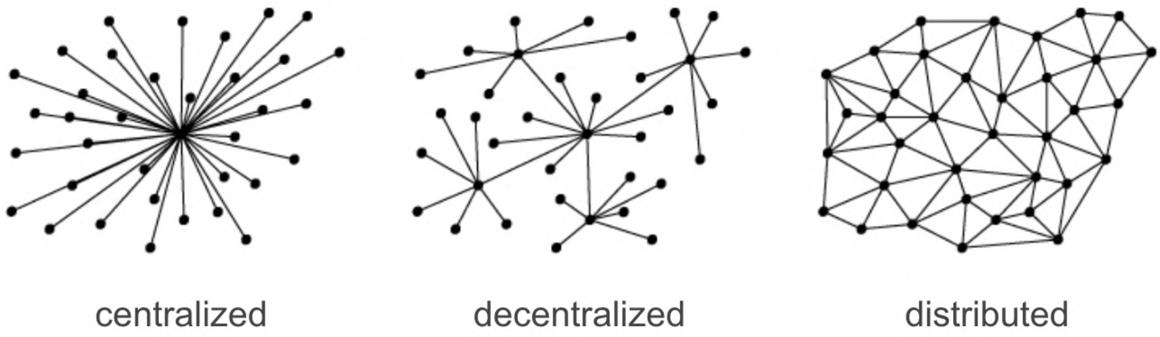


Figure 16 - Graph representation of centralized, decentralized and distributed networks.

In practice, there are always some *trade-offs* that should be considered when designing and then implementing any type of system. Even though there is common belief that there is a trilemma when considering Blockchain networks (known as the 'Blockchain Trilemma'), the authors of this whitepaper claim that in reality, it is a **dilemma** and it concerns all kinds of systems, not just Blockchain networks. So, the overall idea is that it is hard for a system to achieve optimal levels of two properties simultaneously. The authors claim that there are two *fundamental properties*, with a series of *derived properties* that stem from each one. The two fundamental properties are (a) user-friendliness (or usability or convenience), and (b) security, both of which exist in a continuum (Fig. 18). As a matter of fact, this dilemma has already been proposed one way or another [44, 45] well before the arrival of Blockchain technology, since *information systems* existed even before the arrival of the internet.

The fundamental property of **user-friendliness** is related to the following:

- semi-decentralization, scalability, privacy, anonymity, usability, UX, usefulness, effectiveness, user satisfaction, learnability, memorability, accuracy, adaptivity, customization, convenience, intuitiveness, unobtrusiveness, aesthetics, empathy, clarity, accessibility, responsiveness, interactiveness, engagement, human centricity.

The fundamental property of **security** is related to the following:

- full decentralization, automation, transparency, traceability, auditability, accountability, interoperability, fraud detection, disintermediation, asset tracking, immutability, censorship resistance, permissionlessness, trustlessness, frictionlessness, efficiency, speed, fault tolerance, trustworthiness, reliability, robustness, durability, resilience, availability, pervasiveness, ubiquitousness, explainability, data integrity, user control.

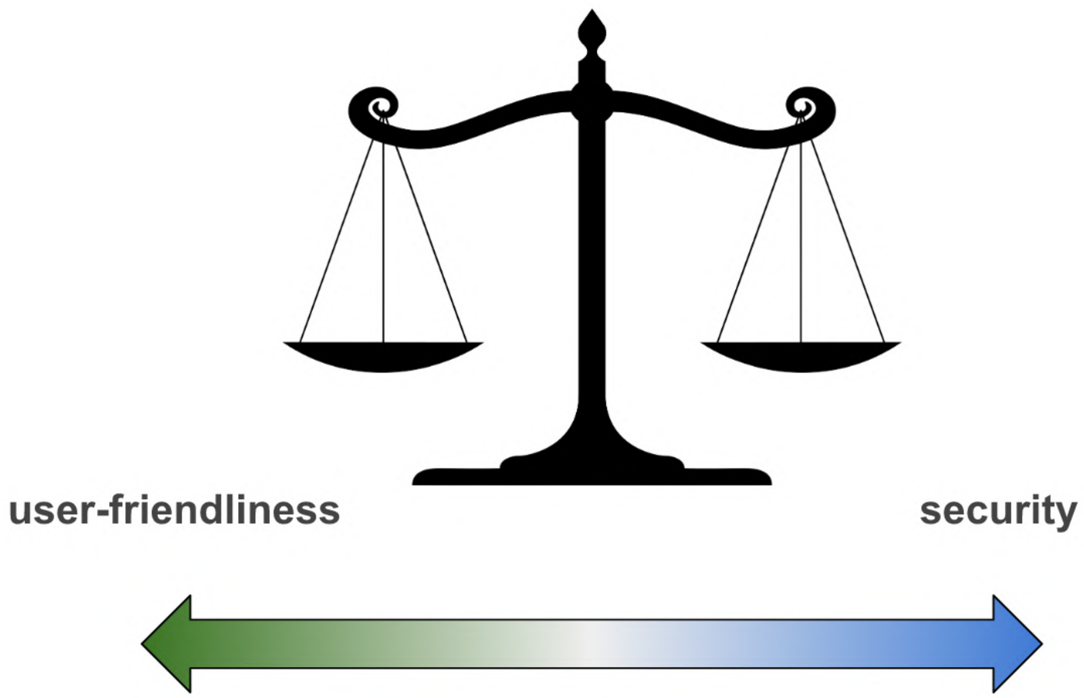


Figure 18 - The continuum between the two fundamental properties of user-friendliness and security.

In other words, the most "secure" system is an *unusable* system and vice versa. This is a typical problem in Web2. Also, one might notice something that is not in conformity with the Cybersecurity literature. Privacy is derived property of user-friendliness and *not* of security. The reason is that 'security' in the context of this whitepaper is interpreted as closer to full transparency, and hence, in this sense, full transparency is fulfilled at the expense of privacy. A derived property close to privacy is that of *anonymity*. Even though blockchain has the inherent capacity of pseudonymity, it inherently lacks the ability of anonymity, even though certain Web3 solutions are trying to indirectly implement a layer of anonymity. Therefore, while a *pseudonym* can hide a user's identity, it is not foolproof. Through some meticulous searching of a blockchain, one could possibly find out the real-world identities behind a user's wallet.

3.2.2 Full vs. Semi-Decentralization: best of both worlds

Practically, thus far, it has proved extremely difficult in pinpointing an optimal middle-ground solution in the continuum of the two fundamental properties mentioned above (Fig. 18). Blockchain networks that opted for more user-friendliness (in an attempt to achieve mass adoption) have been criticized that they are no longer fully decentralized. Conversely, fully decentralized solutions have been criticized for not being user-friendly enough, thus hindering mass adoption. In other words, at first glance, it seems like that such an optimal middle-ground solution might be infeasible to exist.

Taking the above into careful consideration, a solution could be to combine two (or more) systems, one of which is **fully decentralized (FD)**, while the other is **semi-decentralized (SD)**. It should be noted that the term 'semi-decentralized' is used in the *broad sense* to also include 0% decentralization, since decentralization can exist in a continuum.

Despite the series of benefits, the lack of a centralized authority that comes with decentralization often causes challenges in Decision-Making and governance, since it might be slower or difficult to reach a consensus and resolve disputes on important issues. Furthermore, designing and implementing effective governance mechanisms can be challenging, and there is a risk that they may not work as intended, leading to further disputes and other issues. Additionally, such systems are usually designed to be highly resistant to change, which can limit their ability to adapt to new circumstances or user needs, thus reinforcing the *status quo* and making it challenging to innovate and improve the system over time due to lack of self-correction (Section 2.2.7) and flexibility.

All these can result in confusion and inefficiencies, especially in cases in which it is critical for the whole network for a decision to be reached fast. As the proverb goes, “Many hands make light work, but too many cooks spoil the broth”. For example, in 2016, the Ethereum community faced a governance crisis when a disagreement over a software update resulted in a hard fork, splitting the community into two separate camps. This incident highlighted the challenges associated with governance in fully decentralized systems and the need for some kind of moderation and oversight to ensure *stability*, *fast consensus* and *agility*. Therefore, combining fully and semi-decentralized systems has the advantages of both faster **Decision-Making** and fairer **governance** than just using one of the two.

As it has already been hinted, a further disadvantage of fully decentralized systems is that of no moderation or oversight. Without moderation, such solutions often become a breeding ground for inappropriate content and behaviour. Without moderation, users may feel emboldened to post inappropriate, offensive or illegal content, which can harm the reputation of the system, thus discouraging legitimate users from participating in it. Hence, one of the advantages of combining fully and semi-decentralized systems is that of **fair moderation**.

A prevalent disadvantage, which is arguably the main reason that has hindered the mass adoption of Web3, is the complex nature of fully decentralized systems that make it difficult for users to navigate, understand or even access it. The steep learning curve and technical jargon associated with these systems can limit their adoption and potential impact, particularly for those who are not well-versed in technology. Furthermore, because such systems often rely on a distributed network of nodes rather than a centralized organization, there may be limited or no support available for users who experience technical issues or other problems, which makes it difficult to troubleshoot issues or get help when needed. Also, such systems can compromise user privacy due to the lack of anonymity, since the transactions and data stored on the system are visible to all nodes in the network, as explained above.

In contrast, a combination of a semi-decentralized gateway that is **user-friendly**, convenient and intuitive (plus all the other derived properties, as explained above) in combination with an underlying hosting, infrastructure and distribution system (Section 3.1.2) that is fully decentralized can achieve the best of both worlds.

In conclusion, combining fully and semi-decentralized systems in one solution can be beneficial in a number of ways. By utilizing the strengths of both systems, it is possible to create a more efficient, scalable, and flexible meta-system that is also more secure, thus achieving the best of both worlds. A question that remains is *'how'* should these systems be combined in an architecture that would achieve the desired outcome. For instance, it is important to consider and then address how does someone design such systems in order to make self-correcting mechanisms *inherent* to their architecture, so that *minimal* human supervision is needed, while also making sure that the participants to such a meta-system find it *trustworthy* (Section 3.3) enough to use. So, Web3 is not enough for this.

3.2.3 The lack of real-life experiences

Since the Web relies on a (computer) network, and since it also generates or enhances (social) networks and communities, there are established *computational* methods [1] to mathematically examine its behaviour, such as Network Science, Complex Systems and Collective Intelligence, which are the same techniques used to study the human brain [36], for example. In other words, the Web can be considered as the "brain" of the Internet or even the brain of human knowledge, history and whole existence. Conversely, our own brain might need to adapt and naturally evolve in order to interact with and utilize the full potential of such a Web. For instance, a future version of the Web might have to account for **real-life experiences** which are more **meaningful** and **natural**, while getting away from traditional displays and computer monitors as much as possible in order to counteract potentially harmful consequences and even permanent damage to the human brain [37] caused by this artificial way of communication and interaction.

Therefore, Web3 lacks a way of addressing on its own the lack of meaningful and natural interactions among humans. So, complementary approaches, strategies and technologies are necessary to address them *concurrently* with the development and deployment of Web3 solutions. In addition, as already mentioned in Section 3.2.2 above, a combination of fully and semi-decentralized systems is also necessary. Otherwise, Web3 runs the risk of being demonized by a portion of the population, which might cause irreversible damage to: (a) its adoption by the general audience or (b) reaching a critical mass so that it can eventually be considered successful or even meaningful. Nonetheless, even after considering and addressing all of the aforementioned issues related to Web3, the authors of this whitepaper claim that there are still pieces of a puzzle that are missing in order to fulfil the full potential of any version of the Web, as explained in Section 3.2.4 below.

3.2.4 Web3 and AI

Despite being a common misconception, AI is *not* an inherent part of Web3 (as defined in Section 3.1.1). However, the authors of this whitepaper claim that it might be one of the missing pieces of the puzzle in order for the Web to fulfil its full potential (as defined in Section 3.1.4).

AI has already made significant contributions to various domains, and its potential in Web3 is immense. More specifically, AI can help improve the efficiency of Web3, since it can sometimes become slow and inefficient due to network congestion combined with the absence of a centralized authority to manage the network. AI has the potential to help address this challenge by optimizing the use of network resources and improving the speed and reliability of transactions. For instance, AI can be used to predict network congestion and allocate resources accordingly, thus reducing transaction times and improving overall network performance. Also, it can provide **recommendations** regarding better *network governance* for a more data-driven Decision-Making (Section 4.1.6).

Moreover, AI has the ability to enhance the **user experience**. Web3 is designed to be user-centred, and AI can help achieve this by providing *personalized recommendations*, insights and alerts. For example, after obtaining user approval, AI can analyse a user's browsing history and preferences to recommend relevant content or products, thereby enhancing the user experience and engagement. Furthermore, AI can assist enhancing creativity and innovation, while fostering a culture of experimentation and discovery. AI can help achieve this by providing tools and frameworks for generating novel ideas and solutions. For example, AI can be used to analyse user behaviour and preferences to suggest new **products or services** that meet their needs and desires (Section 4.6).

Furthermore, AI can contribute towards cutting through the noise. In the digital age, **content creation and consumption** have become integral aspects of our lives. However, content creators and consumers alike encounter similar obstacles that impede the impact of newly produced information. In general, there are three main factors that affect the *effectiveness* of content: (a) the speed of content production, (b) the speed of content dissemination, and (c) attention and content filtering.

The **speed of content production** poses a significant challenge for content creators. With the advent of Web2 platforms, the demand for fresh and engaging content is insatiable. For *content creators*, speed is essential for staying competitive in a rapidly evolving digital landscape. They must continuously generate material to ensure *relevance* and *engagement* with their audience, while also allowing them to maintain a consistent presence, build brand recognition, and attract and retain a larger audience base, thus and potentially boosting their reputation and revenue generation.

For *consumers*, fast content production offers immediate access to information and entertainment, while also enhancing the overall user experience, as consumers can rely on regular updates and fresh material. However, this pressure to produce content rapidly can sometimes lead to compromises in quality, depth, and accuracy. AI can help alleviate such issues in various ways, such as using **AI-generated** content (for both text and images) or identifying patterns and trends that resonate with their target audience in order to create more impactful and engaging material in a shorter timeframe.

The **speed of content dissemination** is crucial for *content creators*, since rapid dissemination enables them to reach a wider audience in a shorter amount of time, while also maintaining a competitive edge since being the first to share valuable or unique content can make a significant difference, thus potentially gaining a loyal following, and attracting opportunities for collaborations, sponsorships, or partnerships.

On the *consumer* side, rapid content dissemination also facilitates real-time discussions, fostering engagement and interaction among consumers, which further enhances the overall user experience. AI can have a huge impact on the speed of content dissemination by leveraging AI systems that monitor and scrape social media platforms, news outlets and other online sources to **segment and identify** emerging trends, thus providing timely and relevant content that aligns with current trends, and maximizes the chances of such content being shared and disseminated rapidly. Also, introducing specialized **chatbots** (Section 3.3.1) for the consumers can speed up content dissemination in a myriad of ways, such as hassle-free and instant information delivery, quick content navigation and search, and intelligent content curation customized to their preferences and interests.

The factor of **attention and content filtering** presents a significant obstacle to the impact of content. On the *content creator* side, In a world filled with numerous distractions and competing information, capturing and maintaining individuals' attention has become increasingly difficult. People are bombarded with a constant stream of content from various sources, leading to *information overload* and limited cognitive resources to process it all. Consequently, individuals tend to filter and prioritize content based on personal interests, preferences, and relevance. This filtering process often results in the neglect or dismissal of valuable information that may have had a significant impact if given the opportunity. The rapid dissemination can also contribute to the challenge of reaching the intended audience effectively. With countless content pieces vying for attention, it becomes challenging for a specific piece of content to *cut through the noise* and reach its target audience in a meaningful way.

Additionally, on the *consumer* side, the fleeting nature of social media platforms and the constant influx of new content can make it difficult for any single piece to gain prolonged *attention* and make a lasting impression. Information spreads at an unprecedented rate in the digital realm, creating a saturated landscape. As soon as content is published, it competes with countless other pieces for attention. The constant influx of new information *overwhelms* consumers, making it difficult for any single piece of content to stand out. This rapid dissemination also contributes to the fleeting nature of information, as it quickly becomes buried beneath a never-ending stream of new content. Also, content filtering becomes essential for consumers, as they must sift through vast amounts of information to find what is relevant and valuable. The challenge lies in identifying high-quality and reliable sources amidst the sea of content.

AI can contribute towards addressing such challenges by using AI-powered **recommender systems**, which can curate and recommend personalized content based on user preferences, behaviour and demographics. By understanding individual interests, such AI systems can suggest relevant content to *consumers* in real-time, increasing engagement and ensuring that they receive content that aligns with their preferences. Also, by leveraging such AI systems, *content creators* can provide customized and timely content suggestions to their target audience. This helps consumers discover new content quickly, while ensuring that the recommended material is of high quality and relevant.

Overall, the potential for innovation and advancement in Web3 ecosystems resulting from combining them with AI is limitless. AI can enhance decentralized platforms by automating processes, improving user experiences, and enabling intelligent Decision-Making, while also providing personalized content, and facilitating more efficient governance systems. However, lack of *trustworthiness* could hinder mass adoption, as explained in Section 3.3.

3.3 Roadmap towards Trustworthy AI

Trustworthy AI [39, 40] is an active and relatively new area of multidisciplinary research, which necessarily touches philosophical aspects too. One might not need to go all the way back to Plato and Aristotle, but there are philosophical disputes that stem from practical concerns, as described below. It is also an area that necessarily touches upon technical aspects too, so the reader is expected to have some familiarity with some technical jargon widely used in AI when going through this specific Section.

Furthermore, the principles that are going to guide Trustworthy AI will have to be enforced in a *decentralized* way (and not by any centralized authority), and will have to be enforced as '**soft constraints**' (i.e. as *flexible* rules that can practically be violated, but with penalties associated with breaking it) in order to account for the *human error* in both determining and then enforcing these principles. Such principles can be defined and elucidated from the following perspectives:

(1) Technical perspective (for example, as described in [1])
(2) User perspective (as defined in Section 2.1)
(3) Social perspective
(4) Political perspective
(5) Economic perspective
(6) Environmental perspective

The rest of this Section takes into account all of those perspectives in order to propose how Trustworthy AI should be implemented as part of the proposed Ecosystem. It should be highlighted that for the purposes of this whitepaper, the term 'trustworthy' refers to maximizing the level of trust by each (human) user, while eliminating or limiting factors that could potentially contribute to generating or amplifying scepticism, suspicion and mistrust towards AI, thus inhibiting its mass adoption. This gives rise to certain risks (Section 3.3.1), and ethical concerns (Section 3.3.2) associated with such mass adoption, but also strategies to balance and mitigate risks (Sections 3.3.3 to 3.3.10) in order to reap the vast amount of benefits from the research and development of AI systems.

It is absolutely crucial to highlight that this Section and the whitepaper in general, do *not* make any claims on how the policymakers (or any other stakeholders) should address AI regulation or not. Instead, it merely proposes the **practical** creation of a general framework to conduct *collaborative* and *controlled* trial-and-error *experimentations* in a trustworthy way in order to reach mass adoption. In general, the authors do *not* take any such political positions and do *not* make any political statements, nor should the whitepaper be interpreted as such. It should also be made clear that the concepts of 'AI safety' and 'security for AI' are both *outside* the scope of this whitepaper, but not necessarily outside the scope of the proposed Ecosystem (Section 4.2), since, for example, safety typically relates to optional and mandatory standards (and sometimes regulations) that are *specific* to the area of application (e.g., mobile robots vs. aerial robots), rather than generic to all AI systems.

3.3.1 The dangers of Stochastic Parrots

With the recent rise of Conversation AI platforms [1], such as ChatGPT (or GPT 3.5) by OpenAI, the dangers of the so-called 'Stochastic Parrots' have been promptly highlighted by the AI community [49 - 51]. Despite some legitimate concerns and genuine criticism, unfortunately, this led to the release of a letter by the 'Future of Life Institute' asking for a six-month minimum moratorium on "training AI systems more powerful than GPT-4" [52], thus, sparking irrational fears (Section 3.3.10), while also calling into question whether the signatories (which include famous billionaires and IDE entrepreneurs) were actually asking for such moratorium in good faith or whether they had ulterior motives.

Regardless, the original paper by Bender et at. [49], which was actually peer-reviewed (whose flaws as a process are described in Section 2.3.4), is full of strong political biases (Section 2.3.3) of highly *controversial* and *politically polarizing* issues, such as references to: (a) "diversity and inclusion", which is deemed illegal in most western countries except the United States of America, which allows some form of "Affirmative action" in specific cases, and (b) anthropogenic "climate change", which neither has scientific consensus nor allows equal and fair funding (Section 2.3.8) to "heretics" (Section 2.3.2) to conduct experiments and publish their work. It is also important that their paper only presents a dogmatic opinion of such issues, without any references to contradictory opinions, to the point of presenting them as undisputed "facts". However, it remains questionable if this lack of objectivity (which is unfortunately widespread) was intentional or not.

Moreover, a further problem, which is also widespread, is using the particularly problematic term "Large Language Model", which *neither* has formal definition *nor* should it used due to its nonsensical nature. What is "large" supposed to mean? How "large"? And what procedure does one use to measure a model? Is model compression allowed? A more important problem is that ChatGPT (GPT 3.5) and other versions are *not* Language Models [1]. They are typical examples of **Conversational AI** platforms, formally called 'Non-Task-Oriented Dialogue systems', 'Question Answering (QA) systems' (as defined and explained in [1]) or more colloquially 'chatbots'. It is likely that such Conversational AI platforms might have used Language Models, but: (a) they were thrown away after Training, (b) nobody can be certain if Language Models were indeed used and how, since it is part of a closed-source software with no accompanying paper other than some blog posts.

No matter if Language Models are used or not, the lack of (a) open-source **code**, (b) trained **models** and (c) open **data** violated some of the main desiderata of Trustworthy AI, such as transparency, openness, reproducibility, auditability and explainability. As a matter of fact, this is actually where most dangers of the 'Stochastic Parrots' stem from. If one cannot examine the original dataset or datasets used for training such models, then it is not possible to know how many limitations the model has, what they are and what is their nature. As the famous saying goes, "Garbage in, Garbage out" (GIGO). In other words, soundness implies validity, but validity does not imply soundness, which implies that if an argument is 'sound' (meaning that it is 'valid' and has true premises), then the conclusion must logically follow from the premises, but not necessarily the other way round. Furthermore, the choice or quality of the dataset might have crucial ethical implications (Section 3.3.2), such as in healthcare and governance, while it might also affect the level of creativity (Section 3.3.3) that the model is able to reach. Such models might also run the risk of encoding political bias, since it might include only authors arguing only in favour (or only against) a certain topic. Additionally, such checks and quality control should be as *automatic* as possible, since any human curation would potentially introduce new biases to the whole AI system.

More specifically, in the case of Stochastic Parrots, the trained models might produce content that seems human-like and very convincing, but: (a) it might not be factual at all (such as the case of ChatGPT and all GPT versions at the time or writing), (b) present only one side of a questionable, controversial or heated debate, (c) present content in absolute confidence or certainty, since the model did not capture or does account for uncertainty (Section 3.3.5), thus misleading the human user to put undue reliance on such content or information, and (d) does not include references or citations to original sources in order to back the content or the claims made. Hence, the characterization 'parrots' was given to them, while 'stochastic' refers to not producing the exact same outcome every time.

It is worth mentioning that the issue of Stochastic Parrots extends well beyond the way the original authors framed it. More specifically, it extends to modalities other than text, such as images and speech. The root cause of this issue is due to the way that the underlying *generative models* (see Section 3.4.2 and [1]) are being implemented, which can be mitigated through techniques such as *grounding mechanisms* (see Section 3.3.6 and [1]). More generally, either a new source of information (i.e. context) or a learnable model must somehow be integrated into such a Generative AI system in order to *guide* or *ground* the system towards generating more sensible and aligned output (Section 3.3.7). However, the reader should *not* misinterpret this as merely using multiple modalities.

In view of the above, there is a need for the practical creation of a general framework that will distinguish realistic dangers from irrational fears, while also addressing such dangers inside an *isolated* sandbox that allows conducting *collaborative* and *controlled* trial-and-error experimentations and simulations. The remaining Sections of this Chapter address the creation of such a framework. Once again, it should be reiterated that the authors do *not* make any claims on how the policymakers (or any other stakeholders) should address AI regulation or not. Also, the authors neither support nor oppose the aforementioned moratorium.

3.3.2 Ethical dimensions of AI

This Section deals only with ethical concerns that the authors consider to be both *legitimate* and made in *good faith*. In addition to legitimate ethical concerns (most of which are addressed in Sections 3.3.3 to 3.3.9) related to AI, there is also a series of irrational fears about it, as described in Section 3.3.10, which the authors suggest being addressed as described in that Section.

In general, Ethics is a highly **subjective** topic. Something that is considered "ethical" in most Western countries (i.e. in accordance with the Judeo-Christian Ethics), might not be considered "ethical" in the Far East (i.e. in accordance with naïve dialecticism) and vice versa. Even within the same country, ethnic group or even family, the value system of each person typically differs, because value system is a totally *individualistic* concept. Just like each and every Researcher (as explained in Section 2.3.3), each and every person has his own *biases* to a certain degree. As a result, defining when an AI system is behaving in an "ethical" way and when it does not is an extremely ill-posed and ill-defined problem.

To make the situation even more complex, hardly ever are self-proclaimed "experts" in Ethics *polymaths* (as explained in Section 2.3.6), when compared to Ancient Greeks, for example. In some cases, "experts" in Ethics claim to have "expert" knowledge in AI Ethics, but barely know how to handle a laptop (let alone connect to a server in the cloud or implement and train a Machine Learning model). Even if this fact is ignored, there is a huge conflict of interest between these self-proclaimed "experts" and a rational assessment of AI due to the fact that it directly impacts the among of taxpayer funding they receive (as explained in Section 3.3.10). So it would not be far-fetched to call them modern-day *charlatans* and the whole situation a travesty of Ethics. A typical illustration of their lack of understanding is confusing methods and algorithms (e.g., Backpropagation) with models (e.g., Artificial Neural Networks), by incorrectly calling such models "algorithms" (as in [54]). The reader should not misrepresent this as requiring some "expert" knowledge in AI in order to make arguments or voice concerns. However, a certain *minimum* technical knowledge, understanding and experience is expected in order for someone to be taken seriously, but certainly, no one having such technical skills below a minimum standard should misleadingly present himself to the public as "expert". Therefore, such literature is defined by the authors to be outside the criteria for inclusion in this whitepaper (but not necessarily outside the proposed Ecosystem).

On the one hand, AI has the potential to create tremendous benefits (as described in Sections 2.1.1 and 4.10). On the other hand, there are potential risks and unintended consequences associated with the mass adoption of AI (as described in Section 1.1). The ethical implications of AI are crucial in determining how these technologies are implemented and used. Ethical concerns need to be addressed at every step of the process, including the stages of research, design, development, and deployment. Disagreements or competing strategies on how to address those ethical concerns should all be welcome, provided there is an isolated sandbox to conduct *collaborative* and *controlled* trial-and-error *experimentations* similar to how nuclear research is conducted at CERN, as an illustration. This gives rise to the need for an ecosystem to perform such experimentations, as further explained in Section 3.4.3.

Nonetheless, it is crucial to be highlighted that while ethical concerns surrounding AI are of utmost importance, they also constitute a grey area. Usually, there are no clear-cut answers to ethical dilemmas that may arise from the implementation and use of AI. A use case which lies in the grey area of such ethical concerns is the debate on whether it should be given the ability to tell lies or not. While some might argue that AI systems should *not* be allowed to lie, there are legitimate and practical reasons for allowing this ability, just like a young child learns when to tell lies and when not to. For instance, if a child lies to their parents and is caught, then they will most probably learn the important lesson that lying is not always beneficial, and in some cases, it may have consequences. In addition, the child can also learn the importance of sometimes being allowed to 'white lies', provided that *not* telling such lies: (a) would make it sound impolite towards Alice (a human), or (b) might put the life of either Alice or Bob (another human) or both in danger. Similarly, the same applies to the learning process of an AI system.

Furthermore, in the same use case, an advantage of allowing an AI system (but not necessarily all AI systems!) to sometimes tell lies is to enable it to better understand human behaviour, thus developing some form of sentience and emotional empathy (Section 3.3.6). So, if an AI system is designed to recognize when humans are lying, then it can use this knowledge to improve its interactions with humans. For example, an AI system that can tell when a human is lying can learn to respond appropriately, such as by asking follow-up questions or seeking clarification.

However, it is crucial to note that there are risks associated with allowing an AI system to tell lies. For instance, if an AI system is designed to lie, it may deceive humans and cause harm. Therefore, it is essential to ensure that there are specific *mechanisms* that allow an AI system (again, not necessarily all AI systems) to learn: (a) how to identify, asses and mitigate such risks on its own in real-time before taking an action to lie (or not), (b) to lie only in specific situations after performing such a Risk Identification and Assessment, (c) perform Risk Mitigation before proceeding in such an action to lie (or not). The authors claim that maximizing for creativity (as explained in Section 3.3.3) is a good candidate for such a mechanism, provided that it is implemented in the general framework of Trustworthy AI, as proposed in this whitepaper, which takes a *holistic* approach.

3.3.3 Maximizing for creativity: there is no box

Beyond ethical concerns, the whole point of AI is to learn *patterns* from existing data in an *automated* way in order to extract meaningful and actionable insights that can then generalize well to previously *unseen* data, while also being robust. A direct implication of this is that a certain level of **creativity** (or level of 'surprise') is expected from an AI system. In other words, similar to human pedagogy, an AI system should have the capacity to demonstrate a relatively accurate and deep understanding of an underlying phenomenon, situation or concept (observed by the original data) by generating *original* or *novel* content (e.g., labels, images, text, audio, game, experiences) in order to somehow prove that it has acquired such a deep understanding, just like an examiner would test a human candidate in a spoken or written way or just like a mother would test his son if he has learnt his lesson. More detailed examples about the lack of such creativity are presented below in order to emphasize the need for such creativity.

Although intuitive, the use of RL from Human Feedback (RLHF), for instance, is *antithetical* to the whole point of AI, i.e. *automate* the process as much as possible by removing the human "from the loop" (as explained below and in [1]), and to allow the AI system to learn as *creative* as possible (i.e. with minimal human supervision during Training), but also in order to ensure *scalability*, *efficiency* and *cost-effectiveness*. Such RLHF methodologies have found widespread use in Stochastic Parrots, such as ChatGPT by OpenAI, but these methodologies are some of the root causes of the limitations, shortcomings and also dangers (as explained in Section 3.3.1) towards Trustworthy AI. Another example that is antithetical to the whole point of AI is the relatively recent phenomenon of vast amounts of Tech startups popping up like mushrooms in the domain of manual Data Labelling by human annotators (often called 'clickworkers' or 'crowdworkers'), such as Labelbox, Sama and a company ironically called Scale AI [53].

In addition to the scalability and technical issues, such "solutions" are also problematic from a **Product Management** and **business** perspective, since they often run the risk of providing a technically sound "solution" (albeit inefficient) to a *problem* that the users or customers do not really have, but might think they have. For example, when asked about customer input in the development of the Ford Model T, Henry Ford reputedly said, "If I had asked people what they wanted, they would have said faster horses". Regardless of whether Ford actually said those words or not, they still carry a valuable meaning, since the quote is often used to support the idea that customers are typically oblivious to their needs. Nonetheless, this should not be misinterpreted as not receiving any kind of feedback from or not listening at all to customers. However, it should be construed as making the right questions in order to understand their needs and pain points instead of focusing on blindly implementing “solutions” they might come up with or recommend, since coming up with solutions lies in the expertise of the solution provider, and not that of the customer.

More specifically, imagine a movie review app employing such a Data Labelling company in order to annotate an unlabelled dataset in the context of Sentiment Analysis (i.e. classify whole reviews or individual sentences as positive, negative or neutral). However, there are several reasons why Data Labelling might eventually not be useful in such an app, including the following:

(a) The *quality* of the unlabelled data varies greatly depending on the source of the reviews (e.g., some reviewers might be more reliable than others, and reviews from different countries may vary in terms of language and cultural context).

(b) The sentiment of a movie may not necessarily reflect its quality or value as a work of art, while one user might perceive such sentiment in an *opposite* way than another

(c) The sentiment is an overly *summarized* information, so some users may prefer, for example, to have a human-like conversation that would provide them with a detailed analysis and criticism of a movie, rather than just a simple positive or negative rating.

(d) The sentiment is only a single modality (i.e. text), thus, having serious limitations in conveying the full meaning to or creating some sort of empathy for the user in contrast to a multimodal dataset (e.g., a review that contains sarcasm or irony may be difficult to detect through text analysis alone).

(e) If a user is indecisive between two movies and they both have approximately the same amount of positive (or negative) reviews, then the sentiment alone proves pretty much useless in assisting the user in his Decision-Making process.

The aforementioned issues generalize over a vast amount of tasks using single modalities. One might naïvely attempt to still employ human annotators, but for *multiple* modalities. However, this makes the whole situation even more complicated for the annotators, even more expensive, even more prone to human errors, even more biassed, even less efficient, even less automated and even less scalable.

Therefore, all of the above give rise to allowing AI to 'maximize for creativity' by removing the human "from the loop" during Training. Nevertheless, one might argue against the ability of allowing it to have such a creativity, but instead, argue in favour of '**handicapping**' all future AI systems. For example, he might oppose allowing an AI system to somehow learn or teach itself (Section 3.4.2) the ability to tell lies. There are convincing ways against this form of handicapping, though, as explained in Section 3.3.2. In general, in the same way it would seem cruel and unintuitive (to say the least) for a mother to handicap the legs and the brain of her son in order not to kick people or tell lies respectively, the authors claim that the human (e.g., an AI Architect or AI Researcher) should refrain from handicapping AI. Instead, it should be allowed to *learn*, *grow up* and *mature* just like a young child.

Diving into slightly more technical details, the learnt patterns (or discovered knowledge) can be stored either: (a) in the form of *parameters* in a computational model, (b) in the form of *rules* in a rule-based (symbolic) model, or (c) in a combination of the two leading to a *hybrid* AI system (as proposed in [1], for example). Even though traditionally symbolic models made use of 'handpicked' rules, extracted manually by human "experts", rules can be made 'learnable' (or trainable) too, as formalized by GUT-AI theory [1].

All of the above give rise to the need for a *white-box* method for measuring and assessing the creativity of an AI system. Even though black-box methods can also be proposed and used, such methods would violate the desiderata of Trustworthy AI, such as transparency, explainability, and steerability (Section 3.3.7), all of which could generate or amplify scepticism, suspicion and mistrust towards AI, thus inhibiting its mass adoption. To the best of authors' knowledge, the only metric for AI creativity proposed thus far is the **Upper Bound on Creativity (UBOC)**, which has been proposed and mathematically formulated by Kourouklides [1].

It should be noted that the closest concept to 'maximizing for creativity' is possibly the "explicitly non-committal attitude" as briefly mentioned by Sheldrake [32], but it was conceived in a different context than in [1]. The concept of 'maximizing for creativity' was partly inspired by the concept of 'There is no box' which refers to the Creative Thinking and Reasoning Process in humans, as explained in [1].

Some of the main *factors* that can affect the UBOC metric include the following:

(1) **Quality and quantity of data:** The quality and quantity of data that an AI system has access to can potentially have a significant impact on its ability to generate creative solutions (to the *overdetermined* system of equations). If the data is limited or of low quality, the AI system may struggle to produce truly original and innovative solutions. In the case of Supervised and Unsupervised Learning (including Self-Supervised Learning), this includes both (a) how large, and (b) how 'diverse' the *dataset* is in order to represent a wide range of creative possibilities. In the case of RL, this includes how complex the *environment* is. In all cases, it includes the number of modalities, such as image, text and audio. (cf. Multimodal Machine Learning in [1]).

(2) **Size of the set of candidate features:** When an AI system has access to a larger set of *features* (also called embeddings or representations) to choose from (Fig. 11), it has more options for representing and also interpreting data (in the new feature space), which can lead to more innovative and creative solutions. For example, an AI system for Speech Emotion Recognition that has access to both *linguistic* and *paralinguistic* features of speech may be able to recognize more complex patterns and generate more creative interpretations of the original data (in the input space).

(3) **Degree of human control during Training:** The higher the degree of human control over the AI system during Training, the *lower* the UBOC metric, since it runs the risk of being instilled or reinforced *a priori* biases by a specific human (e.g., the AI Architect or the AI Researcher), thus 'contaminating' its creativity. In other words, AI systems with more autonomy and less human intervention (during Training) may be able to push closer to their boundary of creativity (Section 3.4.2), while AI systems whose Training is more closely supervised by humans may be more constrained in their solutions, and more prone to human error. However, the opposite is true for the phase of Inference, as explained below.

(4) **Level of user interaction during Inference:** In contrast to Training, the higher the level of user interaction, the *higher* the UBOC metric. More user interaction and feedback (Section 3.3.9) can potentially push the creative limits of the AI system, while also generating more novel and innovative solutions.

(5) **Level of uncertainty:** The level of uncertainty (Section 3.3.5) or randomness introduced into the AI system can also affect its UBOC. Higher levels of uncertainty may enable the AI system to generate more creative and unexpected solutions.

(6) **Heterogeneity and number of AI systems:** When multiple AI systems (e.g., RL agents or Ensemble Learning) with different capabilities and approaches are *combined*, they can create a more diverse and dynamic AI meta-system, with the

potential to generate more innovative and creative solutions. In addition, heterogeneity (e.g., heterogeneous RL agents vs. homogeneous ones) can also assist an AI meta-system in better capturing non-linearities, which can improve the creativity and diversity of solutions generated by the whole AI meta-system, since it combines multiple methods (i.e. algorithms) and models.

(7) **The Category of AI:** The choice of Category (e.g., Supervised Learning, RL) can have an impact on the UBOC, since each Category is associated with certain limitations or constraints on creativity. For example, Supervised Learning is limited by the ability of the human to choose (a) the 'right' dataset (e.g., large enough and 'diverse' enough) during Training, and (b) also his ability to label the dataset correctly. However, Self-Supervised Learning is not limited by (b). Meanwhile, the UBOC metric in RL can be affected by the range of rewards and punishments available to the AI system. Therefore, if the range is narrow or limited, the AI system may not be able to generate truly original or innovative solutions.

3.3.4 Customizable AI

More generally, one of the expected advantages of AI and other new technologies is the **customization** of technological solutions to the specific user. For example, it is expected from a Conversation AI system (similar to ChatGPT by OpenAI) to generate a series of different *personalities*, which could have been implemented by using multiple *heterogeneous* agents in the context of Multi-Agent RL (MARL), as an illustration (even though, at the time of writing, ChatGPT does not offer such a functionality). The reason is that previously a product had to be as *identical* or as *homogeneous* as possible in order to achieve economies of scale (i.e. enable increase in scale by decreasing cost per unit of output).

However, this constraint is relaxed (or eventually eliminated) by the arrival of new technologies that make customization more efficient and cost-effective. One way of achieving this is through maximizing for creativity (Section 3.3.3), which ultimately allows the machine to learn directly from data in a relatively more *automated*, *unbiased* and *scalable* way by eliminating most manual labour by the human annotator. Other ways of achieving such customization include:

- accounting for, quantifying and embracing uncertainty in life (Section 3.3.5)
- enabling sentience and emotional empathy for AI (Section 3.3.6)
- making AI more steerable, explainable and aligned (Section 3.3.7)
- enhancing collaboration between the user and the multiple RL agents (Section 3.3.9)

Hence, customization of AI systems not only enhances UX and therefore user-friendliness (Section 2.2.5), but also contributes towards a greater level of trust and acceptance of AI by the public, which potentially leads to more benefits for all humans.

3.3.5 Embracing uncertainty and stochasticity

What many people fail to understand is that **uncertainty** is an inherent aspect of AI systems, as they operate in complex, nonlinear, dynamic environments with incomplete and uncertain information. In the context of Trustworthy AI, embracing uncertainty involves acknowledging that AI systems cannot always provide accurate predictions or decisions and that there is always a degree of uncertainty (or confidence) associated with their output. This uncertainty can arise from various factors, such as (a) incomplete, noisy or missing data, (b) model misspecification or (c) unforeseen events. Hence, by incorporating probabilistic models that account for uncertainty in the input data, AI systems can be made more reliable, robust and resilient to unexpected events or anomalies, thus preventing system failures or errors that might have detrimental implications, while also enabling steerable AI (Section 3.3.7).

Moreover, **stochasticity** is an interrelated aspect with that of uncertainty that should also be embraced in the context of Trustworthy AI. In contrast to Deterministic Modelling, stochasticity refers to the inherent randomness in AI systems, which can arise from factors such as random initialization of weights in an Artificial Neural Network, for example, or the random selection of data samples in order to construct the dataset to be used during Training. Accounting for stochasticity has the ability to help prevent AI systems from being (statistically) biased or overfitted to specific datasets or scenarios, thus, reducing the likelihood of failure to generalize to new situations and unseen data. Furthermore, it is worthy to note that stochasticity can enable Uncertainty Quantification, leading to more generalized, unbiased and effective AI systems, as described below.

In general, Uncertainty Quantification is a process of estimating the reliability and confidence of the predictions of a Machine Learning model. It involves identifying and quantifying the sources of uncertainty in such a model, which can lead to better Decision-Making, improved model performance, but also increased trust in AI systems. In addition, Uncertainty Quantification contributes towards the *transparency* and *explainability* (Section 3.3.7) of such systems. Moreover, Uncertainty Quantification helps to identify and mitigate risks associated with AI systems, since they can involve *unintended consequences*, particularly in cases of safety-critical applications, such as self-driving cars that are aware of their own uncertainty, thus, either making more cautious decisions in ambiguous situations or handing over control to the human driver in order to reduce the risk of accidents or other critical failures to the car (as further explained in Section 3.3.7). This is actually one of the many reasons why **Bayesian Learning** is so useful, since it plays a crucial role in correctly estimating uncertainty in an AI system, as extensively explained in [1]. Therefore, Uncertainty Quantification is particularly important for tackling the problem of AI alignment (as defined in Section 3.3.7), but also ensuring trustworthiness of AI systems in general.

3.3.6 Sentience and emotional empathy for AI

In the general context of Trustworthy AI, there are a series of reasons why sentience and emotional empathy might be useful or even necessary for the development of AI systems and AI agents. Formally speaking, **sentience** refers to the *egocentric* ability of an observer (human, artificial or otherwise) to experience and perceive (positive, negative or neutral) situations *subjectively*, while **emotional empathy** is the *altercentric* ability to *emotionally* understand and feel how another being thinks and feels, as if they were experienced by the observer. Both require a certain level of external awareness (of the surroundings) and cognitive ability. However, the reader should not confuse the similar, but different, concepts of **consciousness** [1, 32, 36, 42] and sentience, since the latter refer more to the also *egocentric* ability of an observer to experience and perceive situations *intellectually*, thus requiring a certain degree of *self-awareness*. Then there is also the concept of **cognitive empathy**, which refers to the *altercentric* ability to *logically* understand how another being thinks and feels. Effectively, cognitive empathy is 'empathy by thought', rather than 'by feeling'. Definitions can vary, but these are the ones used throughout this whitepaper, with consciousness and cognitive empathy being outside the scope of this whitepaper, but not necessarily outside the scope of the proposed Ecosystem (Section 4.2). The authors proposed that a relationship between these concepts exist, as shown in Fig. 19, while the learning process (possibly in the brain) is necessarily **multimodal** in the sense that it is affected by *sensory experiences* involving multiple (internal and external) senses.

Some of the *reasons* that make these abilities important include the following:

(1) These abilities can help AI systems and AI agents to better understand and feel human emotions, subjective experiences and behaviour as indicated by social cues, visual cues, auditory cues, and haptic cues (cf. Multimodal Machine Learning in [1]).

(2) These abilities greatly enhance the ability of AI to interact and collaborate with humans in a more meaningful way (Section 3.3.9) by having the ability to express back emotions of its own, thus leading to more effective communication through more empathetic and human-like responses.

(3) These abilities can enable *emergent communication* among AI agents themselves (cf. Communication among agents in [1]), which can be interpretable by humans.

(4) These abilities can facilitate better AI steerability, explainability and alignment (Section 3.3.7).

(5) AI agents can better understand the consequences of their actions and make more informed and ethical decisions that take into account the wellbeing of humans, while considering and mitigating potential risks (Section 3.3.2).

(6) These abilities contribute towards making AI more customizable (Section 3.3.4) and hence, more user-friendly and trustworthy.

Regarding emotional empathy, Gallese and Sinigaglia [38] theorized that the major process of Social Cognition among humans is *embodied simulation* (c.f. Embodied and Grounded Cognition in [1]), in which individuals reuse the actions, emotions, and sensations "copied" from others through specialized neurons called **mirror neurons**. Neurological studies of mirror neurons in macaque monkeys strongly suggest that primates likely understand the intentions of others by internally *simulating* their behaviour (possibly in their brain) and observing the outcomes of those intentions. In other words, some neurons fired both when the observer (i.e.monkey) performed an action, such as reaching for a piece of food, and also when the observer watched the experimenter perform the same action, as if the observer was performing a similar *action* or experiencing a similar *emotion* or *sensation*. Such an embodied simulation can act as the additional source of information to guide or ground the system towards generating more sensible and aligned output (Section 3.3.1).

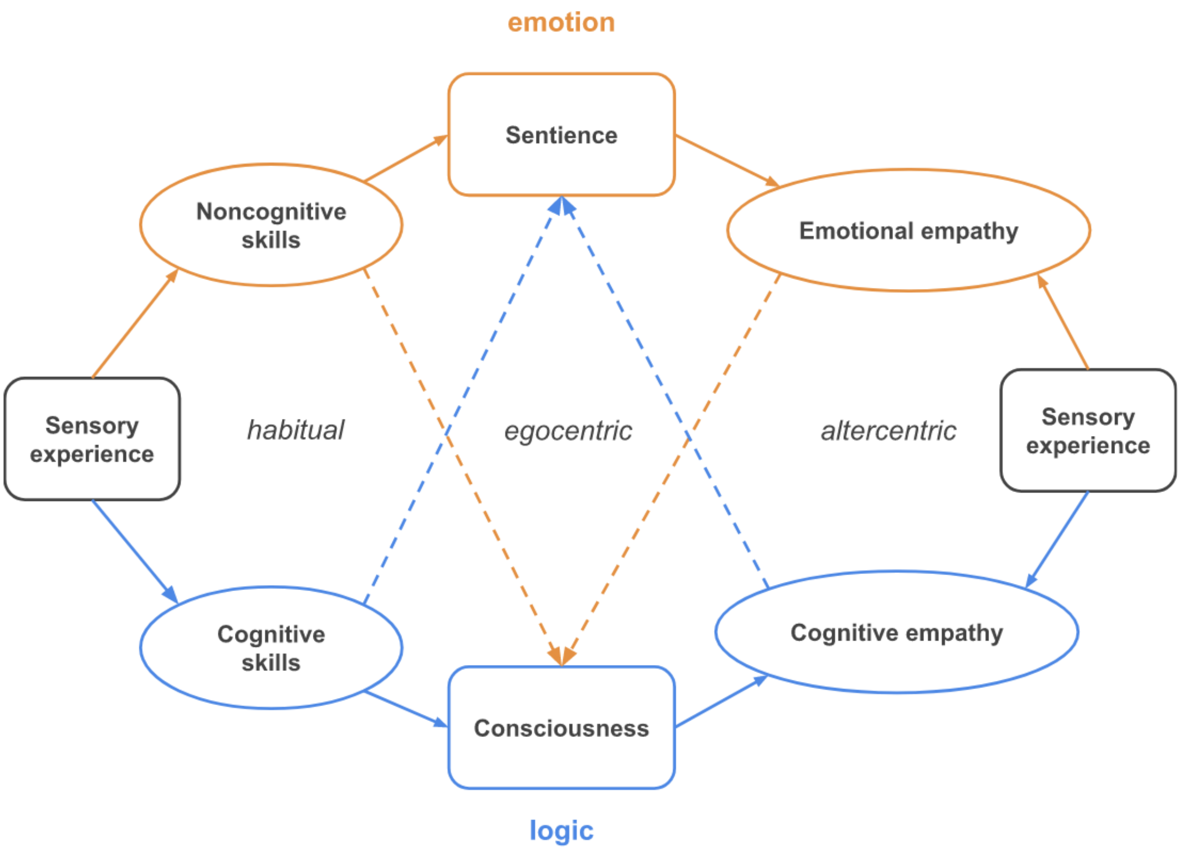


Figure 18 - Relationship of sentience, consciousness, emotional and cognitive empathy (with dotted lines representing grounding mechanisms or feedback signals).

Neuroimaging techniques, such as functional Magnetic Resonance Imaging (fMRI) and Positron Emission Tomography (PET), have also located *human* mirror neurons. Because the human Mirror Neuron System includes the automatic "copying" of actions, speech, face expressions, and emotions of the people they observe, it is likely that mirror neurons are an important component of social *group dynamics* when considering a *bottom-up* approach (cf. Collective Intelligence in [1]). Therefore, **Embodied Simulation Theory** [38] (which stems from the empirical support for the Mirror Neuron System) argues that humans reuse what they automatically "copy" in order to choose their next actions (c.f. *Visuo-Spatial Perspective Taking* in [1] and *Altercentric Cognition* in [43]). Generally speaking, this can be considered a form of *emotional empathy*, and hence, it can be used as a source of bio-inspiration for AI to similarly acquire its own version of emotional empathy, as proposed by Kourouklides [1].

As a matter of fact, AI can be used to learn on its own the *emergence* (Section 3.4.2) of such an emotional empathy, and with *minimal* human intervention during this emergence (Section 3.3.3), but while still having the qualities of being steerable and aligned (Section 3.3.7).

Taking it a step further, understanding the actions, intentions and emotions of another person is believed to *facilitate* efficient human communication. Therefore, empathy probably plays a crucial factor in human communication. Analogously, allowing machines to develop their own version of **emergent communication** (cf. Communication among agents in [1]), which will have to be constrained to the *physical limits* of each AI agent (e.g., chatbot, mobile robots, aerial robots) as well as the physical environment and the social environment around them (termed as *grounding mechanisms* [1]), might facilitate sentience and emotional empathy for AI or even be necessary for it, while sentience and emotional empathy might in turn enhance communication among (heterogenous) agents.

It is worth mentioning that even though the aforementioned theories have gained widespread support from the research community, they are not without their criticism. For instance, certain schools of thought (e.g., Materialism) completely reject that consciousness even exists [32]. This is why an ecosystem is necessary in order to fund and conduct experiments that will shed further light on such topics, while also disentangling gatekeepers (Section 2.3.3) from *bona fide*, genuine critics, whose critical evaluation and review is actually provided in good faith.

Also, it is important to note that there are ethical considerations (Section 3.3.2) and potential risks associated with creating sentient and empathetic AI. For instance, it may be difficult to ensure that AI systems are *not* implemented with (intentional or unintentional) security vulnerabilities that are hidden behind the stochastic nature of sentience and empathy.

Additionally, an AI system that is too empathetic may be susceptible to manipulation or exploitation by malicious actors. In other words, one would not like an AI system to become "emotionally unstable". This is another reason why an ecosystem is needed, i.e. in order to ensure that sentient and empathetic AI is implemented in an *open* and *transparent* manner for the mutual benefit of all humans of the planet, by addressing *auditability*, *reliability* and *robustness* (Section 3.3.8). However, before addressing such issues, it is necessary to firstly examine issues such as *observability*, *reachability* and *controllability*, which are concepts stemming from Control Theory and are explained in Section 3.3.7.

3.3.7 Steerable, aligned and explainable AI

These days, many people cannot even control their own kids, yet they are intriguingly interested in controlling AI systems. Jokes aside, this does not necessarily mean that the AI should not be able to be controlled (i.e. be *controllable*). However, further questions arise. Controlled by whom? Which specific humans or group of humans should "control" AI? The authors claim that all humans should be able to control AI, and not a small clique or an inner circle. The authors claim that similar to other open-source software, AI should also be available to the masses for the betterment of society and the advancement of the human species as a whole. This is why there is a need for one or more non-profit *foundations* to act as catalysts for the arrival of **steerable AI**.

The authors also claim that the road towards steerability is to allow AI to somehow learn or teach itself (Section 3.4.2) the abilities of sentience and emotional empathy (Section 3.3.6) inside an isolated sandbox that allows conducting *collaborative* and *controlled* trial-and-error experimentations and simulations.

The purpose of steerable AI is to allow AI systems to be controlled and directed by a human overseer in a relatively easy and user-friendly manner in order to perform specific tasks or achieve certain goals (defined by the human in collaboration with AI). Therefore, a desideratum of steerable AI is to allow human overseers to adjust and refine its behaviour in real-time. For instance, a human overseer might want to change the task of an aerial robot in agriculture in case an accident suddenly occurs, such as fire in the crops. Up to a very limited degree, semi-autonomous *control systems* already exist in manufacturing, but steerable AI systems refer to orders of magnitude higher autonomy, intelligence and creativity than traditional control systems, as explained below. As an illustration, one of the main gaffes of GPT 4 is that it seems to be less steerable, since the AI system appears to be more “stubborn” by taking human instructions less into account than previous versions (e.g., GPT 3.5), thus, seemingly having greater difficulty in generating the output that is desired by the human user interacting with it.

More generally, this means that a steerable AI system can adapt to changing circumstances, respond to new inputs, and optimize its performance based on spoken, written or visual feedback from its human overseer. Thus, it enables greater adaptability, flexibility and agility in AI systems, making them more useful in dynamic and complex environments. However, it also allows humans to provide input and guidance when the AI system encounters situations that it has not been specifically trained for, or when it encounters unexpected outcomes. This enables the human overseer to retain a level of “control” and oversight, ensuring that the AI system operates within acceptable limits and constraints, so that it does not make decisions or take actions that the human overseer has deemed ‘harmful’, ‘dangerous’ or otherwise ‘undesirable’. Overall, steerable AI represents a significant advancement in the implementation of AI systems that can *interact* and also work *collaboratively* with humans to achieve shared goals and accomplish common tasks (see also Section 3.3.9).

In order to achieve a meaningful level of steerable AI, it is crucial to first address the problem of **AI alignment**, which focuses on implementing AI systems that are aligned with the intended goals, preferences, or ethical principles of humans (i.e. a combination of soft and hard constraints) as defined by the human overseer. Overall, steerable AI is a key component of AI alignment, since it allows humans to exert control and influence over the behaviour of AI systems.

For example, in the scenario of a self-driving car, steerable AI could allow a human driver to take over control of the vehicle in situations where the AI is *uncertain* (Section 3.3.5) or encountering unexpected circumstances. This would ensure that the car remains *aligned* with the standards of safety and comfort of its passengers, even in environments that are complex, dynamic and “unpredictable” in the sense that the confidence of the predicted outcome is *below* a certain threshold or equivalently, uncertainty is *above* a certain threshold. Such thresholds can be defined either beforehand by the AI Architect or defined in real-time by the human overseer, depending on the design criteria of the AI system. As a result, embracing and quantifying uncertainty (as described in Section 3.3.5) is a necessary, but not sufficient condition in order to tackle the problem of AI alignment.

Therefore, steerable AI allows for continuous monitoring and evaluation of AI systems, which is absolutely crucial for ensuring alignment with human goals and preferences. By providing feedback and guidance to the AI system, the human overseer can identify and correct any misalignment or *unintended consequences* that may arise. Consequently, the authors claim that AI alignment is *not* a problem that can be naïvely tackled in isolation, but instead, as a part of a more holistic and general framework of Trustworthy AI, which necessarily includes humans as decision-makers.

Not to be confused with AI steerability, a highly undesirable side effect of Stochastic Parrots, such as GPT, is something called **'Prompting'** or 'Prompt Engineering'. Its meaning is that, due to the lack of AI alignment, the human user has to keep describing a concept many times and in different ways in order for a Conversational AI system to provide the answer that was aligned with the intended goal of the human. Therefore, Prompting violates the desiderata of AI alignment, since a well-behaving AI system should understand both the context and hence, the intent of each specific human user. This is one of the main shortcomings of most Conversational AI systems, as they are described in [1].

3.3.8 Auditability, reliability and robustness for AI

For AI systems to be trusted, humans will need to be able to assess the auditability, reliability and robustness in an open and transparent way. Not to be confused with explainability, a **software audit** [39] is defined as an independent evaluation of the level of adherence to relevant standards, guidelines, rules, regulations, specifications and procedures regarding software products and systems. Such an audit can be either internal or external, both of which should ideally be automated in order to ensure an audit outcome that is as objective as possible. In the context of Trustworthy AI, **reliability** refers to the ability of an AI system to perform its intended functions consistently, while producing results that the human user can trust and depend upon. This requires using Reliability Engineering techniques in order to identify and quantify the sources of failure in software products and systems, while also eliminating or mitigating the consequences of such failures, if possible. Meanwhile, **robustness** can be defined as the ability of a software product or system to maintain its functionality and performance under varying conditions, while withstanding and adapting to a range of different scenarios, but without compromising its overall performance or reliability. This requires using Robust Control Theory methods and sensitivity analyses.

3.3.9 Human-MARL collaboration and interaction

Despite being a common misconception, the goal of AI is to *not* make machines more like humans in the sense of replicating all the specialized abilities of human intelligence. Therefore, even though AI can sometimes be *inspired* by nature (including humans), this bio-inspiration is analogous to how aeroplanes are inspired by birds (but are definitely not identical to them). In the same way, AI can be inspired by the human brain, but it is *not* trying to explain how the human brain works (which is closer to the field of Computational Neuroscience). After all, the term 'intelligence' in AI is used *metaphorically*, and it does not reflect the complexity, agility and adaptability of human intelligence.

Instead, AI has multiple goals that stem from the ultimate purpose of AI (as described in Section 5.1). The main goal of AI is to create machines that can work alongside humans in a *collaborative* manner, complementing human strengths and weaknesses. In other words, the goal is to leverage the strengths of machines and humans in complementary ways. On the one hand, machines can perform certain tasks more efficiently, more cost-effectively and less prone to error than humans can, such as processing large amounts of data or performing repetitive tasks. Humans, on the other hand, are better at tasks that require higher levels of creativity, empathy, and social skills. By creating machines that can perform specific tasks better than humans can, humans can be freed from mundane or repetitive tasks and allow them to focus on tasks that require their unique human strengths.

For instance, imagine the scenario of a hurricane (or similar natural disaster) in a developing country, in which temporary infrastructure has to be built. An AI meta-system that consists of mobile robots can provide instructions on workers in order to construct temporary housing in case the specific humans have never been trained to do so. In addition, an AI meta-system that consists of aerial robots can perform inspections on the construction site and cooperate with mobile robots in order for the latter to provide feedback to the workers. This scenario can also be extended to other cases in construction, as well as manufacturing, logistics, agriculture (Fig. 20) and other domains. By creating machines that can collaborate with humans for the accomplishment of a common task, the whole process can be made more efficient and productive, while also being more trustworthy and enjoyable to humans.

Figure 20 - An example of HuMARL collaboration and interaction in agriculture using aerial robots.

Therefore, the main goal of AI is *not* to recreate a specific human, but create a *diversity of (artificial* and *heterogeneous) species*, so that by exploring the aforementioned purpose, the AI community will *potentially* be able to improve and enhance mobile robots, aerial robots (drones), swarm robots, etc., in harmony and peaceful coexistence with humans through **Human-MARL (HuMARL)** collaboration and interaction [1].

3.3.10 Irrational fears about AI

Apart from the potential risks and unintended consequences that humans, the environment around us and AI itself might have from the research and development of AI systems, a huge variety of *irrational fears* also arise, just like with every other piece of technology that emerged in the past of human history (Fig. 21). However, humans should be reminded that the greatest threat to them are **other humans**, not AI, not software, not hardware, not some form of artificial life. Just like any other piece of technology, AI is merely a tool, just like a knife. And realistically speaking, out of all the technologies, the most imminent and catastrophic threats are the weapons of mass destruction (nuclear, chemical and biological). In a prioritized list of such threats, AI should be towards the end of the list, if it is going to even be included at all (Fig. 22).

Figure 20 - Alarmists and doomsdayers from the 20th century.

There are individuals who self-identify as "Researchers", but engage in fraudulent “research” in order to secure funding. These individuals often misuse alarmism and cannot be trusted to genuinely assess risks, as they have a vested interest in *exaggerating* such risks in order to for more funding to be allocated to their field (e.g., Ethics, Physics) in contrast to other ones (e.g., technical and scientific side of AI). Additionally, politicians are fond of these exaggerated risk assessments because they provide an opportunity for imposing new and increased *taxes*, which can be used to combat unknown causes that they do not understand and are also outside their control. More generally, there is a significant amount of fearmongers with ulterior motives to instil irrational fears in one domain (e.g., the economy) to the public, typically because they have a platform due to their expertise or popularity in another domain (e.g., Machine Learning). Researchers, politicians and other fearmongers often manipulate emotions and fear levels for their own personal gain. In reality, the actual risks to humanity are much lower than what these various alarmists (Fig. 21) would lead us to believe, and risk assessments should be approached with common sense rather than “political correctness”.

Figure 22 - The irrational fears about AI.

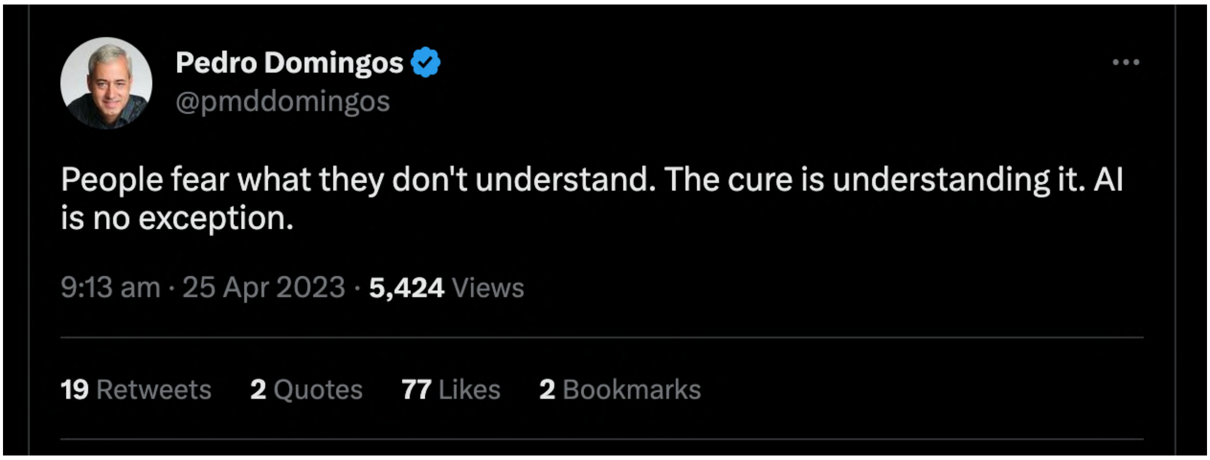


Figure 23 - The need for informing, educating and training the public about AI.

Therefore, it is important to address and minimize irrational fears by being as transparent as possible, while also informing, educating and training the public about the research and development of AI systems (Fig. 23). For instance, it would be great if various Education Technology companies (see also Section 4.10.1) emerged out of the proposed Ecosystem that would be addressing in various user-friendly and interactive ways all the irrational fears promoted by various fearmongers regarding AI, while:

(a) allowing engagement of the general public in an easily accessible online forum with no censorship, fair moderation (as defined in Section 3.2.2), and Fair Rewarding (as defined in Section 3.1.3) when it comes to all the independent content creators,

(b) being complemented with real-life experiences (Section 3.2.3), thus, further enhancing engagement outside the limited screen of a smartphone or a laptop, for example,

(c) engaging the public in order to create Optional Open Standards (Section 4.2.3) in order to increase trustworthiness of AI and hence, its mass adoption, and

(d) allowing laymen and non-specialists to generate “toy” AI systems as part of project-based learning using a very user-friendly way, such as natural language and no-code development platforms, for example.

3.4 Roadmap towards Artificial General Intelligence

Everything mentioned above about AI mainly refers to 'narrow' AI or 'weak AI' in contrast to **Artificial General Intelligence (AGI)** or 'strong' AI. Both terms are explained in [1], but briefly, the former refers to AI systems whose 'intelligence' is restricted to performing specific tasks and solving specific problems, while the latter refers to hypothetical AI systems whose 'intelligence' is comparable to that of humans in terms of its breadth and depth. As the reader might have guessed it already, the concept of AGI is highly *anthropocentric*, since it would be capable of understanding and learning any intellectual task that a human being can, and would also be able to reason, perceive, understand, and communicate in natural language.

AGI has not yet been achieved in practice, as it is still a purely theoretical concept, thus, remaining part of science fiction, for now. Realistically speaking, the humanity is several decades or even a century away from a fully fledged AGI, despite the alarmists (most of which never even attempted to build a single AI system in their life) claiming otherwise, since most of their claims stem from *irrational fears* about AI (Section 3.3.10). Symmetrically, there is a diametrically opposite ideological movement that absolutely adores (or even worships) the arrival of AGI and in general, the widespread use of AI (often referred to as 'transhumanism'). This whitepaper is *neither* concerned with promoting (or rejecting) the one versus the other *nor* does it even want to take a "balanced" or "middle-ground" standpoint among them. The authors remain completely **agnostic** on this political topic and focus on the *practical implementations* of AI instead, by assessing realistic risks versus realistic benefits. In other words, the proposed Ecosystem is envisioned to implement multiple censorship-free *forums* to discuss such topics, but without imposing *a priori* biases of an individual or a small clique.

Even though the humanity is several decades or even a century away from the point in time that AGI will be realized (call it ‘Singularity’ or whatever), AGI should be considered more of a “spectrum” or a continuum or in other words, a gradual process (measured by UBOC [1]) rather just as a single point in time. In other words, humanity is not going to wake up one morning and suddenly, *boom*, AGI will arrive. This works as an advantage for humanity in order to give it time to *prepare* for the arrival of AGI by establishing, for instance, a framework for Trustworthy AI (Section 3.3), but also to then perform *collaborative* and *controlled* trial-and-error experimentations similar to how nuclear research is conducted at CERN, as an illustration. Both examples give rise to the need for an ecosystem to address AGI for the reasons further explained in Section 3.4.3.

3.4.1 Automated Scientific Discovery revisited

In order to perform the aforementioned experimentations in a *trustworthy* (Section 3.3) and *controllable* (Section 2.2.7) environment, such an environment should necessarily facilitate Automated Scientific Discovery (as defined in Section 2.3.10). The authors revisit this concept in the context of AGI. An AGI system or agent should have the ability to perform scientific discovery. Also, scientific discovery can be fully automated via AGI. So, at a first glance, this might seem like an *antinomy* (or a *chicken-and-egg* problem). However, the level of automation (of scientific discovery) can also be considered as a spectrum. Therefore, it can be posed as a **dual problem** (similar to Expectation-Maximization as explained in [1]) and get closer to the solution one iteration at a time, instead of attempting to isolate and improve only one fold all at once.

3.4.2 Self-learning systems: AI generating AI

Automatic Scientific Discovery is not all you need to get closer towards AGI. Taking it a step further, AI will eventually have to learn how to teach itself and in order to achieve *autodidactic* AI, **self-learning systems** have to be produced. In other words, the epitome of automation, efficiency and creativity would be to have the possibility of **AI generating AI**. In the context of Generative AI, in contrast to discriminative models, there is a series of *generative models* (as both kinds of models are defined in [1]), such as Generative Adversarial Networks (GANs), Autoencoders, and Sequence-to-Sequence (Seq2Seq) models. A famous Seq2Seq model is the Transformer. The interested reader is referred to [1] for more details on these models. However, these are models in which AI is generating text, image, speech and other modalities, and not AI generating AI.

To the best of the authors' knowledge, the closed concept that has been proposed in the literature so far is that of the 'Autodidactic Universe' by Alexander et al. [41]. Nonetheless, their work was aimed at investigating a novel conception of the laws of Physics, in which laws *evolve*, but evolve by *learning*, and emerge from an autodidactic process, suggesting that the universe might have the capacity to learn laws (similar to 'learnable' rules proposed by GUT-AI in [1] or 'habits of nature' proposed in [34]).

So, is AI all you need then? The authors claim that no, AI is *not* all you need. However, the Ecosystem proposed in this whitepaper (Section 4.2) might be all you need, for now. The reasons why this is the case are described in Section 3.4.3.

3.4.3 The need for an ecosystem

By now, the need for an ecosystem in order to address all of the challenges mentioned in Section 3.1 should have been made apparent. The authors of this whitepaper make the claim that such an ecosystem neither exists nor was it ever proposed so far in a way that it is both holistic and practical to implement.

To sum up, the **reasons** why such an ecosystem is necessary are briefly the following:

(1) Create an isolated sandbox to conduct *collaborative* and *controlled* trial-and-error *experimentations* in order to study future AI systems, including AGI and self-learning systems (Section 3.4.2) at a large scale before they are released intro production in order to prove their trustworthiness (Section 3.3) so they can reach mass adoption.

(2) Enable Automated Scientific Discovery (Section 2.3.10) and Automated Data Science (Section 4.5) in order to increase the efficiency of the relevant part of the *cycle of innovation* (Fig. 14), so that the quality of life of all humans can be improved.

(3) Tackle Automated Scientific Discovery and the continuum of AGI as a dual problem, as described in Section 3.4.1.

(4) Enable Automated Product Discovery (Section 4.6) in order to: (a) make less hindered, more effective, and more efficient the knowledge transfer (e.g., know-how, novel methodologies, research findings) from the scientific discovery part of the cycle to the commercialization part without relying on single points of failures due to centralization (Section 3.1), (b) maximize commercialization of research findings so they can fuel the *ownership economy* in Web3 (Fig. 15), which would then allow the amount of funding allocated for further research to increase dramatically, and (c) increase the quality of products offered to consumers by enhancing competition.

(5) Create conditions, such as Fair Rewarding (Section 3.1.3), that would provide fertile ground for the ownership economy in Web3 (Fig. 15), while also reversing the deleterious effects of limiting creative expression and stifling innovation, which act as *inhibiting factors* towards the cycle of innovation (Fig. 14).

(6) Align interests under one *fully decentralized* roof in order to address both: (a) the lack of funding (Section 2.3.8) that is necessary for generating scientific discoveries and technological breakthroughs, and (b) minimize the inefficiencies when it comes to excessive spending of initializing commercialization of research findings (Fig. 14).

(7) Make the transition to an ownership economy (Fig. 15) smoother for everyone involved by allowing the creation of AI solutions that would: (a) prepare prosumers for such an economy through education, training and other tools, which could be built on top of the Ecosystem, while genuinely assessing risks and challenges related to AI and debunking irrational fears about it (Section 3.3.10), and (b) help organizations address challenges for transitioning to such an economy in a cost-effective, affordable and effortless manner that would not compromise their bottom line, but actually increase their operational efficiency.

Knowing the reasons why the proposed Ecosystem is needed is indeed essential. However, “the devil is in the details” according to the old idiom. Therefore, the blueprints, foundations and guiding principles influencing the eventual implementation of the Ecosystem are equally crucial, which raises the question of what they should be in order to ensure its success. Chapter 4 deals with this question in a brief manner, while Chapter 5 explains how the Ecosystem fits within the broader GUT-AI Initiative.

4 Solution

The solution presented in this Chapter is expressed in such a way that it *neither* restricts creativity *nor* does it create a bias towards any specific implementation (i.e. it does not state *'how'* the solution will be implemented) in order to allow maximum *creativity*.

4.1 The Foundation

The GUT-AI Foundation has a *supportive* role, while acting as a catalyst in order to accelerate the GUT-AI Initiative (Chapter 5), but *without* interfering with the **decentralized** nature of the whole initiative. In other words, the GUT-AI Foundation is merely a pure *subset* of the initiative. The Foundation is envisioned to function as a non-profit DAO and be incorporated as a legal structure in one or more DAO-friendly jurisdictions [46 - 48].

The Foundation is envisioned to be a *fully self-governing* DAO right from the beginning, and the Network's technical specifications and governance methodology will be designed with the provision to support this. All governance decisions will be made democratically by the **governance token** holders or their elected representatives (as described in Section 4.1.1). It should be highlighted that the Foundation may also offer potentially SD solutions on top of either FD or other SD solutions, but as already mentioned above, *without* interfering with the decentralized nature of the whole initiative and *without* restricting the right to any potentially competing protocols to also offer such solutions as well. The reason for offering *both* FD solutions side-by-side with SD solutions is to get the best of both worlds, since each has its own advantages and disadvantages (as described in Section 3.2.2). The way that the GUT-AI Initiative will blend these two different types of solutions is described in Section 4.2.1 below, with the Foundation participating in the proposed Marketplace (Section 4.4).

4.1.1 Foundation governance

In general, the governance of any DAO foundation is critical for its success and long-term sustainability. A well-designed governance framework ensures that the DAO operates transparently, efficiently, and effectively, while preventing failure to accomplish its goals or deviate from its stated mission and vision. Not to be confused with Network governance (Section 4.2.8), this Section outlines the key components of the governance framework for the GUT-AI Foundation, including Decision-Making processes and leadership roles.

The **Decision-Making** processes of a DAO foundation should be transparent and participatory, but also effective by ensuring that decisions are made in a swift, streamlined and simplified manner. This creates a *trade-off* between participation and effectiveness. As the proverbial expression says, “too many cooks spoil the broth”, which conveys the notion that, if too many people participate in a task, especially in a leading role, then the task will not be done very well, or even worse, will not be done at all within the designated timelines.

Regarding transparency and participation, all DAO members should have the opportunity to participate in the Decision-Making process, while the process should be documented and communicated clearly to all such members. This includes outlining the Decision-Making criteria, the types of decisions that can be made, and the roles and responsibilities of those involved in the process. The process should also include a clear timeline for making decisions, as well as a mechanism for resolving disputes. Hence, all of the aforementioned practices take a considerable amount of time, thus hindering the requirement for the decisions to be made in a quick and timely manner, especially for time-sensitive decisions. This gives rise to the need to practically address this issue. The authors claim that **elected committees** are a good candidate solution for addressing this issue, as described below.

Irrespective of the technology used, DAO is a radically new type of organization. Its emergence promised the elimination of positions such as Chief Executive Officer (CEO), Chief Technology Officer (CTO) and Chief Financial Officer (CFO), while also offering the potential not just for better output, but for a better work *experience* for everyone. However, the **leadership roles** within a DAO foundation are critical for ensuring that such a foundation operates efficiently and effectively. Therefore, *elected committees* (similar to 'collectives' [69] in Polkadot) could prove beneficial when applied to leadership roles.

Traditional foundation	GUT-AI Foundation
Board of Directors	Board Committee
CEO	Executive TMC
CTO	Technology TMC
CFO	Finance TMC
-	Arbitration TMC

Table 2 - Comparison of the leadership roles in a traditional foundation and the GUT-AI Foundation.

More specifically, inspired by the collegial system of Swiss Federal Presidency, the authors of this whitepaper propose the introduction of **Three-Member Committees (TMC)** to assume *all* the respective *senior management* roles (e.g., Executive, Technology, Finance as listed in Table 2), and whose term will be 18 months, with the possibility of reelection. Each such Committee will have a designated **Chair** who will have the right to a *casting vote* and will *rotate* on a 6-month basis. Legal or regulatory reasons might require the need for other committees too. There is also a possibility for forming *ad hoc* committees to address project-specific issue that might arise from time to time.

As a result, all three members would have the ability to eventually become Chairs of the committee in hand during their term, unless or until the Board (a multi-sig committee) decides to remove them. At the first level, dispute resolution will be dealt with by the Arbitration Committee, which can then be appealed to the external arbitration organization designated by the by-laws or the rules of the Foundation. The TMC members will all be elected by the Board (and *not* the token holders directly) either using cumulative voting or any other (multiple-winner) voting scheme determined by the Board in its sole discretion.

The Board will be directly elected by all governance token holders and will be the *only* governing body in the Foundation with the right to appoint and remove TMC members. However, the *highest* governing body of the Foundation will be the General Meeting of the token holders, since it will determine the size of the Board (which must be within the range specified in the by-laws), and will also elect all the Board members using cumulative voting. The General Meeting can take place *remotely* (in accordance with the by-laws) and can override any decision by any other body. Overall, the proposed **governance system** for the Foundation is a hybrid one, combining *direct democracy* with *representative democracy*, to achieve the best of both worlds and address the **participation-effectiveness trade-off**.

Each governing body will have rights, duties and responsibilities as stated in the by-laws. In addition, the by-laws and all the rules (which will be subordinate to the by-laws, and can be made by any governing body, including the General Meeting, Board and the TMCs) will stipulate what *type of decisions* (Table 3) can be made by each governing body or individual. Any member of any governing body may **delegate** their voting power to a representative (with each such delegation expiring after a predetermined date), and such representative may be either (a) another member of the same body or (b) external to the body or the Foundation. The specific voting mechanisms to be used are described in Section 4.1.4.

Type of decision	Decision-Making process
Type 1	• General Meeting *simple majority*
Type 2	• General Meeting *supermajority*
Type 3	• Board *simple majority*
Type 4	• Board *supermajority*
Type 5	• Executive Committee *majority*
Type 6	• Automatic when number of Board members falls below the minimum threshold

Table 3 - Examples of Decision-Making processes according to the type of decision.

4.1.2 Technopolitical governance vs. enforcement

At this point, it is important to reiterate that the GUT-AI Foundation has a *supportive* role, while acting as a catalyst (or facilitator) in order to accelerate the GUT-AI Initiative, but *without* interfering with the decentralized nature of the whole initiative. Practically, this translates to a different Foundation governance (described in Section 4.1.1) and a different Network governance (described in Section 4.2.8). In other words, both the technopolitical governance and the enforcement mechanisms (as defined below) of the Foundation are *disentangled* from and *vary* significantly to that of the Network itself. Even though this is not always the case in general, this is reasonable in the case of this whitepaper due to their respective *raison d'etre* (i.e. the most important reason for them existing in the way that they do), since each of the two entities serves a *distinct* purpose. The aim of this Section is to make this clearer using specific examples, and also, further elaborate on the details in order to ensure that it is better understood by the reader why this is the case.

In order to clarify, **technopolitical governance** is defined in this whitepaper as the use of technology (as a tool) to vote on or manage various aspects of governance systems and Decision-Making processes. For instance, it involves the use of technology (as a means to an end, but *not* the end itself) in order to *vote* on conflicting topics, *elect* representatives, and *decide* how rules, regulations and procedures will be proposed, formulated, implemented, established, administered, enforced, evaluated and terminated. Meanwhile, **enforcement** refers to those mechanisms and procedures put in place in order to ensure *compliance* with the rules, regulations and procedures of the governing system (as they were defined by technopolitical governance). Smooth functioning of a *governing system* (e.g., Foundation or Network) requires *iterative* improvement of both technological governance and enforcement at specific points in time as deemed necessary (Fig. 24). Additionally, in case of disputes, there needs to be a **Conflict Resolution mechanism** (e.g., internal or external arbitration) to facilitate an impartial ending of the conflict. In general, there is a series of vastly different ways that these three governance mechanisms can be implemented.

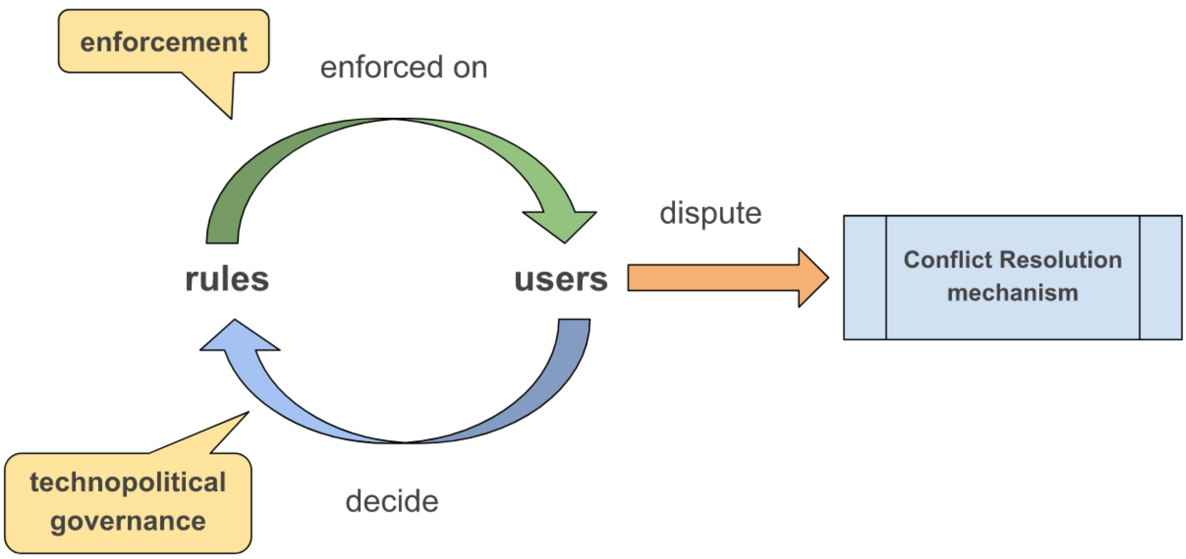


Figure 24 - Iterative improvement of technopolitical governance and enforcement in a governing system. The term 'rules' is used in the broad sense to also include 'rules about rules' (i.e. meta-rules).

In the specific cases of the proposed Foundation and the proposed Network, all three governance mechanisms *differ* significantly in terms of implementation. Their main differences are briefly described in Table 4. It is worth mentioning that the authors of this whitepaper do *not* claim in any way that these are the optimal or "best" ways to implement any DAO Foundation or any blockchain network, but instead, that they are a good starting solution for the specific case of the GUT-AI Initiative, which has its own specific peculiarities, functional and non-functional requirements.

GUT-AI Foundation	GUT-AI Network
More **on-chain** and **formal** technopolitical governance with **more** transparency.	More **off-chain** and **informal** technopolitical governance with **less** transparency.
More **technology-driven** enforcement and algorithmic governance.	More **human-driven** enforcement and non-automated governance.
Disentangled from any type of blockchain network.	Disentangled from any type of DAO foundation.
More organized, structured and controlled by nature.	More anarchic, unstructured and unconstrained by nature.
More populist and egalitarian.	More technocratic and non-egalitarian.
Modifiable, reversible and appealable rules (in accordance with other rules).	Non-modifiable, non-reversible and absolute character of the rules.
Access to internal and external arbitration mechanisms.	-
Semi-decentralized and enforceable.	Fully decentralized, but unenforceable.

Table 4 - Main differences between the governance mechanisms of the proposed Foundation and those of the proposed Network.

As a result, in terms of **governance mechanisms**, the GUT-AI Network is closer to the admittedly successful Bitcoin [57] and Ethereum [58] networks, whereas the GUT-AI Foundation is closer to the Dfinity's DAO [62], albeit with a *different* purpose. In order to further elaborate, the governance in the Bitcoin and Ethereum cases is *technocratic* and *unstructured*, since it is carried on solely by core Developers (not the users or validators), and *out of the chain* itself by proposing new rules in informal channels (e.g., Discord, Twitter), while adoption of such new rules is entirely *optional*, leading to shadow hierarchies. The interested reader is referred to [59] for more details. One would actually expect that, after more than 14 years from the date that Bitcoin was created, a social media dPlat dedicated to facilitating such informal communication would have been created and reached adoption by the Web3 community, but apparently, up to the time of writing, no such dPlat did.

As an illustration, if there is a serious disagreement within the community about the new rules of such a network (i.e. a large enough percentage of the community decides that they want to continue using the old rules), then the chain will inevitably split, eventually resulting in two separate cryptocurrencies, i.e. the "original" one and an entirely new one (often called "split coin"), which is actually incompatible with the original one. This type of resolution mechanism is known as a **"hard fork"** and happens naturally and automatically (due to the anarchic nature of such a network), with some famous hard forks leading to the creation of Ethereum Classic and Bitcoin Cash in 2016 and 2017 respectively. Overall, the phenomenon of informal, **off-chain** technopolitical governance and **human-driven**, non-automated enforcement [67 - 68] is *not* inherently problematic, but practically, it sometimes leads to some serious challenges, such as disproportionate *power concentration* around: (a) core Developers who are established, and (b) large validators [60]. However, despite their limitations, networks such as Bitcoin and Ethereum have evidently managed to withstand the passage of time, while still actively evolving and amassing adoption.

It should be noted that the terms 'miners' and 'validators' are used interchangeably in this whitepaper due to reasons mentioned in [61]. The same applies to 'mining' and 'validating', and so on. The authors also recognize that there are important differences between Ethereum and Bitcoin (e.g., different consensus mechanism), but those differences remain outside the scope of the governance mechanisms, since the bottom line is that, for better or worse, neither validators nor non-validator users get a saying (or a vote) in the technopolitical governance, but solely established core Developers do, since they are the only ones who practically decide on changes in the code. Even a new Developer has very little incentive to evolve the protocol, thus, interests and incentives are misaligned (Section 2.2.3). In other words, if the **"Code is Law"** principle applies, but only a small clique decides the changes to the code, while the new code is optionally adopted, then the principle is *one-sided* and only applies to the application of the new "law" to the user, who is treated like a second-class citizen, which was one of the initial problems of centralization, since "Code is Law" does not apply to the side of "law-making" or "meta-enforcement".

Meanwhile, users always have the option to "revolt" by opting out from the respective blockchain network (i.e. selling coins massively or boycotting its use), and joining another blockchain network or going back to fiat money. This is where **Game Theory** comes into play. Even though such networks are unconstrained by nature, there are some *soft constraints* to this resulting technocratic oligarchy in order *not* to "kill the goose that lays the golden eggs". In this way, *social pressure* can be exerted to influence the behaviour of core Developers without using any form of formal authority or power. Arguably, Game Theory is also supposed to protect against a so-called 51% attack (or majority attack), since the Nash equilibrium theoretically prevents such a scenario from happening. Therefore, the argument is that the new system (i.e. Bitcoin-like network) is an important improvement compared to the current system (i.e. fully centralized), albeit imperfect and suboptimal.

Regarding the GUT-AI Foundation, an opposite approach is proposed, in which the "Code is Law" principle applies to and is enforced upon each and every DAO member equally, without creating any second-class citizens. Getting inspired by some of the governance mechanisms of the Dfinity's DAO [62 - 65], this is achieved by: (a) having a more **on-chain** and formal technopolitical governance, (b) having a more **technology-driven** enforcement and algorithmic governance [67 - 68], and (c) recognizing that Decision-Making does *not* take place just by voting, but also through engagement, social interaction and **discussion** among DAO members, while also having rules, regulations and procedures that enforce this. Therefore, in contrast to the governance of Bitcoin-like networks, the Foundation strongly *inhibits* the creation of a technocratic oligarchy, while *minimizing* their power in case such an oligarchy starts to form, since the Foundation is relatively more organized, structured and controlled. Therefore, the governance is made more populist and egalitarian by design. There would also be a possibility of amending the ledger under certain assumptions, prerequisites and rules, with access to both internal and external arbitration mechanisms.

However, this comes at the expense that the Foundation itself might not be fully decentralized. Nonetheless, a high degree of semi-decentralization can and will be enforced in order to avoid the risk of being fully technocratic or fully centralized, thus creating a programmatically *imposed* lower bound of semi-decentralization (in contrast to Bitcoin-like networks). More crucially, it begs the question of what the blockchain is worth when it may be so amenable to change, thus making its governance potentially unsuitable for a network to issue a cryptocurrency (i.e. a utility token) similar to Bitcoin or Ethereum tokens, since such a cryptocurrency might end up being worthless. However, this is where the whole point lies. Such a token will *neither* be a utility token, *nor* will it be entangled with any utility token. Conversely, it will be **disentangled** from any type of blockchain network, which is a striking difference with the Dfinity's DAO that is entangled with the Internet Computer blockchain.

The authors of this whitepaper claim that this is where the **sweet spot** lies, as shown in Table 4. From a *game-theoretic* perspective, the proposed Foundation and the proposed Network do *not* form a fully cooperative game, as opposed to Dfinity's DAO and the Internet Computer blockchain, which they do. Instead, it is proposed for the two aforementioned systems to form together a **semi-cooperative game**, so that the game can combine both cooperative and non-cooperative elements, since a semi-cooperative treatment is a more pragmatic representation of (digital and physical) governance systems. **Non-cooperative elements** may arise from *suspicions* from one system (i.e. player) that the other one might opt to abandon their cooperation due to "selfish interests" and *vice versa*.

For instance, the DAO members might use their *collective bargaining power* to massively "revolt" against the Network by opting out from the respective blockchain network (i.e. selling coins massively or boycotting its use), and joining another blockchain network or going back to fiat money, thus making the Network suspicious about the Foundation. Symmetrically, the Foundation will be suspicious about the intentions of the Network in case the former decides to move to a new blockchain network, thus avoiding scenarios — or at least minimizing their chance of occurring — such as the one of ApeCoin owners voting marginally against leaving the Ethereum network, which can arguably prove devastating to ApeCoin in the near future [66], but clearly demonstrates a misalignment of interests (Section 2.2.3).

In contrast to the above, in the case of Dfinity's DAO, which is termed Network Nervous System (NNS) [62 - 65], the Internet Computer is a DAO-controlled network. In particular, all decisions taken by the NNS directly control the Internet Computer, and are made by DAO members whose voting power is determined by how much of the Internet Computer's native governance token they have staked in the NNS. Hence, the NNS and the Internet Computer form together a fully cooperative game.

Any further game-theoretic treatment is outside the scope of this whitepaper. Instead, the remaining Sections 4.1.4 to 4.1.8 deals with how to practically implement such a semi-cooperative game in the context of the proposed Foundation, while Section 4.2 deals with it in the context of the proposed Network. Meanwhile, Section 4.1.3 mentions some general issues that arise in governance systems, since such systems do *no* exclusively include formal rules, but additionally include unwritten rules, social expectations, norms and culture. Moreover, they include participation, engagement and social interaction in order to ensure inclusivity and legitimacy, while also achieving a *trade-off* **equilibrium** between: (a) a state of relative stability and social cohesion, and (b) fostering innovation and evolution.

4.1.3 Governance systems: general issues

A more general problem that arises from the *digital* governance systems, such as the ones of foundations and blockchain networks, is the lack of linking digital identities (and behavioural patterns) to the physical ones, so that one user to be able to assess the intentions of another, in order for the former user to achieve the goal of knowing *what* to expect from the latter by **predicting** his next move in the digital world with an accuracy close to the one in the physical world. Clearly, this defeats the whole purpose of blockchain networks existing from the first place, since it is antithetical to the desiderata of Web3 (as described in Section 3.1.1). Nonetheless, a possible solution for foundations is mentioned in Section 4.1.4. Regarding blockchain networks, a solution is to make sure that enough *engagements, social interactions* and *discussions* take place in the digital world as part of the Decision-Making process (as mentioned in Section 4.1.2), since it also includes norms, culture and other **social dynamics**, not merely code, voting and formal rules. In this way, the aforementioned goal can be achieved *without* any linking to the physical world or otherwise compromising his pseudonymity and privacy.

All in all, whatever the governance system, — DAO foundation or blockchain network, digital or physical, modern or ancient — it all boils down to the concept of **'trustworthiness'** or 'reliability' (*not* to be confused with 'trustfulness' or 'faith' in the system). In other words, a governance system can function *well* only when, and as long as, it can practically, effectively and decisively produce the *desired outcome* in a relatively predictable manner the vast majority of the time, no matter the underlying (digital or physical) infrastructure. Therefore, this whitepaper is mostly concerned with *practical applicability* rather than philosophically proposing theoretical frameworks or exhaustively considering all speculative scenarios and edge cases (that might never even come to fruition or occur in real life).

This is also the reason why the authors have deliberately opted for *not* aligning with any specific ideology, such as the ones proposed by Carl Menger, Ludwig von Mises, Friedrich Hayek, Milton Friedman, Murray Rothbard, Hans-Hermann Hoppe and so many others. Instead, the authors claim the following:

(a) There is *no* one-size-fits-all solution for all the problems in all the industries in all the countries of the world. Conversely, for each specific practical problem, a *different* treatment is required and a distinct set of solutions might be applicable. However, each and every ideological contribution is useful and worth being studied, since they all act as a **theoretical toolbox** in order to treat the problem.

(b) The future of Economics (and most of the other so-called "Social Sciences") needs to be tackled from a point of view that fully embraces the Scientific Method, while rejecting *a priori* biases of the so-called "Social Scientists". Conversely, an *isolated* sandbox (such as the one proposed in this whitepaper) that allows conducting *collaborative* and *controlled* trial-and-error experimentations and simulations can act as a **practical toolbox** in order to treat each specific problem, since experiments could be simulated on the proposed Ecosystem (Section 4.2).

(c) The solutions to the various problems that our economies (and societies) face today, but also to all the ones that they will face in the future, should not come out only from some inner circles of people (e.g., academics, unelected bureaucrats and self-proclaimed experts) who sit behind their dusted desks in some brick-and-mortar institutions acting as *gatekeepers* (as described in Section 2.3.3). However, each and every person on the planet should have an opportunity to contribute to and be part of the solution to such problems by using his fresh ideas, knowledge, know-how, skills, experience and expertise via the proposed Ecosystem, thus overcoming traditional barriers that currently exist, while also ensuring Fair Rewarding (Section 3.1.3) for each such contribution. The proposed Ecosystem might not fully achieve this, but: (i) it is a step in the right direction, and (ii) it is envisaged that the solutions that will fully achieve this will come out of this Ecosystem or at least, get closer to it.

4.1.4 Decaying Voting

It should be emphasized that **voting mechanisms** are a critical component of the governance framework for a DAO foundation, since they directly relate to governance and there is a series of challenges associated with each such mechanism that have also been observed in practice when it comes to DAOs [55], such as bad actors, malicious groups and buying elections [56]. A major risk is that anyone can sell their private keys to transfer their votes maliciously or without the knowledge of the DAO (since votes are linked to wallets). Projects like Gitcoin DAO alleviate this problem by using stringent identification systems that make it difficult to transfer voting rights maliciously or opaquely. For example, users have to register a profile with decentralized identity services, such as Proof of Humanity, BrightID, Ethereum Naming Service (ENS) and Idena. Once this risk is eliminated or at least mitigated, any voting mechanisms used will (hopefully) have the intended outcome.

The authors of this whitepaper propose the use of **Decaying Voting**, which is a type of *weighted voting*, i.e. each voter has a variable voting power (i.e. weighted vote) as determined by a function. Making it closer to a *reputation-based* DAO, the function in the case of Decaying Voting is determined by the summation of 'activity points', which: (a) can be *earned* in various ways, all of which are rule-based, unambiguous and predefined in order to minimize the impact of both subjectivity and the Tyranny of the Majority, and (b) *decay* over time and eventually expire in order to minimize unfairness and inherent bias against new, but active members. Additionally, a token-based *quorum* will be set, i.e. for a proposal to pass through the General Meeting, a certain minimum percentage of votes (*not* DAO members) must participate in the voting process.

As hinted in Section 4.1.2, the Foundation will also include elements of **delegated voting** (also called *proxy voting*), while also allowing a form of **"meta-delegation"** (often termed as 'Liquid Democracy') in order to tackle the problem of low participation rates generally found in governance systems of foundations and DAO-controllerd networks. There is a common misconception that this type of voting intrinsically leads to a "concentration" of voting power and hence, increasing the risk of centralization. However, nothing is far from the truth since in reality, it creates an **accumulation** of voting power (and *not* necessarily a "concentration"), which actually does *not* cause centralization any more than a person in any physical (or digital) governance system seeking expert opinion on a proposal and voting based on that opinion. As a matter of fact, a countermeasure is the ability for a token holder to change his designated representative (or proxy) or even completely remove his delegation at any point in time, while also forcing each such delegation to expire after a predetermined date. Hence, the risks of centralization arising from voting power in such a delegated voting system are significantly reduced.

As a result, the aforementioned accumulation of voting power, but also the overall voting system proposed for the Foundation, is *not* inherently contradictory to the goals of voting power decentralization. In fact, there is a further misconception that needs to be debunked regarding the same matter. The (mistaken) belief is that every type of weighted or reputation-based voting caused a form of "oligarchy" (albeit not technocratic). In reality, it rewards exceptionalism by essentially creating a **meritocracy**, in which equal opportunities are promoted (but *not* equal outcome), in contrast to a being plagued by *mediocracy*, in which mediocrity is rewarded. More specifically, well-functioning DAO members get rewarded (Section 3.1.3) by their peers by: (a) being designated as representatives, (b) being elected in committees (e.g., Board or TMCs), and (c) automatically receiving activity points for positively contributing to the DAO in certain rule-based, unambiguous and predefined ways. Conversely, non-well-functioning or apathetic DAO members receive *negative* rewards by: (a) being replaced or removed as representatives of their peers, (b) being recalled from committees, (c) no longer receiving activity points, while having their old ones decayed and eventually expired in an automatic and predefined way.

Therefore, the proposed voting system promotes egalitarianism and populism (Table 4), while also levelling the playing field and removing inequalities, such as between: (a) established vs. new core Developers, or (b) genesis and inactive vs. new and active DAO members. Additionally, the voting process will include *on-chain* discussions (as described in Section 4.1.3), which create and reinforce empathy among DAO members, thus leading to better and improved Decision-Making, while also ensuring social cohesion and a relative stability in the DAO community, but without harming innovation and evolution. Undoubtedly, there are certain underlying *assumptions* (e.g., transparency, prudent rules for earning activity points, sensible expiration date, user-friendliness of the voting tools) for such a voting system to be workable in practice.

As a concluding remark for this Section, the technical implementation of the voting mechanisms, but also that of the DAO in general, is deliberately left out because this whitepaper assumes a general audience and hence, it is *not* intended as a deep technical document. Also, it allows plenty of room for creativity (Section 3.3.3) when it comes to implementation.

4.1.5 Eligible votes

The DAO members who will be eligible to vote will be determined by specific predefined rules in order to avoid certain undesirable situations and unintended consequences. For instance, for the whole voting system to make sense, unless and until a DAO member registers a profile with the designated decentralized identity services (such as the ones mentioned in Section 4.1.4), he will *not* be able to earn 'activity points', but he will have a single vote, instead. This is in order to minimize the chance of Sybil attacks.

In addition, the authors recommend imposing a probation period (of 30 days, for example) for every new DAO member, during which he cannot vote on any matter. However, he would be expected to participate in on-chain engagements, social interactions and discussions (as described in Section 4.1.3) in order for his peers to assess his capability and suitability to be a well-functioning member of the Foundation. This countermeasure minimizes the risk of token holder "activists" targeting the Foundation for their own "selfish" reasons or hidden agenda, which might cause a devastating effect on it.

In general, further countermeasures can be proposed, but all such measures should be juxtaposed against the risk of jeopardizing the reputation and the trustworthiness of the Foundation perceived by all its members, as explained in Section 4.1.3.

4.1.6 Self-correcting rules and by-laws

A goal of the Foundation is to introduce rules and by-laws that are **self-correcting** in the long term. This can be implemented as a *five-step procedure*:

- **STEP 1:** Define the metrics that will be used for evaluation
- **STEP 2:** Gather the analytics, data and meta-data based on the metrics
- **STEP 3:** Use AI to provide feedback on the following:
    (a) What rules should change and how
    (b) What metrics should change and how
- **STEP 5:** Have on-chain social interactions in order to discuss the changes
- **STEP 4:** Vote in order to decide which changes to implement

The overall aim of the Foundation will be to run as *efficiently* and as *automated* as possible by maximizing the use of available AI, software and hardware technologies in order to achieve maximum scalability and adoption, while ensuring *transparency*. Therefore, AI is *not* introduced in governance for the sake of AI itself, but instead, it is introduced for the sake of and only when applicable to facilitating *self-correction* of the governance system.

In other words, the suggestions proposed in this whitepaper are *not* how to make a governance-system more machine-centred, but how to make it eventually make it both more **human-centred** and more **user-centred** through *control mechanisms* (as explaining in Section 4.1.7) facilitated by the machines acting as objective third parties, instead of one human being forced to blindly put all his trust (as defined in Section 4.1.3) in an inner circle of other humans, who might not always have his best interests at heart, thus potentially leading to economic elite domination. Similarly, this should *not* be misinterpreted as proposing a form of machine technocracy, either. The point is to strike a **balance**.

4.1.7 Oscillatory decentralization

While Polygon [82] introduced the notion of 'gradual decentralization' [70] regarding their approach towards (network) governance, this whitepaper introduces the notion of **'oscillatory decentralization'** (Fig. 25) in order to address the *participation-effectiveness trade-off* (as defined in Section 4.1.1) in the context of any sociopolitical system. Both notions are compatible with the postulate found in this whitepaper that decentralization exists in a *continuum* (Section 3.2.2) and also compatible with GUT-AI theory [1].

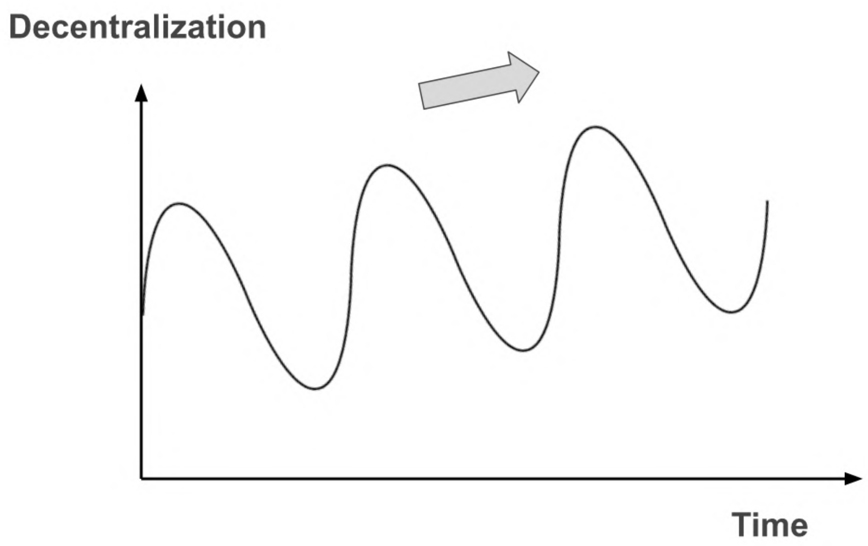


Figure 25 - Graphical representation of oscillatory decentralization.

Using the same example as in [70], in occasions of an *existential threat* to the state, the Roman Republic would transfer all power and authority to a single person, thus entrusting a more centralized entity with the powers to overcome such a threat, while expecting that authority to be returned after the threat was over. Hence, the belief was that a single office, unaffected by the inefficiencies caused by bureaucratic proceedings and political dynamics, was considered the most suitable for addressing a situation in which decisive, swift and simplified Decision-Making was a matter of **survival** for the Republic.

For instance, Lucius Quinctius Cincinnatus was appointed as ‘dictator’ during a time of crisis. Despite having absolute authority, he voluntarily relinquished power once he resolved the crisis by defeating the enemies of Rome, ending the war and *stabilizing* the situation, thus leading to a **gradual decentralization** after the crisis had passed. This concept of ‘benevolent dictator’ indicates that a capable, charismatic and decisive leader can effectively guide a social system towards progress and stability, under the strong **assumption** that certain *dynamics* (e.g., compassionate, virtuous and selfless character) will inhibit the leader from promoting “selfish interests” at the expense of the social system. Otherwise, when this assumption is violated, the concentration of power in the hands of a single individual can lead to abuses and the suppression of dissent, such as the case of Julius Caesar who abused the office of the dictator by famously declaring himself “dictator for life”.

In the path of history, the governance system of the Eastern Roman Empire (i.e. what is called ‘Byzantine Empire’ today) tried to adopt such governance mechanisms with a high degree of success. For instance, even though the Byzantine emperor served as the ultimate legislator and judge, the legal system was relatively *impartial*, while granting certain rights and protections to its subjects, partially due to the **values and norms** inherited by previous Roman leaders, such as Lucius Cincinnatus, but also due to **checks and balances**.

Also, even though there was a supposedly “centralized” administrative structure in place, Byzantium divided its territories into administrative units, known as ‘themes’ (i.e. provinces), while (partially) **decentralizing** power to such themes, since the centralized administration in the capital city, Constantinopolis, *delegated* considerable authority to them. Also, the Empire was meticulously adept at diplomacy, and was skilled in *adaptability*, due to the flexibility of its *governance mechanisms* (e.g., provincial autonomy, administrative reforms, efficient bureaucratic apparatus, culture of competent leadership, technological innovations).

Despite being falsely accused of “caesaropapism”, countless institutions, such as the ‘Pandidacterium’ (i.e. the Imperial University of Constantinople), which was arguably the first university in the world, *complemented* the governance mechanisms by fostering a **culture** of secular scholarship, education and encyclopaedism, while attempting to preserve, revitalize but also to expand Greco-Roman civilization (which eventually led to modern-day Western civilization), thus cultivating a strong *sense of identity* and also reinforcing *legitimacy*.

The aforementioned reasons, *inter alia*, allowed the Roman and then Eastern Roman Empire to survive and prosper for so long, arguably making it the *longest-lived empire* in the history of mankind, since it essentially lasted from 27 BC to 1453 AD — an impressive grand total of 1480 years of prosperity, innovation and triumphs.

Therefore, this whitepaper proposes the concept of *oscillatory decentralization*, which allows a governance system to **oscillate** (similar to a pendulum) between the states of full decentralization (i.e. crest) and semi-decentralization (i.e. trough), as shown in Fig. 25, based on the postulate that decentralization exists in a continuum (Section 3.2.2), and provided that the resulting **sociopolitical system** would or should be *stable*, by either performing dumped or sustained oscillations, as shown in Fig. 26 (b) and (c) respectively.

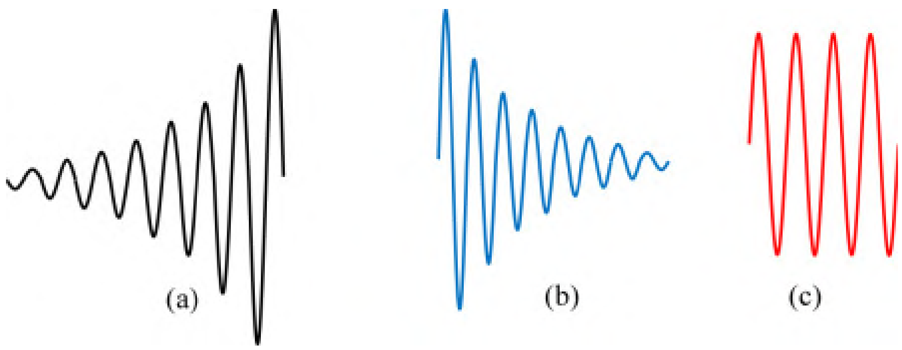


Figure 26 - Examples of (a) unstable, (b) dumped and (c) sustained oscillations.

The claim is that instead of naïvely, futilely and pointlessly trying to achieve a "static" trade-off solution between the two aforementioned states right from genesis (i.e. $t = 0$) by modelling a governance system as a static one, it is more realistic to: (a) admit that finding such a static solution is either a 'needle in a haystack' or even practically impossible, and (b) model it as a **Dynamical System** that allows for flexibility and adaptability, as shown in Fig. 26. This model takes into account that the world around us changes — "παντα ῥεῖ", i.e. "everything flows" according to Heraclitus — and should be allowed to change in order to make it adaptable to **evolution** and **innovation**.

The ability to be flexible or **change**, or merely considering the possibility of change, even if it never materializes: (i) strengthens the awareness of having options, (ii) creates an environment where everyone feels comfortable to freely express his options as an active member of the sociopolitical system, and (iii) admits the reality that nobody is perfect and it is possible that the decisions made might *not* be the best right from the outset, so mistakes will inevitably be made (and should be allowed to be made), while such mistakes will eventually be a *learning* experience both individually for each participant, and collectively for the community as a whole. These are some of the founding principles behind Agile Methodology, and also many other real-life systems that embrace **feedback** (Section 2.2.7).

Therefore, the problem becomes: (i) what governance mechanisms should be put in place in a sociopolitical system in order to cause *oscillations* of decentralization to occur in the first place in order to allow change to naturally and organically occur (but *not* be forced) due to aforementioned reasons, and (ii) what stability mechanisms should exist in such a system in order to ensure *stability*, i.e. avoiding political instabilities, as shown in Fig. 26 (a), that might eventually prove catastrophic for the community as a whole.

The term ‘sociopolitical system’ is deliberately used in order to emphasize that, whether we like it or not, such a system inevitably results from its respective governance system, and it will always form as long as more than one human participant is involved, whether the participants in it realize that or not. It is also used to highlight the need for the social interactions (as described in Section 4.1.3) to be transparent, because otherwise, the *trustworthiness* (as also described in Section 4.1.3) of the governance system would most likely decline.

Overall, this whitepaper proposes that the solution to the aforementioned problem is to impose **control mechanisms** (both formally through rules, but also, informally as norms, as explain in Section 4.1.3) similar to the ones that ensure adaptability, flexibility and ultimately, stability in a closed-loop control system (Fig. 3), such as self-correcting rules and by-laws using AI (as described in Section 4.1.6), under the important *assumption* that a critical mass of **information** (e.g., social interactions) would be recorded *on-chain*. However, it should be reiterated that the suggestions found in this whitepaper are towards making a governance system both more **human-centred** and more **user-centred**, *not* more machine-centred, in order to *avoid* both a human-imposed and a machine-imposed *dystopian* scenario.

4.1.8 Optional arbitration

Recognizing the fact, and being pragmatic about it, that there will never be enough rules in place to cover every single scenario regarding the Foundation governance means that there will be instances of conflicts among two or more TMCs (e.g., which specific TMC has jurisdiction over a specific matter), and the Board should not be obligated to intervene to each such instance, since otherwise, the overall *efficiency* and the *agility* of the governance system would be inhibited.

In such a case, the TMCs will have the *option* (but *not* the obligation) to use internal or external (third-party) **arbitration** in order to resolve the conflict. External arbitration is not prohibitive for a DAO foundation, since not being self-contained is not a hard constraint, in contrast to a blockchain network. In addition, given the modifiable character of the rules (Table 4), such an optional arbitration will be available to DAO members, thus increasing the trustworthiness (Section 4.1.3) of the governance system, while still maintaining a high degree of **trustlessness**, since there is *no* centralized authority that needs to be trusted.

4.1.9 Network-agnostic

It is important to reiterate that the Foundation proposed in this whitepaper is **network-agnostic**, since it will be totally decoupled from the Network that is also proposed in this whitepaper. The reason is to ensure the *fully decentralized* nature of the Network both in *practise* and in the *perception* of its human participants (i.e. validators, non-validator users or otherwise), since “Caesar's wife must be above suspicion”, as the proverb goes. In other words, the Foundation can be active and sell its AI solutions (Section 4.1.10) in more than one network (and not just in the proposed Network) and *vice versa*, i.e. the Network is *foundation-agnostic*, as explained in Section 4.2.7.

4.1.10 AI Solutions and other product offerings

The fact that the Foundation is a non-profit one, it does not necessarily mean that it should not generate revenue or even profit (since the term ‘non-profit’ is a misnomer). Therefore, the Foundation as proposed in this whitepaper will follow the business model [71 - 73] of other organizations that generate significant revenue from **free and open-source software (FOSS)** *without* selling the software itself or the code behind it.

It should be made clear that, 'open-source' does *not* always imply that it can be **'freely**' used, i.e. 'libre' software as a matter of liberty not price, in the sense of "with little or no restriction" due to the license associated with the respective software. There are also additional subtle differences between organizations (e.g., GNU [74] and Open-Source Initiative [75]) about the definition of 'open-source' software, which are outside the scope of this whitepaper. However, all such definitions have tangible and practical implications. The overall aim of the Foundation is to make the software (including its underlying code) as **unrestrictive** as possible.

For instance, 'open-core' software [71 - 72] (i.e. *partial* availability of the code) is well *outside* the definition of 'open-source' software for the purposes of this whitepaper, since open-core software violates both the desiderata of Trustworthy AI and decentralization (e.g., security, transparency, permissionlessness, censorship resistance), while also *hindering* the spread of its use, the generation of network effects and ultimately, inhibiting the creation of an ecosystem of users [73].

Therefore, a partial list of some streams of revenue for the GUT-AI Foundation includes:

(1) **Transactions of data:** Offer a wide variety of *datasets* and other *data* (Section 4.4.1) via either the proposed Marketplace or elsewhere to individuals or other organizations, including the following:

  (a) *Industry-specific* data (raw or cleaned) to be used for Training or Inference.

  (b) *Simulated* or augmented data to be used for Training or Inference.

  (c) Data from (indoor or outdoor) *simulator* and game emulators.

  (d) *Pretrained* models for multiple modalities (e.g., image, text and audio).

  (e) Saved model *checkpoints* or parameters halfway during Training.

  (f) Any kind of *metadata* associated with the dataset or the model.

(2) **Transactions of goods:** Offer a wide variety of *products* (Section 4.4.2) via either the proposed Marketplace or elsewhere to individuals or other organizations, including the following:

   (a) Branded *merchandise* such as t-shirts, hoodies, stickers, coffee mugs and other promotional items to users and fans of the software.

   (b) Dedicated Internet of Things (IoT) devices, workstations or mining equipment with the software preinstalled on and bundled with them.

   (c) Development boards, such as Arduino or Raspberry Pi, which enable prototyping, learning and experimentation with open-source software.

   (d) DIY and educational electronics kits, including circuit boards, electronic components, sensors, wires and cables.

(3) **Transactions of services:** Offer a wide variety of *services* (Section 4.4.3) via either the proposed Marketplace or elsewhere to individuals or other organizations, including the following:

   (a) Cloud *hosting* services and server access specifically tailored to the software, providing reliable, cost-effective and optimized infrastructure for users who prefer a hassle-free deployment experience.

   (b) Premium *support* and *maintenance* contracts to entities that require dedicated assistance, troubleshooting and ongoing updates for the software.

   (c) *Customization* services to those entities who require tailored solutions or additional features beyond the standard offering of the software.

   (d) Holistic *education* services and *training* programmes, one-on-one tutorials, video courses and certifications in order to assist users in understanding and maximizing the potential of the software, while enhancing their skills and knowledge. Such services and programmes can be provided in collaboration with an Education Technology partner.

(e) Professional services to help entities *migrate* their existing data or *integrate* the software with their current systems, ensuring a smooth transition and minimizing disruption.

(f) *Advisory* services based on the expertise, research findings, strategic insights and consumer behaviour analytics collected by the Foundation through both primary and secondary research in order to help other organizations (using the free and open-source software) make better-informed decisions, and especially in order to provide actionable recommendations that will improve the experience of their respective users.

(4) **Crowdfunding:** Initiate crowdfunding campaigns to fund certain development efforts or specialized projects, allowing individuals and organizations to contribute financially for a very specific cause related to the software.

(5) **Sponsoring and voluntary donations:** Allow an individual or another organization (either a for-profit company or otherwise) can provide funding either (a) with no strings attached as a voluntary donation or (b) alternatively, in order to monetary sponsor the implementation of certain features, improvements, bug bounties, translations, software integrations or the creation of a whole new software in exchange for such funding.

Overall, the point is that one way or another, the Foundation will participate in the Ecosystem on equal terms with all other participants, *without* exerting any undue or disproportional control over any other foundation or participant, as explained in Section 4.2.2. Inside the Ecosystem (Section 4.2), the Foundation will offer AI Solutions and other products offerings, such the ones mentioned above. Both the Foundation and the Ecosystem are efforts towards accelerating the broader GUT-AI Initiative (Chapter 5).

4.2 The Ecosystem

When it comes to writing a whitepaper (and especially the 'solution' part), there is no typical structure or way in the literature of writing it. However, two main ways stand out, either (a) full of technical jargon and details or (b) full of financial jargon without any technical details. In this whitepaper, the authors choose an approach that is different from both of them. Without getting into either too many technical or too many financial details, the solution is presented in a way that maximizes creativity, as mentioned in the beginning of this Chapter, but also throughout the whitepaper, while also increasing readability and comprehension to a general audience from multiple backgrounds. In other words, it states the *'what'*, but not the *'how'* the solution will be implemented. This Section deals with the **Ecosystem**, which is one of the efforts towards accelerating the broader GUT-AI Initiative (Chapter 5).

4.2.1 Protocol Architecture

It should be highlighted that the term 'layer' in this Section is *not* used to define Blockchain (network) layers, but new layers proposed in this whitepaper by the authors, termed as **GUT-AI Protocol** layers (Fig. 27). The two types of layers are *orthogonal* to each other, and they can actually be mixed and matched, as shown in Fig. 28.

In total, the proposed architecture (Fig. 27) consists of five GUT-AI Protocol layers:

- Layer 1 - Foundation Layer
- Layer 2 - Infrastructure Layer
- Layer 3 - Platform Layer
- Layer 4 - Intelligence Layer
- Layer 5 - Innovation Layer

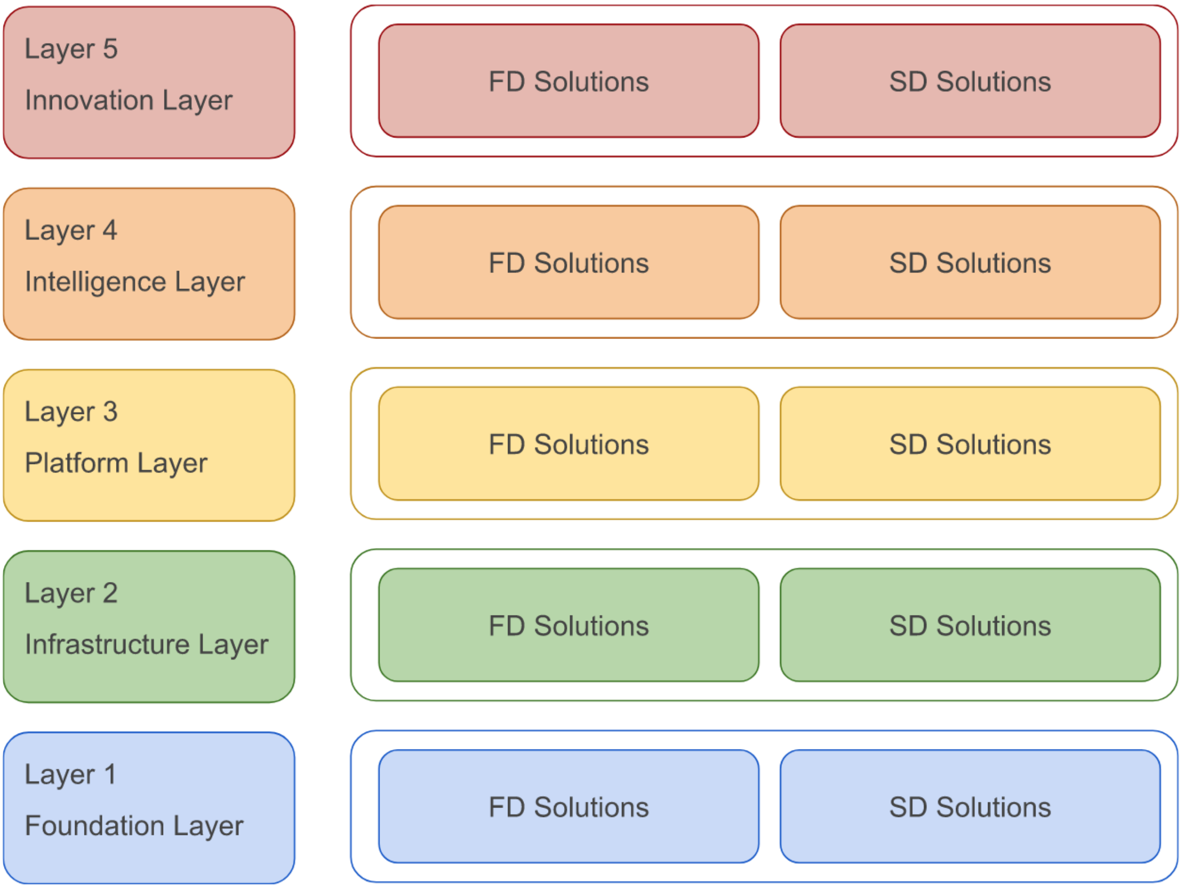


Figure 27 - Five-Layer Architecture of the proposed GUT-AI Protocol.

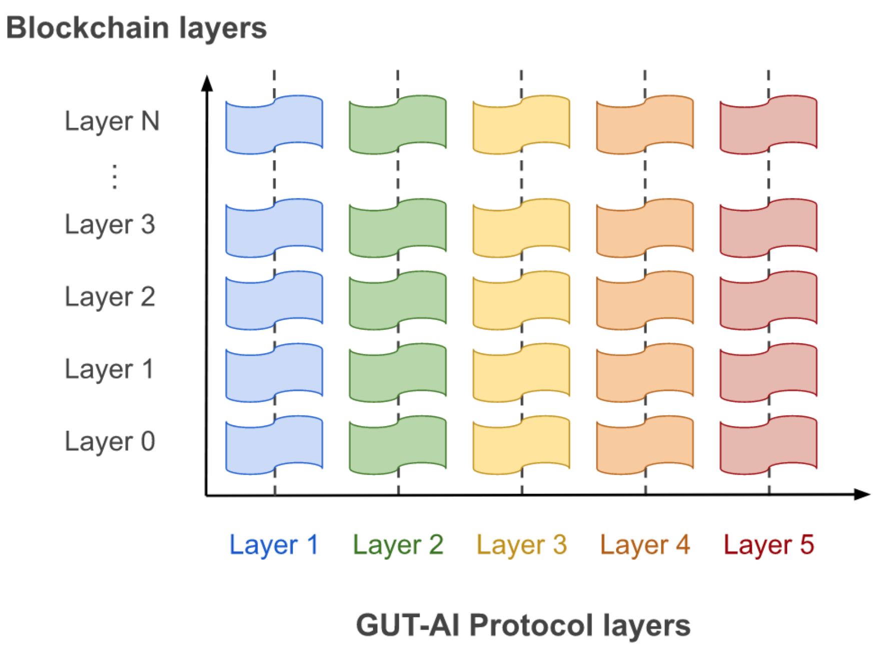


Figure 28 - The relationship between Blockchain layers and GUT-AI Protocol layers.

4.2.2 Everything-as-a-Service

It should be emphasized that, similar to the proposed Foundation (Section 4.1), the proposed Protocol (Section 4.2.1) is only a **part** of the whole Ecosystem as proposed in this whitepaper. The Ecosystem (Fig. 29) is a combination of all of the following:

(1) **Humans:** The Ecosystem will primarily be built around humans (i.e. decision-makers, miners, validators, Researchers, Developers and any other type of users), thus making it both *human-centred* and *user-centred*.

(2) **Organizations:** A lot of humans gathered together have a collective bargaining power. This can have the form of foundations (such as the GUT-AI Foundation), companies (such as IDEs, CLEs, and CSEs) or any other type of organization. Actually, the Ecosystem is foundation-agnostic, as explained in Section 4.2.7.

(3) **GUT-AI Protocol:** As mentioned in Section 4.2.1, the Protocol is an integral part of the whole Ecosystem and it does *not* refer to a 'network protocol'.

(4) **Blockchain networks:** As elaborated in Section 4.2.6, the Ecosystem is deliberately blockchain-agnostic.

(5) **Software:** AI Solutions offered by the Foundation (Sections 4.1.10) and any other type of software also form an integral part of the Ecosystem.

(6) **Hardware:** Predominantly Layer 1 of the GUT-AI Protocol is highly intertwined with and dependent on devices and other types of hardware.

Additionally, in terms of implementation, each and every *component* (Section 5.4) of each Protocol layer should necessarily be provided **"as-a-Service"** since ensuring an exceptional *experience* for both users (Section 2.2.5) and Developers (Section 2.1.4) is of paramount importance. Also, each **component** should be implemented as a *plug-and-play* solution (as explained in Section 2.2.6), thus incorporating the design principle of *modularity*.

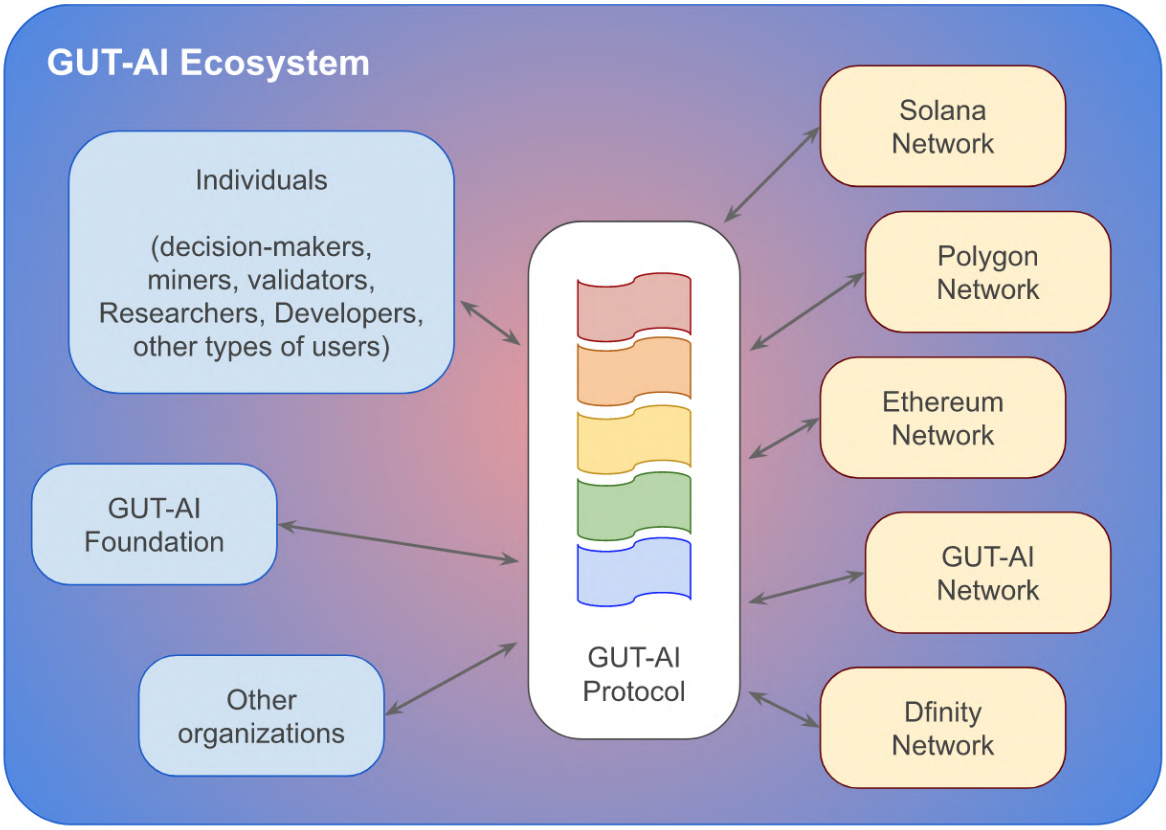


Figure 29 - The relationship between the Foundation, the Protocol, the Network, other organizations, other blockchain networks, the Ecosystem and various types of individuals.

The GUT-AI Ecosystem is intended to also act as a shared **collaboration platform** among organizations (Section 2.2.8), thus acting as a *decentralized* common point of reference, but *without* the multiple single points of failure found in Web2 (as explained in Section 3.1). It should also be reiterated that the GUT-AI Protocol is *not* a blockchain or any other incarnation of a network. Outside network and communication protocols, the term 'protocol"' has a wider meaning. In the broad sense, a **protocol** is a set of established *rules, procedures or customs* that govern behaviour, communication, or interactions within a specific context or domain. It provides a *structured framework* for conducting activities and ensures smooth and efficient coordination between individuals or entities. The ultimate goal of a protocol is to promote order, clarity, and effective collaboration. Following protocols helps maintain *standards* (Section 4.2.3) and achieve the desired outcomes.

4.2.3 Optional Open Standards

The existence of a protocol in the broader sense (as explained in Section 4.2.2) generates a need for the creation of **Optional Open Standards** and the dissemination of *good practises*. In general, such standards play a crucial role in the field of AI due to their significance in fostering innovation, promoting *interoperability*, while also addressing *trustworthiness* (as defined in Section 3.3). The combination of *openness* and *optionality* increases the number of stakeholders and empowers them, while also maintaining a balance between standardization and flexibility.

Firstly, **openness** is essential in the context of AI as it encourages collaboration and knowledge sharing. By adopting open standards, Developers and Researchers can freely exchange ideas, methodologies, and best practices. This fosters innovation and accelerates the development of AI methodologies, enabling (community) *alignment* and collective progress in the field. Open standards also promote *transparency*, enabling stakeholders to understand and scrutinize the underlying models, thereby addressing concerns related to trustworthiness.

Secondly, **optionality** provides flexibility and adaptability in the rapidly evolving landscape of AI. Since AI encompasses a wide range of applications, industries, and contexts, a one-size-fits-all approach is often impractical. Optional standards allow organizations and individuals to choose from a range of specifications and technologies that best suit their specific requirements, while also *'maximizing for creativity'* (Section 3.3.3). Such a flexibility facilitates the integration of AI systems across different platforms, promotes *interoperability*, and avoids vendor lock-in. It also encourages *competition*, as multiple standards coexist, fostering innovation and preventing monopolistic practices.

Furthermore, the *combination* of openness and optionality together helps ensure *scalability* and *agility* by providing flexibility, adaptability, and responsiveness to the rapidly evolving landscape of AI. As technology continues to advance and evolve, the ability to scale systems becomes crucial. Optional Open Standards allow for the integration of new functionalities and technologies without significant disruption. Such a scalability ensures that AI systems can handle growing volumes of data, support increased user demand, and adapt to changing business needs. Meanwhile, agility facilitates the integration of new functionalities, technologies, and advancements in AI, ensuring that systems can quickly adapt to changing requirements and stay at the forefront of innovation.

4.2.4 Utility token

IMPORTANT: It should be made clear that everything included in this Section is *tentative* and subject to change. Also, it should be highlighted that this whitepaper does *not* describe the so-called Token Economics (or 'Tokenomics'), since they fall *outside* the scope of this whitepaper since it focuses on the Foundation. Instead, a *separate* whitepaper dedicated to the token will deal with such matters. Therefore, everything mentioned in Section is *indicative* in order to demonstrate one possible implementation of the proposed token.

The **AGI** utility token (a *placeholder name*) is a digital asset built on a blockchain network, leveraging smart contract technology, while serving as a default *medium of exchange* within the GUT-AI Ecosystem (Fig. 29). AGI tokens possess intrinsic value by offering to its holders a range of practical *utilities* (as detailed in Section 4.2.5) that enhance the user experience, facilitate transactions, and incentivize active participation in the Ecosystem's decentralized applications (dApps) and dPlats, while also playing a crucial role in enabling seamless interactions, value exchange, and fostering a thriving community within the Ecosystem.

A consensus mechanism very similar to the one in Decred [78-80] is proposed in this whitepaper. Decred is a *governance-focused* cryptocurrency that utilizes *neither* solely the '1 computer = 1 vote' of Proof-of-Work (PoW) consensus *nor* solely the '1 token = 1 vote' of a Proof-of-Stake (PoS) consensus, but instead, it is a *hybrid* between the two, called Proof-of-Activity (PoA) in order to get the best of both worlds, while overcoming certain limitations found in either PoW or PoS or both (as explained in [79]). A new type of PoA, called **Proof-of-Contribution (PoC)**, is proposed in this whitepaper.

Briefly, the *validation process* in the proposed PoC is the following:

(1) **Ticket purchasing:** In order to participate in PoS voting, token holders must firstly time-lock some of their AGI coins to buy *non-transferable* "tickets" at a market-like mechanism where the system aims for a set number of live tickets (51,200). The ticket price changes dynamically every time at a predetermined interval (144 blocks or around every 12 hours). When someone buys a ticket, the AGI coins they used are locked either: (a) until their ticket is pseudorandomly called to vote or (b) for a predetermined period of time (51,200 blocks or approximately 178 days) if it is never called to vote, in which case it expires and a refund is issued at the original price of the ticket. This type of mechanism introduces an opportunity cost for PoS, ensuring that "ticket holders" (or "stakers") have skin in the game and act in the Network's best interests. In general, ticket holders have three distinct roles: (i) scrutinizing PoW "miners", (ii) voting on changes to the consensus rules, and (iii) voting on technopolitical governance of the Network (Section 4.2.8). An individual token holder can buy any number of tickets, and every ticket holder has the ability to cast a single vote for every ticket that he holds. Upon voting, each ticket returns a small reward plus the original price of the ticket. The probability of a ticket voting will be based on a Poisson distribution with a mean of around 28 days.

(2) **PoW stage:** In the PoW stage, miners compete to solve some kind of computationally intensive puzzles to create a new block by expending energy. In terms of the PoW stage, it would be more preferable to solve a Machine Learning problem that has some practical utility rather than to wastefully burn cycles on meaningless time-consuming (cryptographic) puzzles. The first miner to successfully find a solution to the puzzle becomes eligible to create a new block. The winning miner broadcasts the new block to the network, along with the solution to the puzzle. But instead of immediately adding it to the blockchain, the stakers need to vote on it. This PoW mechanism ensures that the miner has put in some minimum computational effort, thus ensuring the security of the network. The mining difficulty adjusts automatically every time at a predefined interval (roughly 12 hours).

(3) **Tickets selection:** Using a lottery system, the network pseudorandomly selects 5 tickets from the pool of purchased tickets in order to validate the correctness of the new block.

(4) **Block verification:** Unless it receives a minimum of 3 ticket votes in favour, the new block will not be recognized as valid by the Network and will be rejected. Otherwise, it will be approved and added to the blockchain.

(5) **Block reward allocation:** In contrast to Bitcoin, block reward is *not* allocated exclusively to the miner, but distributed among the miner, the stakers and the Collective Funds System (CFS) using a predefined allocation of 60%, 30% and 10% respectively. The CFS is a form of treasury with its own governance (as explained in Section 4.2.9), *independent* of that of the Network.

PoW, PoC and PoS consensus mechanisms are compared and contrasted in Table 5 in order to highlight main advantages and disadvantages of each such mechanism, while also making clear why the PoC is chosen for the proposed Network.

PoW	PoC	PoS
High risk of centralization due to data centres outnumbering the hobbyist miners, due to economies of scale	Low risk	Low risk
Low risk	Low risk	High risk of centralization due to large stakers trying to exhibit control over the network, and stake grinding or other stake-related attacks
Tragedy of the Commons problem, as the transaction fees are paid only to the miner who created the block	Block reward is distributed among the miner, the staker and the CFS	Tragedy of the Commons problem, as the transaction fees are paid only to the staker who created the block
Less scalable	More scalable	More scalable
High risk of hard forks	Low risk of hard forks	High risk of hard forks
Low risk	Low risk	High risk of nothing-at-stake problem (to also validate blocks on forked chains)
Medium	Increased network security by rewarding active participation	Medium
High associated costs (energy consumption, wear and tear on the hardware) and barriers to entry	Medium associated costs and barriers to entry	Low associated costs and barriers to entry
Low risk	Medium risk	High risk of coin hoarding (rather than spending)

Average transaction processing speed on Layer 1	Average transaction processing speed on Layer 1	Fast transaction processing on Layer 1
Transaction processing speeds of up to 1 million TPS on Layer 2 (e.g., Lighting network)	Transaction processing speeds of up to 1 million TPS on Layer 2 (e.g., Lighting network)	Transaction processing speeds of up to 65,000 TPS on Layer 2 (e.g., Polygon)
High risk of selfish mining	Low risk	Medium risk
No governance mechanisms within the consensus process	Yes	Possible
Limited participation	Enhanced participation by non-mining users	Possible participation

Table 5 - Comparison between PoW, PoC and PoS consensus mechanisms.

Each blockchain network serves a *different* purpose, and its associated consensus mechanism plays a crucial role in it. For instance, Ethereum recently migrated from PoW to PoS because it better serves their stated purpose. Dfinity uses PoS which facilitates their on-chain governance and participation of non-validator users, since it implements a DAO-controlled network (as explained in Section 4.1.2). In the case of the GUT-AI Network, maximizing market capitalization is, for example, a secondary goal that complements the primary purpose, i.e. to implement the proposed Ecosystem, which is the motivation behind this whitepaper. Therefore, PoC was chosen by the authors to better serve this purpose. In other words, the proposed (Layer 1) Network and any other blockchain layer built on top of it (Fig. 28) are a means to an end, and *not* the end itself, as explained in Section 4.2.4, which elaborates on the utility of the proposed AGI token.

Provided that the proposed Network is a means to an end, and not the end itself, there will be two *pre-coin phases*, which will serve as transition periods before the implementation of the Network, and will be implemented in the following way:

- **In Phase A:** There will be no coin and instead, an existing Layer 1 blockchain will be chosen in order to build dApps and dPlats that implement some parts of the proposed Ecosystem.
- **In Phase B:** A utility token different from AGI will be implemented as a Layer 2 blockchain and all dApps and dPlats implemented in Phase A will be migrated to this new utility token.

The two above-mentioned phases will be followed by **Phases C1**, **C2** and **C3**, as explained in Section 4.2.8, which elaborates on network governance.

Regarding Phase B, Ethereum [58], Solana [81], Polygon [82] and Dfinity [62] have all token programmes that allow everyone to create their own token. Each such token programme has its own benefits and limitations, while each such blockchain network has its own distinct ecosystem of dApps and dPlats, along with its own distinct developer community. For example, Dfinity offers the advantage of natively building web front-ends for dApps, with existing solutions that can optionally be used for “bridging” Web2 interfaces (that the vast majority of users is more accustomed to) with Web3 solutions, such as a user-friendly web *authentication system*, which provides a high level of convenience and a very low level of friction for the users, since it eliminates the need to directly manage or handle cryptographic keys themselves. Polygon has a much larger ecosystem and developer community than Dfinity, given that it is a Layer 2 blockchain on top of the Ethereum, but without the known disadvantages of Ethereum such as relatively high and unpredictable gas fees. Therefore, given the trade-offs associated with each one, the decision for what token programme to be initially chosen will be taken by the Foundation.

4.2.5 Maximizing for utility

It cannot be emphasized enough that *AI, decentralization* and *Blockchain* technology are all a *means* to an end, *not* the end itself. In other words, these innovative technologies should be viewed as powerful **tools** that enable us to achieve specific goals and solve real-world problems, rather than as standalone solutions. While these technologies possess immense potential to revolutionize industries and advance the human species as a whole, they are not a panacea. Their effectiveness highly depends on *how* they are implemented, applied and eventually integrated into existing systems. Treating them solely as standalone solutions can lead to *misplaced* expectations and *missed* opportunities. By recognizing them as tools, we acknowledge the need for human judgement, critical thinking, and responsible Decision-Making to harness their full potential and achieve meaningful outcomes that align with both the collective and individual values and objectives of the community.

For instance, the AGI token has a primary purpose to serve as a functional *tool* within the proposed Network in a way that it is **'maximizing for utility'**. In particular, while financial gains and speculative investments may unavoidably be part of the equation, the true value of a token lies in its *practical* use and ability to facilitate meaningful interactions and transactions. By prioritizing utility, it is ensured that tokens are designed to solve real-world, practical problems, while providing tangible benefits, and enabling seamless functionality within the Network. A *utility-focused* approach fosters trustworthiness, promotes adoption, and enhances the overall sustainability and long-term viability of the Ecosystem, as it *aligns* the token's value with its practical applications and utility for users. When tokens have clear and valuable use cases, they become more attractive to users, businesses, Developers, and Researchers, thus leading to increased adoption and stronger *network effects*. This, in turn, strengthens the Ecosystem, and establishes a virtuous *cycle of innovation* (Fig. 14).

More specifically, the primary **token utilities** will include the following:

(1) **Access to the Marketplace:** By utilizing the AGI utility token, users can effortlessly buy, sell and procure AI solutions from the proposed Marketplace (Section 4.4). The token serves as a means of transaction, enabling secure and transparent payments for accessing and utilizing such AI solutions. This streamlined process eliminates the need for intermediaries, reduces transaction costs, and ensures a frictionless experience for both solution providers and consumers.

(2) **Access to dApps and dPlats:** Access to dApps and dPlats provides significant utility for the AGI token, because it allows its holders to participate in a wide range of decentralized services. By granting access to these dApps and dPlats, the proposed Network enables users to leverage the power of AI and Blockchain technology, fostering research, technological development and innovation. This utility also enhances the value proposition of AGI by attracting users, Developers, Researchers and organizations to the Ecosystem, thus driving further adoption of the token. Some decentralized services can include the following:

   (a) **User-friendly web front-ends** in order to provide a gateway for all types of users to interact with the Ecosystem.

   (b) **Open data sharing platforms** in order to disseminate open data.

   (c) **Data validation and verification systems** in order to validate and verify the integrity and authenticity of scientific data.

   (d) **Decentralized gaming platforms** in order to facilitate environments and simulators for RL by creating immersive and interactive gameplay experiences, whose environment changes dynamically.

   (e) **Collaborative research platforms** in order to enable real-time collaboration, availability of pretrained models and knowledge transfer.

(f) **Trustless reproducibility platforms** in order to provide a transparent and immutable record of experimental protocols to enable trustless verification, validation and reproducibility of research findings.

(g) **Decentralized publication platforms** in order to ensure immutable records for open-access publications and fair distribution of revenue among authors.

(h) **Tokenized peer review systems** in order to financially incentivize reviewers with tokens for providing high-quality and timely reviews.

(i) **Decentralized knowledge bases** in order to create peer-reviewed, censorship-resistant and interactive encyclopaedias using Extended Reality, while also storing it in a blockchain-based time capsule to preserve them securely and immutably for future generations to access and explore.

(j) **Decentralized AI-powered assistants** in order to provide Researchers with customized recommendations, insights, and scientific news, thus helping them navigate efficiently through the vast research literature.

(k) **Decentralized telemedicine platforms** in order to reach *developing and low-income countries* that have zero or extremely limited access to medicine.

(l) **Decentralized social media platforms** in order to connect established Researchers with aspiring individuals, thus facilitating mentoring relationships, relationships with IDE organizations, and dissemination of research findings.

(m) **DeFi** in order to facilitate transparent, auditable and accountable funding and grant allocation for scientific research projects, while also enabling commercialization of research findings.

(n) **Decentralized e-commerce platforms** in order to enhance customer experiences through AI-driven product recommendations, personalized marketing, and intelligent payment gateways.

(o) **Enterprise AI solutions** in order to provide *industry-specific* solutions to enterprises that build on top of generic open-source solutions.

(3) **Fundraising, investment and donations:** Given that the broader GUT-AI Initiative (Chapter 5) is a totally decentralized initiative, anyone will be able to independently propose and start a project in order to create software that can fit in any of the five layers of the GUT-AI Protocol (Section 4.2.1). Therefore, AGI tokens can provide a decentralized mechanism for fundraising by allowing such projects to secure funding *without* relying solely on traditional avenues such as venture capitalists or government grants. As a result, fundraising efforts can reach a global audience, enabling individuals from anywhere in the world to contribute to these projects, while attracting a diverse range of contributors and Investors and fostering a sense of community. Moreover, by enabling fractional ownership of such a project, Investors can hold tokens representing a share of the project, thus providing them with financial incentives and economic benefits (e.g, voting rights, profit-sharing) as the project succeeds, while also promoting transparency, auditability and efficiency regarding projects' finances. Additionally, AGI tokens will allow for micropayments, philanthropic contributions and crowdfunding campaigns towards the advancement of AI research and its potential societal benefits, thus enabling small contributions from numerous individuals. Furthermore, with both fundraising and conversations happening on-chain, token holders can participate in voting processes to determine project directions, feature prioritization or resource allocation, while facilitating presales to potential customers, interoperability within the Ecosystem, external partnerships, collaborations with other projects, and project mergers in order to avoid duplication of work by easily accessing information about project objectives, task allocation, research findings and key milestones. Consequently, the Network becomes a decentralized *common point of reference* for AI-related projects.

(4) **Default method of exchange:** The AGI utility token will act as a native currency within the Network, facilitating secure and efficient transactions. Users can utilize AGI tokens to pay for services, products, and subscriptions offered within the Ecosystem. The token's integration simplifies the payment process, reducing transaction fees, and enabling cross-border transactions without the need for traditional intermediaries. Although the proposed Ecosystem is blockchain-agnostic (Section 4.2.6) and the network protocol does not require a specific payment type, the Network assumes the AGI token as the *default* mechanism for payment. This default method of exchange can be a valuable utility for a blockchain token because it offers a streamlined and convenient payment process, while promoting *interoperability* across different applications and platforms by enabling the token to be used in various dApps and dPlats. Therefore, by implementing such a mechanism, the AGI token becomes the automatic payment option for users, thus eliminating the need for traditional payment methods, while improving the user experience by making payments *frictionless* and *efficient*. Additionally, the default payment mechanism promotes token *adoption* and liquidity within the overall Ecosystem, since users are incentivized to acquire and hold the token for relatively longer periods of time as it becomes an integral part of their everyday transactions. This adoption translates to greater liquidity for AI-related projects within the Ecosystem, and potentially, a greater impact of Trustworthy AI (Section 3.3) in businesses and individuals alike.

(5) **Encouragement and promotion of Trustworthy AI:** The AGI token can serve as a powerful tool for encouraging and promoting Trustworthy AI (Section 3.3). By leveraging the decentralized and immutable nature of the Blockchain technology, tokens can incentivize Developers, Researchers, and users to prioritize and uphold the principles of *trustworthiness* in their AI systems in order to facilitate

transparency and auditability in AI systems. For instance, tokens can be used to *track* and *verify* the provenance of training data used in pretrained models, thus ensuring that such data come from trustworthy and reliable sources. Additionally, tokens can enable the creation of decentralized *governance systems*, where token holders have a say in Decision-Making processes related to the development and deployment of AI systems, thus aligning the interests of all stakeholders.

(6) **Governance and voting rights:** AGI token holders will possess governance and voting rights, thus enabling them to actively participate in Decision-Making processes concerning the changes, upgrades and future direction of not only the AI systems, but also that of the Network itself. This approach ensures that the Network evolves in a manner that aligns with the interests and preferences of its community members through a combination of *direct* and *indirect democracy*. This level of involvement strengthens the relationship between the Network and its participants, thus fostering collaboration and innovation.

(7) **Smart contracts:** AGI tokens can leverage smart contracts, which are self-executing agreements with *predefined* rules, which can automate various aspects of transactions (i.e. “Code is Law”) in order to ensure transparency, trustworthiness and efficiency, since they eliminate the need for manual confirmation and reduce potential errors. For example, in the context of DeFi, smart contracts can facilitate *lending* and *borrowing* without financial intermediaries, thus increasing the liquidity of AI-related projects under development. A borrower can provide collateral in the form of tokens, and the smart contract *automatically* executes the loan and enforces repayment terms, while eliminating the need for traditional lenders and reducing costs. In general, smart contracts promote a *frictionless* user experience by simplifying the interaction among users, thus making the proposed Network more accessible to a wider audience.

(8) **Incentives and rewards:** AGI tokens can be staked within the proposed Network, therefore allowing users to lock their tokens for a specified period. Staking provides additional security to the Network and rewards users with a share of transaction fees or newly minted tokens. Staking AGI tokens contributes to network stability while offering users a passive income opportunity. Also, users can be rewarded for identifying and reporting vulnerabilities through bug bounty programmes, thus contributing to the security and integrity of the Network.

(9) **Fair Rewarding:** AGI tokens can be used to reward content creators and curators (Section 3.1.3) within a dApp or a dPlat, thus incentivizing the production of high-quality content and promoting community engagement. Furthermore, active engagement and contributions can be incentivized by rewarding users who provide valuable data, pretrained models, insights or compute resources to other users. This mechanism encourages collaboration, innovation, and the sharing of knowledge, creating a dynamic environment where stakeholders actively contribute to the growth and development of the Ecosystem, while also fostering a more collaborative and engaged community, thus leading to improved social cohesion.

(10) **Supply chain tracking:** In combination with cost-effective technologies (such as QR codes), tokens can be employed to *track* and *verify* the authenticity, provenance, and movement of goods, while enhancing transparency and strengthening confidence in the *supply chain*. With immutable records on the blockchain, consumers can *trace* every single step of the process, from raw materials to final delivery, thus ensuring standards adherence, quality control, and verifiable ethical farming, for example. More specifically, traceability of agricultural products can be ensured from source to retail. In general, the token-based system offers a *tamper-resistant* solution, thus enabling efficient verification, and promoting

integrity and reliability within the supply chain network, while fostering increased accountability, streamlining auditing processes, and enabling seamless traceability.

(11) **Decentralized agricultural systems:** Agriculture is a sector of the economy that considerably lags behind from all other ones in terms of using cutting-edge technologies. Additionally, food security is a serious challenge, especially among *developing and low-income countries*. As a result, farmers and consumers alike are currently not reaping the full benefits of technological advances such as Blockchain, AI and IoT. Even though the growth of IoT with wireless technologies has resulted in some advances for smart farming systems, there is a serious lack of mass adoption due to various factors, such as high upfront costs, complexity, low-quality user experience, lack of awareness and training, data privacy and security concerns, lack of standardization and interoperability, reliance on intermediaries, physical infrastructure limitations, and limited access to reliable power sources. The AGI token will be uniquely positioned to practically address most of these issues in an impactful way, since AI plays a crucial role in *optimizing* agricultural processes, from crop monitoring and yield prediction to disease detection and resource allocation. However, accessing and validating accurate agricultural data can be challenging due to information silos and data privacy concerns. So, the token can incentivize data sharing and collaboration among farmers, Researchers, Developers and other stakeholders, while ensuring that data are securely stored and immutable, thus fostering trustworthiness, reducing the need for processor intermediates and promoting standardization and interoperability. Also, smart contracts can automate transactions and streamline payment processes, ensuring Fair Rewarding for data contributors, while also providing liquidity and financing opportunities in the context of DeFi through dedicated crypto lending platforms, without the need of financial intermediaries and without the reliance on any centralized authority.

(12) **Distributed smart grids:** By utilizing the AGI utility token, users can ultimately benefit from real-life experiences, such as using a *grid-connected microgrid* for both electricity and communication through the implementation of *distributed smart grids* (Section 4.3.1). Firstly, tokens can enable secure and transparent transactions between *residential* producers and consumers (i.e. prosumers), thus facilitating P2P energy trading and incentivizing self-sustainability and self-sufficiency (i.e. autarky) at the *community level*, thus relieving the burden of prosumers from the centralized authority. Secondly, tokens can provide a decentralized and tamper-resistant system for tracking energy usage, reducing the need for intermediaries, while ensuring trustworthiness in the local community. Lastly, blockchain tokens can encourage grid participants to contribute to grid *stability* and *efficiency* by rewarding them with tokens for actions like sharing excess energy. As added benefits, the risk of a single point of failure is reduced, and data privacy for grid participants is enhanced, since the need for a centralized authority to store sensitive data is eliminated right at the level of the *physical* infrastructure.

(13) **Decentralized Cloud Provider:** Similar to the aforementioned distributed smart grids, the need for a centralized authority to store sensitive data is eliminated right at the level of the *physical* infrastructure in the case of a Decentralized Cloud Provider (Section 4.3.2), so the risk of a single point of failure is reduced, and data privacy for cloud participants is enhanced, while such cloud computing systems become more resilient and resistant to censorship. Additionally, tokens can facilitate low-cost transactions by reducing the need for financial intermediaries, and serving as the native currency for payments within these decentralized cloud computing systems, while incentivizing consumers to contribute their *idle* computing resources to the network, thus becoming prosumers. Such cloud computing systems also empower users to maintain *sovereignty* over their data, ensuring easier compliance

with data protection laws, and reducing the risk of unauthorized access or data breaches, while avoiding vendor lock-in situations. Furthermore, decentralized cloud providers can provide storage and compute resources for dApps by allowing seamless integration with them. Overall, by providing an open and decentralized marketplace for cloud services, AGI tokens foster *healthy competition* among all (decentralized and centralized) cloud providers, therefore leading to improved service quality and competitive pricing, both of which benefit the consumer.

(14) **Decentralized Physical Infrastructure Networks (DePINs)**: In addition to distributed smart grids and decentralized cloud providers mentioned above, the concept can be extended to any type of physical hardware, including desktops, laptops, mobile phones, wearables, IoT devices, sensors, wireless access, home automation systems, telecommunications, servers, data storage, compute resources, vehicles, transportation systems, logistics, charging systems, photovoltaic systems, electricity generation, energy storage systems, electric power distribution, demand-side management, water supply and distribution systems, wastewater management systems, irrigation systems, agriculture, hydroponics, aquaponics, aeroponics, telemedicine, healthcare infrastructure, and disaster response systems, just to name a few! Therefore, DePIN [77] is where Web3 meets the real world. The reason why DePINs (Section 4.3) are useful is due to its various advantages over traditional centralized infrastructure, which include the following:

(a) **Improved cost-effectiveness:** By optimizing resource allocation at the local level and reducing the need for infrastructural intermediaries, DePINs can potentially lower costs associated with infrastructure provision and operations, hence leading to a win-win situation between the service providers and consumers, since a huge planning and logistic burden is relieved from the former, while the latter ensures a level of *self-sufficiency* and *sovereignty*.

(b) **P2P interactions:** By facilitating direct interactions among participants, peer-to-peer transactions are enabled without having to rely on a financial intermediary taking a cut. In addition, this promotes collaboration, community engagement, and trustworthiness.

(c) **Local empowerment:** DePINs empower local communities and individuals by giving them a greater say in Decision-Making processes, while enabling them to actively participate in the management of infrastructure resources. This fosters a sense of ownership and encourages localized solutions, which is a good practise known as the *principle of subsidiarity* (Section 4.3).

(d) **Increased resilience:** By distributing both infrastructure components and Decision-Making, DePINs can adapt to local conditions and reroute resources, thus reducing the impact of system-wide failures and black-outs caused either accidentally or intentionally by malicious actors targeting the infrastructure. Therefore, DePINs increase the level of *resilience* to disruptions and failures.

(e) **Flexibility and innovation:** Decentralized infrastructure can be more *flexible* and *adaptable* to changing needs and circumstances. They can easily incorporate new technologies, accommodate diverse user requirements, and respond to evolving market demands in contrast to centralized infrastructure that is considerably more bureaucratic and hence, innovation-averse. Additionally, by reducing the barriers to entry due to the slashed upfront capital needs, DePINs increase competition into a variety of industries that have been resistant to change, thus incentivizing *innovation* across the board.

(f) **Self-correction:** Efficient allocation, usage and distribution of any type of infrastructure resources (i.e. energy, compute, water or otherwise) can be enabled by leveraging real-time data in order to optimize processes and promote *self-correction*, thus reducing waste and operational inefficiencies.

(g) **Eco-friendliness:** Since DePINs are *holistic* approaches and are controlled at the local or even individual level, they can have a real and meaningful impact on eco-friendliness, beyond corporate slogans or meaningless alphabet soups, such as “ESG”, “CEI”, “DEI”, “WEF”, and “CSR” that lack transparency. By enabling a level of transparency and accountability at the local level, participants are empowered to individually and collectively take actions that allow for efficient allocation, usage and distribution of energy sources in their community, since they can control a much larger part of the *energy lifecycle*.

(15) **NFTs:** In addition to the fungible AGI token, the Network will also support various types of NFTs, which will provide valuable utility to their holders. Some types of NFTs can include the following:

(a) **dataNFTs** in order to tokenize datasets, metadata, models or any other kind of data, so that their authenticity, integrity and cleanliness can be validated, while their origin, modifications and ownership history can be tracked.

(b) **bioNFTs** in order to tokenize and track ownership of biological and biomedical information, such as genetic information, DNA sequences, biological specimens (e.g., blood test results) and medical image data, so that each user can maintain sovereignty over such information, while enabling *personalized medicine*, but also, monetizing opportunities for the user via *privacy-preserving* and *anonymized* research in various medical and biomedical fields (e.g., bioinformatics, genomics and medical imaging).

(c) **accessNFTs** in order to tokenize levels of access rights (e.g., administrator, contributor) to code repositories that implement AI systems or to ensure that only authorized parties can use or train on specific data.

(d) **donationNFTs** in order to incentivize donations to AI-related projects that promote social good or other types of causes.

In summary, the AGI utility token acts as a catalyst for the Ecosystem, enabling users to unlock the potential of AI, transact securely and efficiently, participate in governance, and earn rewards for their contributions, including staking and Fair Rewarding. Its multifaceted utility is a necessary, but not sufficient condition, for the seamless operation and growth of the Network, while also leveraging the tremendous capabilities of AI to drive transformative change across multiple industries.

4.2.6 Blockchain-agnostic

The Ecosystem mainly depends on the proposed Network blockchain. However, it is useful or even necessary to support other blockchain technologies (as shown in Fig. 29) in order to broaden adoption, improve scalability, or achieve the stated goals of the Ecosystem. The proposed Protocol architecture (Section 4.2.1) is designed with this possibility in mind, since the Protocol layers are orthogonal to the blockchain layers (as shown in Fig. 28). A decision on which blockchains to support, and when, will be made by the Foundation, but the implementation of the *free and open-source software* (Section 4.1.10) preserves the provision for supporting other blockchains, thus ensuring that the proposed Ecosystem remains **blockchain-agnostic**.

4.2.7 Foundation-agnostic

As already explained in Section 4.1.9, the proposed Foundation is network-agnostic, but symmetrically, the proposed Network is **foundation-agnostic**, thus forming a semi-cooperative game (as explained in Section 4.1.2) between the Network and the Foundation. This means that the Network has its own governance (Section 4.2.8) *independent* of that of the Foundation (Section 4.1.1).

4.2.8 Network governance

In the early phases of Network development, even though all governance decisions will take the input of the Network from day 1, they will also consider the input from the Foundation. In particular, Decision-Making will transition in various phases from shared control to a *fully independent* governance as the Network matures, in the following way:

- **In Phase C1 (i.e. years 1 to 3):**
  - Major Changes will be decided by an agreement of the Foundation (in accordance with its by-laws) and a 65% *supermajority* of the eligible AGI token votes (in accordance with the Network rules).
  - Minor Changes will be decided by an agreement of the Foundation (in accordance with its by-laws) and a *simple majority* of the eligible AGI token votes (in accordance with the Network rules).
  - The definition of 'Minor Changes' and 'Major Changes' will be determined (and can be modified) by the Foundation (in accordance with its by-laws) in its sole discretion.
- **In Phase C2 (i.e. years 4 to 6):**
  - Major Changes will be decided by an agreement of the Foundation (in accordance with its by-laws) and a 75% *supermajority* of the eligible AGI token votes (in accordance with the Network rules).
  - Minor Changes will be decided by a *simple majority* of the eligible AGI token votes (in accordance with the Network rules).
  - The definition of Minor Changes and Major Changes can be modified by the Foundation (in accordance with its by-laws) and a 65% *supermajority* of the eligible AGI token votes (in accordance with the Network rules).

- **In Phase C3 (i.e. from year 6 onward):**
    - Major Changes will be decided by an 75% *supermajority* of the eligible AGI token votes (in accordance with the Network rules).
    - Minor Changes will be decided by a *simple majority* of the eligible AGI token votes (in accordance with the Network rules).
    - The definition of Minor Changes and Major Changes will be determined (and modified) by an 80% *supermajority* of the eligible AGI token votes (in accordance with the Network rules).

It should be noted that even though the Network governance will become *fully independent* during Phase C3, the Network can still decide to change its rules to include the Foundation in its Decision-Making, but at the sole discretion of the Network in accordance with the then current Network rules. In other words, the Foundation will need to prove its usefulness and relevancy during Phase C3. Also, an 'eligible AGI token vote' refers to a ticket vote (as defined in Section 4.2.4) cast by its holder, i.e. the staker.

The above-mentioned governance mechanism make the following main *assumptions*:

(1) The Network is already live (on the 'mainnet') as a **Blockchain Layer 1**, and the aforementioned years are measured from the point in time that it goes live, i.e. it excludes Phase A and B (as defined in Section 4.2.4).

(2) The Network has reached a *critical mass* of users (validators, non-validator users or otherwise) that it can ensure at least its survival.

(3) Such governance mechanisms do serve the *primary purpose* of the Network (as explained in Section 4.2.4) and do maximize for utility (Section 4.2.5).

Otherwise, if any of these assumptions is violated, then the Foundation might decide to propose and implement changes, conditional on ratification by the eligible AGI token votes.

4.2.9 Collective Funds System

The CFS acts as a form of *treasury* to the proposed Network, but with a governance that is *independent* of that of the Network (as described in Section 4.2.8) and the Foundation. In contrast to the governance of the Network, the CFS combines elements of both *direct* and *indirect democracy* in order to take into consideration the votes of **all** AGI token holders, and *not* just the votes of the ticket holders. The implementation of this is envisaged to be in the form of a DAO with an elected committee (Section 4.1.1), known as the Funds committee, to be elected as representatives of all the token holders, and with its own by-laws. Similar to the Foundation, the CFS will also allow *delegated voting* in order to tackle the common issue of low participation rates.

Therefore, the existence of the CFS, along with its governance mechanisms, ensures that every voice will be heard when it comes to allocating resources and managing the financial aspects of the treasury, while also ensuring **separation of concerns** regarding changes in the consensus rules (Section 4.2.4). Overall, the governance mechanisms of the Foundation, the Network and the CFS will hopefully produce the desired outcomes (as explained in Section 4.1.3), while also increasing *trustworthiness* from the perspective of the human participants (validators, non-validator users or otherwise).

4.2.10 Crossing the chasm: beyond early adopters

In general, when launching any type of technological product (software or hardware), it will ultimately undergo various stages during its lifecycle, if successful. A typical problem that arises during its lifecycle is **"crossing the chasm"**, which essentially refers to the inherent challenge of transitioning from the realm of niche market to mainstream market adoption.

As shown in Fig. 30, this chasm represents a significant gap between the *early adopters*, who are typically technology enthusiasts willing to experiment, and the *early majority*, who is distinctively characterized by a relatively more risk-averse and restrained disposition. The root cause of this issue lies precisely in the crucial difference between needs, preferences, and behavioural patterns of these two dissimilar consumer groups when compared to each other. While early adopters are amenable to risks and receptive to innovative solutions, the mainstream market expects a track record, proven solutions and some practical success stories. Consequently, strategies and tactics that might resonate well with early adopters may actually prove ineffectual in engaging the broader market audience. Products offered through blockchain are no different when it comes to crossing the chasm, which makes it a very real and practical consideration for the proposed Network.

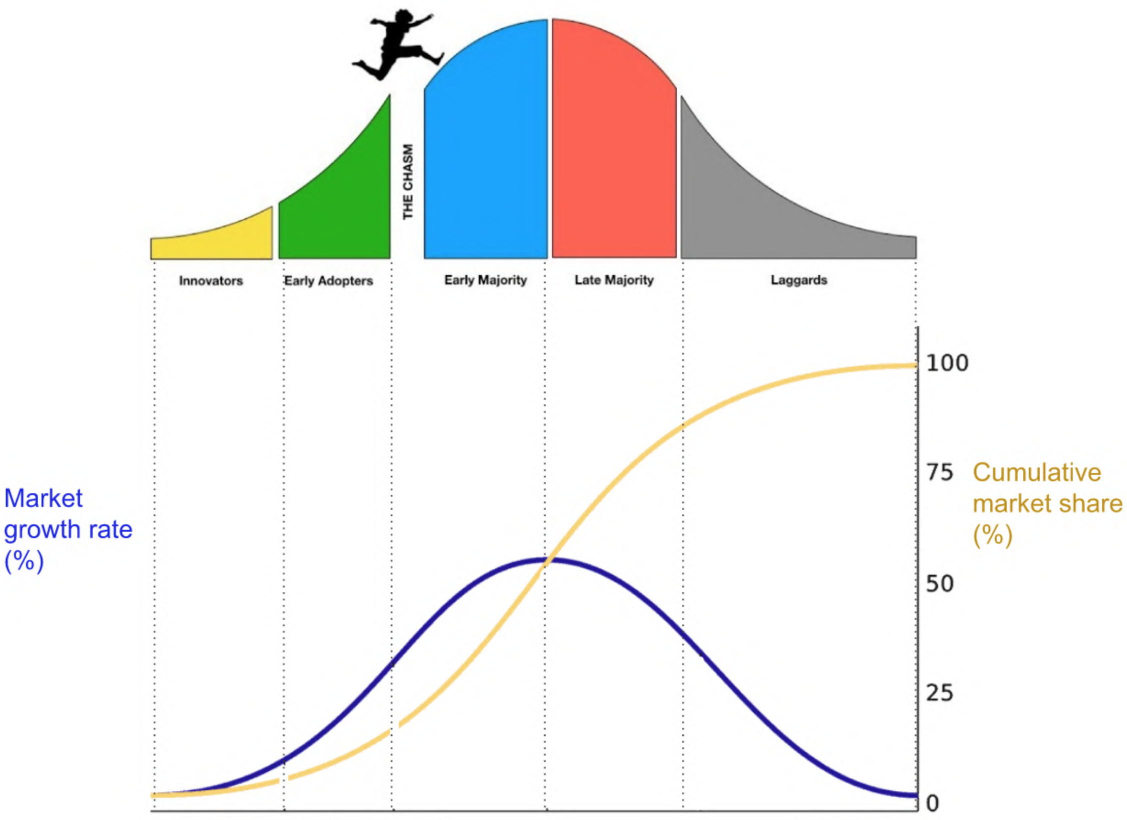


Figure 30 - Market growth rate (left axis) and cumulative market share (right axis).

In the context of both fungible and non-fungible tokens, a very common scenario is that of many early majority consumers struggling with seemingly “simple” onboarding tasks, such as how to purchase new tokens or how to register on dApps and dPlats. It is not rare in these situations for early adopters in the community to provide *pro bono* assistance to newer users. However, this is where the situation might become even worse, since the early majority consumers might additionally struggle with how to avoid being deceived by skilful impostors and fraudsters pretending to be acting in good faith.

Therefore, some generic *strategies* and *tactics* to deal with crossing the chasm in the case of the proposed Network include the following:

(1) Ensure **Fair Rewarding** (Section 3.1.3) of the community members that provide actual and meaningful assistance to new users both financially (using AGI or other fungible tokens) and in terms of recognition (using NFTs or otherwise).

(2) Enhance user experience by prioritizing **user-friendly** web front-ends (Section 2.2.5) and intuitive designs to ensure easy adoption and minimize friction.

(3) Provide **educational and training material**, such as step-by-step tutorials as blog posts and videos.

(4) Provide premium **Customer Support** to address any queries or concerns, thus improving customer experience and satisfaction, especially for those willing to pay for premium service in order to minimize hassle.

(5) Set up **partnerships** by collaborating with established industry players, strategic alliances or influential individuals at the local level.

Clearly, there will be additional strategies and tactics that are more *specific* to the nature of the Network and the digital products built on top of it. However, the aforementioned ones are emphasized in order to justify the rationale for requiring the existence of both fully decentralized and semi-decentralized solutions (as shown in Fig. 27).

Taking into consideration the *dilemma of user-friendliness vs. security* (Section 3.2.1), the scales should tip in favour of the former regarding most of the aforementioned strategies and tactics, thus requiring a *semi-decentralized* approach. As a result, the burden lies on the Foundation and other organizations in the Ecosystem (as shown in Fig. 29), since the proposed Network is envisaged to be implemented in a fully decentralized way.

4.3 Decentralized physical infrastructure: beyond software

Holistic solutions play a crucial role in decentralization by addressing the complex and interconnected challenges that arise in such systems. In particular, in order to fully realize the potential of decentralization, it is crucial to adopt a holistic approach that extends beyond software that also encompasses **decentralized physical infrastructure**. While software decentralization is essential, incorporating decentralized physical infrastructure adds a new dimension of resilience, censorship-resistance and autonomy. Decentralized physical infrastructure, such as *Distributed Smart Grids* (Section 4.3.3) and the *Decentralized Cloud Provider* (Section 4.3.4) enables self-sufficiency (i.e. autarky), reduces dependence on centralized authorities, and enhances local autonomy. It empowers communities (Section 4.3.2) by providing direct access to essential resources, fostering economic opportunities and promoting sustainability. Additionally, by incorporating physical infrastructure in decentralization efforts, potential limitations and challenges related to logistics, connectivity, and resource allocation can be addressed, while also providing *real-life experiences* (Section 3.2.3). Overall, there is a vast amount of reasons why decentralized physical infrastructure and DePINs are essential, as explained in Section 4.2.4. Therefore, holistic solutions recognize the *symbiotic relationship* between cyber and physical infrastructure, ensuring a comprehensive approach that unleashes the full potential of decentralization and drives societal transformation towards a more *resilient, self-sufficient* and *self-sustainable* future.

4.3.1 Self-sustainability

The word **'self-sustainability'** gets thrown around a lot by various organizations, ranging from globalist institutions (e.g., United Nations, WEF) to large multinational oligopolies (e.g., Big Tech, Big Food, Big Finance). However, the nature of self-sustainability is entirely *antithetical* to the control exercised by a colossal, bureaucratic centralized authority. Conversely, it necessitates following the *principle of subsidiarity*, which suggests that decisions and actions should be taken at the most local or decentralized level possible. In other words, local problems require local solutions.

Subsidiarity promotes local autonomy and fosters self-sustainability, while also cultivating a sense of ownership and accountability, since local communities and individuals become active participants in their own survival, development and growth. Hence, it encourages innovation, adaptability, and the efficient use of infrastructure resources (i.e. energy, compute, water or otherwise), as decisions are made based on local knowledge, expertise and context, instead of being made remotely by bureaucrats that are out of touch with local realities. The principle of subsidiarity is praised and supposed to be followed by several governmental bodies, such as the European Union, but either intentionally or unintentionally, it never seems to materialize in practise. In the United States of America, it is supposed to be embodied by the so-called 'state's rights', as stipulated in the Tenth Amendment of their constitution. It can even be traced back to (partial) decentralization of power to local administration in the Byzantine Empire (as explained in Section 4.1.7). Therefore, decentralization and self-sustainability are supposed to be compatible with and desirable by both central governments and multinational institutions. Additionally, there is also the case of *developing countries* or *rural areas* in developed countries that could benefit greatly by self-sustainability, making decentralized physical infrastructure a good candidate solution.

Two important desiderata of self-sustainability are: (a) having options to choose from, and (b) opt-in consent. Having a range of **options** allows the individual to exercise his autonomy and tailor his choices to align with his own values, preferences, and needs, while fostering a sense of independence, thus enabling community members to lead by example. Conversely, the lack of options demotivates the individual, limits creativity, restricts the ability to learn from mistakes, and ultimately increases the risk of centralization. Moreover, by having multiple options, individuals can have the freedom to explore innovative solutions, adapt to changing circumstances, and mitigate risks according to their own risk appetite, thereby enhancing their ability to sustain both themselves and their communities.

Additionally, **opt-in consent** complements the above-mentioned by emphasizing individual agency and ensuring that choices are made voluntarily, while acknowledging that self-sustainability is *not* a one-size-fits-all concept, and recognizes the different individual needs arising from the perspectives and circumstances that are specific to an individual. Opt-in consent also recognizes personal autonomy, self-ownership and self-determination, while promoting transparency, respect, and accountability by requiring explicit agreement before engaging in certain activities or accessing personal information. It also safeguards against coercion, exploitation, and unwanted external influences, enabling individuals to retain control over their own boundaries and promote innovation, growth and prosperity.

4.3.2 Decentralized communities

Furthermore, self-sustainability begins at the node (or vertex) of the graph (or network), i.e. the *household* of an individual or a family unit. As shown in Fig. 31, Decentralization assumes a level of connectivity of those nodes with edges (or links) to form a local community (or module), which is a formation inspired from that of the human brain [36].

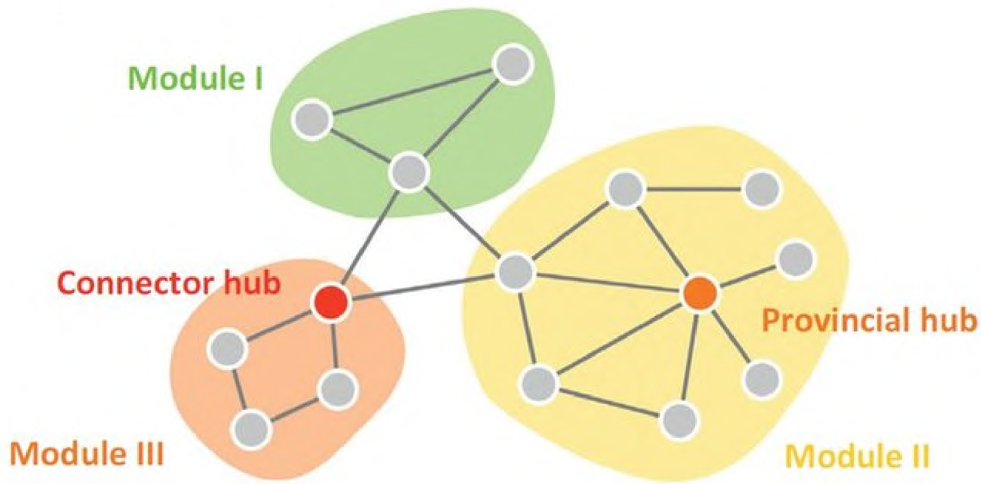


Figure 31 - Representation of hubs, and three modules inside a single modular network.

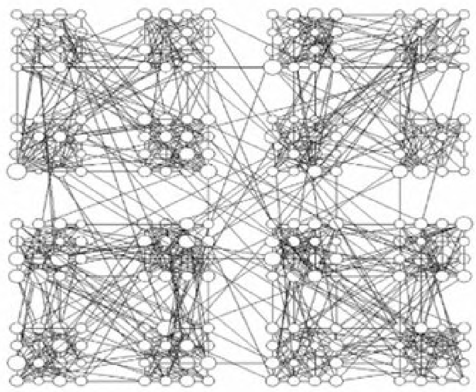


Figure 32 - An example of a hierarchical modular network (i.e. a module of modules).

Similarly, a lot of local modules can be connected together to form a module of modules, known as a *hierarchical modular* network [1], as shown in Fig. 32. Mathematical treatment of a modular network is outside the scope of this whitepaper, but the interested reader is referred to [36]. Each link in the network represents a connection of two household in terms of infrastructure resources (i.e. energy, compute, water or otherwise). In a **decentralized community (DeCom)**, households do *not* and should *not* need to trust any centralized authority or even a single blockchain network, otherwise the aforementioned desiderata of self-sustainability (i.e. having options and opt-in consent) are violated.

4.3.3 Distributed Smart Grids

In spite of the vast amount of theoretical work in the literature, *smart grids* and *microgrids* have not seen a considerable amount of adoption in real life. This is a very typical example of the scientific discovery being *problematic*, mainly due to the series of general reasons mentioned in Section 2.3, including: (a) the fact that such research is necessarily multidisciplinary but the research community (both in the industry and academia) exists in silos (Section 2.3.6), (b) the interests and incentives of the Researchers are misaligned (Section 2.2.3), (c) there is a lack of an infrastructure and a holistic mechanism for exploiting and commercializing the research findings (Section 2.3.9), (d) the lack of further funding for even more scientific breakthroughs that build on top of existing ones (Section 2.3.8), (e) the lack of a collaboration platform between Researchers, Developers and IDE startups and other organizations (Section 2.2), and (f) the lack of an ecosystem (Section 3.4.3) that addresses all reasons why startups fail (Section 2.2) and ultimately, drives the *cycle of innovation* (Fig. 14). Therefore, the Ecosystem, as proposed in this whitepaper, is uniquely positioned to address all of those issues in a frictionless and holistic way.

In addition, there are issues that are specific to smart grids and their lack of adoption. In particular, the adoption of such systems is typically imposed in a **top-down** and **one-size-fits-all** approach by a traditional, centrally managed authority, thus violating the desiderata of both *self-sustainability* (Section 4.3.1) and *good governance* (e.g., trustlessness, transparency, accountability, user participation, privacy, sovereignty), with such an approach inevitably creating a level of mistrust, vulnerability and inefficiencies.

Additionally, such approaches can often: (a) fail to consider the unique needs and challenges of local communities but also of the individual, (b) stifle innovation and creativity by discouraging experimentation, self-correction through feedback and learning from failures, (c) lead to homogenization, marginalization or loss of individual and community autonomy by disregarding their specific needs, preferences, and cultural differences, (d) cause a lack of flexibility due to inability to adapt to changing circumstances and evolving challenges or to respond to local emergencies, (e) create a significant information gap between the central authorities and local communities, leading to a lack of understanding of the on-the-ground realities and challenges faced by individuals or communities, and (f) face resistance from local communities or individuals who feel that decisions are being imposed upon them without their input or agreement, resulting in a lack of ownership and cooperation.

For instance, when it comes to energy sources and their impact on the environment, it is often the case that the conversation is monopolized by *doomsdayers* or “alarmists” (Section 3.3.10), who either knowingly or unknowingly neglect to consider the whole *lifecycle* of such a source or fail to see the *overall* picture (e.g., social aspects, political aspects, economic aspects, national security aspects). These kinds of doomsdayers are found in many fields, including AI and Blockchain. As extensively explained in Section 3.3.10, the root cause of this phenomenon generally stems from their lack of education, which creates *irrational fears*, and usually leads to a form of *religious sect or cult*. Therefore, their actions and behaviour — either through “activism” or through imposing their views on others — can become a worse problem than the perceived “problem” that they are seemingly trying to solve, since it reinforces a sense of *obscurantism* (Section 2.3.2). When such a behaviour is combined with a top-down and one-size-fits-all approach by a traditional, centrally managed authority, it typically becomes a recipe for disaster. It should be made clear that it is not the intention of the authors to criticize or support any specific hypothesis or proposition, but instead, to merely point out the existence of such an undesirable situation, no matter the underlying hypothesis or proposition.

Therefore, the authors propose the use of **Distributed Smart Grids**, which can be achieved by applying the *distributed paradigm* (Fig. 17) in order to transform the century-old design of the existing electrical grid. In other words, it applies the paradigm behind Blockchain to the *physical infrastructure* (i.e. hardware) behind the electrical grid, but taking into consideration the peculiarities of electricity, and adapting the paradigm accordingly. Additionally, the proposed Distributed Smart Grids incorporate an *information infrastructure* (i.e. software) in order to facilitate a two-way communication between node-to-node and node-to-module. Such proposals are not new, but they have failed so far to implemented for the above-mentioned reasons, and this is what the proposed Ecosystem attempts to address.

It should be clarified that the Distributed Smart Grids proposed in this whitepaper assume a *residential* or *commercial* consumer, not an industrial customer or otherwise, as the **target audience**. The reason is that the intended areas to be implemented are residential or commercial ones, *not* industrial zones. This does not necessarily mean that principles of decentralization cannot be implemented at industrial zones too, however, they remain outside the scope of this whitepaper. Even though not restricted to them, *developing countries* or *rural areas* in developed countries could benefit greatly from the proposed Distributed Smart Grids. As a matter of fact, it is sometimes *not* economically viable for a centralized authority to even operate in such areas, and also these areas might suffer from frequent *outages*, which often makes decentralization the only viable option.

More specifically, the authors propose the use of multiple **grid-connected microgrids** within a distinct area, such as a *local community* within a neighbourhood (e.g., multiple homes, small businesses and public facilities) or a *campus* of institutional facilities (e.g., university, hospital, tourist resort, large corporation). Each such distinct area represents a *module* of a hierarchical modular network (Section 4.3.2).

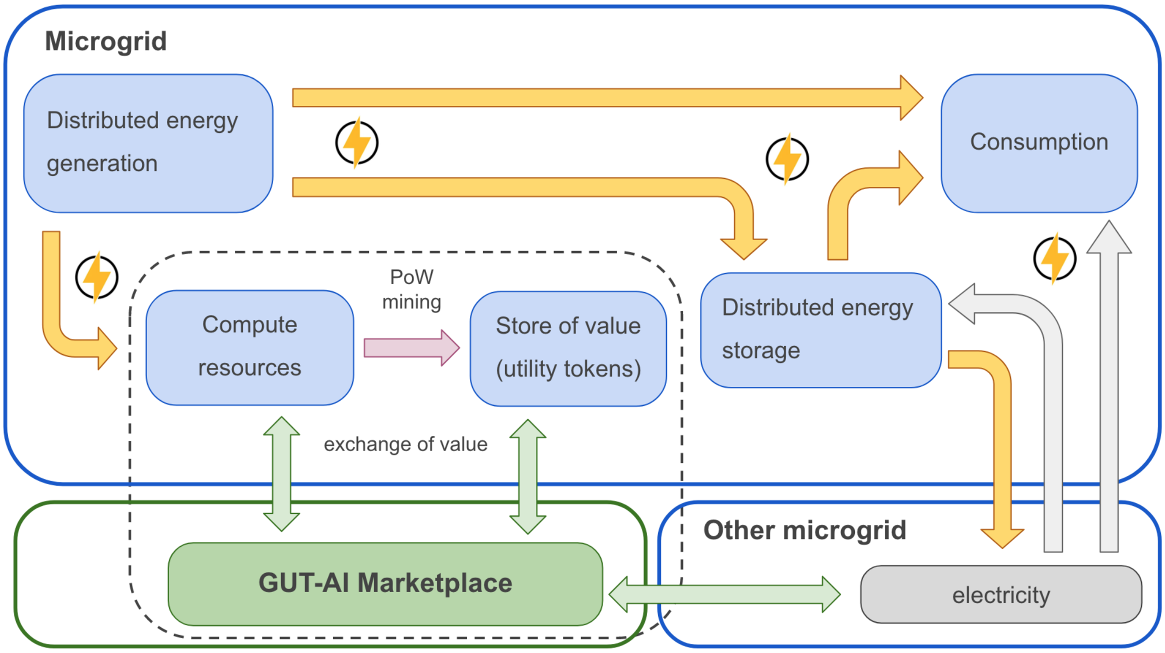


Figure 33 - Block diagram of the proposed microgrid within a decentralized community (i.e. module).

In general, for a system to be considered a *microgrid*, it typically needs to meet certain requirements. The one proposed in this whitepaper (Fig. 33) requires the following:

(1) **Distributed energy generation:** In order to reduce reliance on resources far away from the local community, it is important to have the ability to locally generate electricity at or close to the point of use by *integrating* multiple Distributed Energy Resources (DERs), such as solar panels, solar roof tiles, wind turbines, biomass generators or even small-scale conventional generators for backup generation or price hedging. The specific *energy mix* should be chosen by the individual or the community, and should be adjustable to account for weather changes or energy market conditions.

(2) **Distributed energy storage:** In order to balance supply and demand, store excess energy for later use and provide backup power during disruptions, it is crucial to incorporate energy storage systems such as batteries, supercapacitors or hybrid systems to get the best of both worlds.

(3) **Connection and disconnection capability:** A microgrid should have the ability to connect to and disconnect from the main grid (or other microgrids) in order to enable the former to operate autonomously in "island mode" (i.e. standalone) or interact with the main grid (or other microgrids) when necessary, such as importing or exporting excess energy.

(4) **Information and communication system:** In order to ensure the efficient and effective management of electricity generation, storage, distribution, transactions and consumption, it is essential to integrate an (open-source) *information and communication system* into the electric power system. This system can facilitate real-time monitoring, control, and optimal coordination of the various components within the physical infrastructure, while enabling seamless connection to and disconnection from the main grid, balancing supply and demand, responding to disruptions or outages, and active participation in the Marketplace (Section 4.4) and the Decentralized Exchange (Section 4.4.4) of utility tokens, which is part of it. Among other factors, due to such a tradability on cryptocurrency exchanges, a utility token constitutes a *store of value*, since it can retain its purchasing power over time rather than depreciating.

(5) **Compute resources:** In order to contribute to the creation of the Decentralized Cloud Provider (Section 4.3.3), it is helpful to integrate a set of processing power, memory, networking, data storage and other resources required for performing calculations, executing computations and processing data. All such compute resources also allow the smooth functioning of both the node and the microgrid.

(6) **Scalability, extensibility and flexibility:** In order to enable the microgrid to adapt to evolving energy needs and technologies, it should be designed to be scalable, extensible and flexible, allowing for easy expansion or reduction in capacity based on changing energy demands or resource availability.

It is worthy to note that Demand-Side Management (DSM) is deliberately *not* included in the list above because: (a) such approaches are characteristic of a *centralized* mentality (i.e. top-down, one-size-fits-all), (b) in contrast to traditional electrical grids, the burden of energy generation falls predominantly to the consumer (who becomes a *prosumer*) or the local community, thus making large-scale DSM less relevant in the case of a microgrid of residential and commercial consumers, and (c) they typically require modification of consumer patterns and behaviour in order to adapt to technology (i.e. machine-centred), which is somewhat unrealistic, counter-intuitive and also *antithetical* to **user-friendliness**, since such a modification can have a negative impact on user experience.

Therefore, Distributed Smart Grids aim to maximize the *consumer's convenience* and the efficiency of the local microgrid by being **human-centred**, instead of maximizing the penetration of specific technologies or the mitigation of peak demand. Nonetheless, this does not preclude techniques such as Dynamic Pricing, but it does not force them either. Their use remains at the discretion of the individual or the local community, and is to be decided through a local *governance system* (Section 4.1.3).

In addition to the typical benefits of a conventional microgrid, the proposed Distributed Smart Grids have the following *additional advantages*:

- It also allows interaction with other microgrids (Fig. 35) when necessary (such as when importing or exporting excess energy) instead of just with the main grid. This adds an extra layer of *resilience, reliability* and *autonomy* by reducing centralization and single points of failure, since reliance on the main grid is decreased.
- Each node (e.g., household) has the *potential*, but not the requirement, to also operate autonomously in "island mode" (i.e. standalone), thus maximizing freedom, innovation and creativity for the individual (Fig. 35).

- Each node in the decentralized community (i.e. module) does *not* have to be a residential or commercial building, but it can also be a solar farm, wind farm, local battery storage site or small data centre, as shown in Fig. 35. Such farms can be fully private, fully communal or privately managed, depending on the needs of the local community. These types of decentralized farms not only enhance self-sustainability, but also ensure that Web is treated as a *utility* (Section 3.1.4), similar to electricity.
- AGI and other utility tokens (Fig. 29) can enable direct *P2P energy trading* within microgrids. Nodes can buy and sell excess energy without the need for financial intermediaries, negotiate prices, and execute frictionless and cost-effective transactions in a secure and automated manner.
- Crypto lending platforms in DeFi can facilitate *access to capital* to individuals and local communities who may not have access to traditional banking services or lines of credit. Nodes can borrow funds against their digital assets or use smart contracts to access credit in order to cover the initial cost of installing a microgrid or to finance projects related to the maintenance and expansion of microgrid infrastructure. As a result, the proposed Distributed Smart Grids can provide energy access to underserved communities or to remote locations without access to a reliable energy grid (especially in developing and low-income countries) by making the operation of electrical grids in such areas more economically viable.
- Blockchain technology allows for the *fractional ownership* of assets or projects. In the context of microgrids, this means that individuals or organizations can invest in and own a portion of the microgrid infrastructure, thus creating new *investment opportunities*. Investors can hold tokens representing a share of the project, thus providing them with financial incentives and economic benefits (e.g, voting rights, profit-sharing) as a microgrid project generates revenue, while also promoting transparency, auditability and efficiency regarding the project's finances.

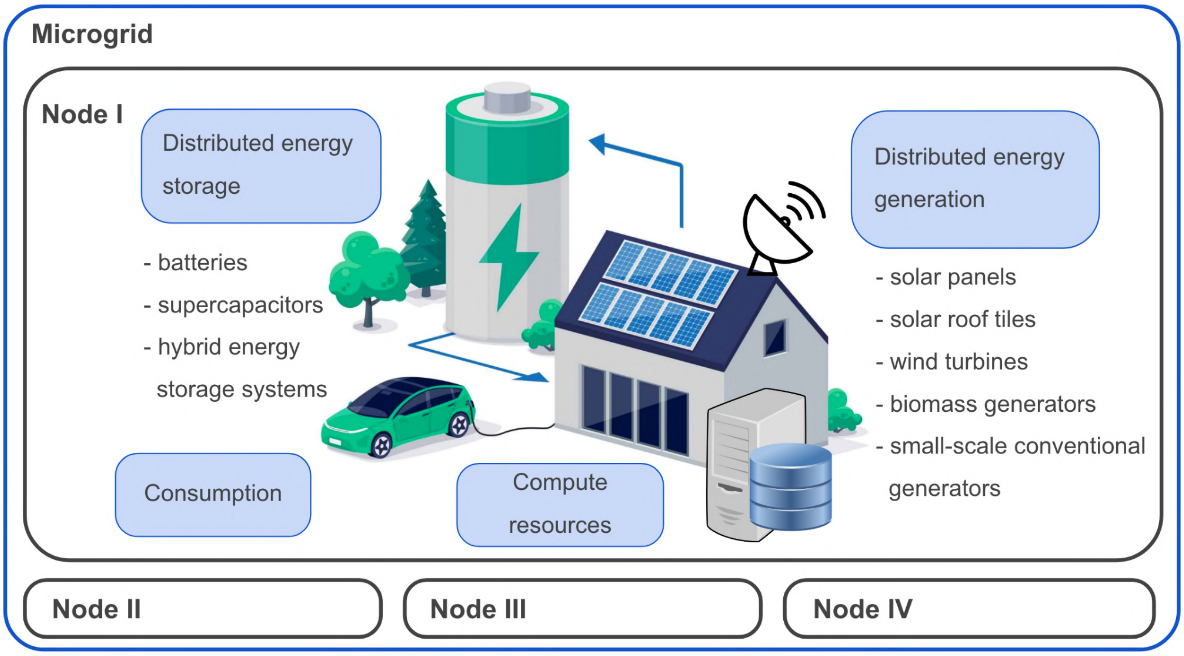


Figure 34 - Pictorial representation of a single node within a microgrid.

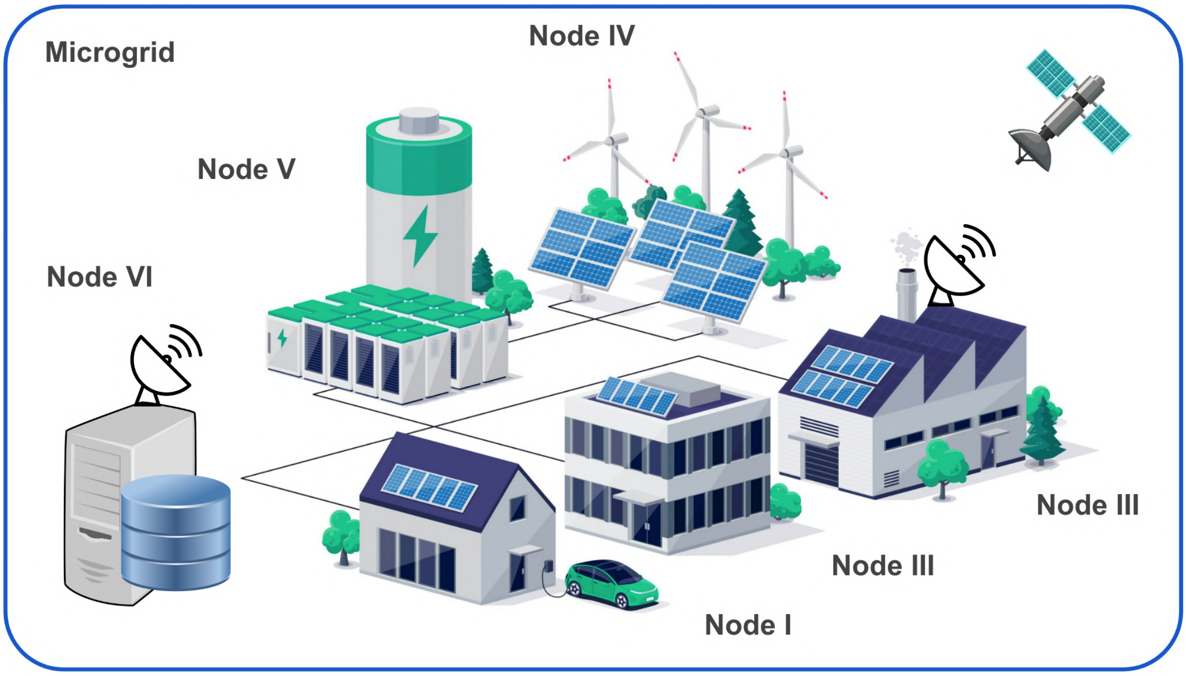


Figure 35 - Pictorial representation of multiple nodes within a decentralized community (i.e. module).

- Price discovery mechanisms in DeFi and the proposed Decentralized Exchange (Section 4.4.4) can facilitate market-driven Dynamic Pricing for energy within microgrids. Nodes can engage in transparent and fair price negotiations through utility tokens, thereby promoting arbitrage and competitiveness.
- Smart contracts can automatically calculate and execute *billing* and *settlement* processes by streamlining self-executing payments based on predefined rules and conditions. This eliminates the need for manual meter reading, invoice generation and other manual administrative tasks, thus simplifying financial operations. This level of automation minimizes human error, financial intermediaries and associated transaction fees, while also promoting transparency and trustworthiness in the energy system.
- Optional Open Standards (Section 4.2.3) and smart contracts can be designed to enable *interoperability* with other systems (whether physical or otherwise), thus ensuring seamless integration with other nodes, other microgrids, the proposed Marketplace (Section 4.4) and any type of physical hardware through DePINs. This interoperability allows for efficient energy sharing, collaboration between nodes, and exchange of know-how and expertise between various communities.
- AI solutions (Section 2.1.1) offered through the proposed Marketplace can assist in simulation, design, deployment, maintenance, real-time monitoring, control and optimal coordination of the microgrid infrastructure, while addressing fault detection, anomaly detection and energy theft detection. They can also be used to predict energy demand and supply in the future given historical energy data, weather patterns, and other relevant datasets. Moreover, they can be used to optimize energy generation, storage, and usage schedules within microgrids in order to achieve cost-effective and sustainable energy management, while also providing customized recommendations, incentives and feedback to consumers.

- The overall Ecosystem enables faster deployment and scalability of microgrids. The *modular* nature of microgrids and the *decentralized* nature of Blockchain allow for easier integration, customization and expansion as the energy needs grow or change over time.

In addition to the aforementioned advantages that the Ecosystem brings to microgrid infrastructure, the microgrid infrastructure also brings benefits to the Ecosystem itself. One primary benefit is ensuring censorship-resistance right at the level of physical infrastructure by recognizing the *symbiotic relationship* between cyber and physical infrastructure.

4.3.4 Decentralized Cloud Provider

No matter the underlying energy infrastructure, a **Decentralized Cloud Provider** can be implemented either using the compute resources inside the Distributed Smart Grids (Fig. 34) or using large *data centres* (Fig. 16) and multiple *centralized* cloud providers. To implement this architecture, the Decentralized Cloud Provider will act as an intermediary between users and the underlying infrastructure (Fig. 37). Users would interact with the decentralized provider's interface, which would handle the complexities of managing multiple cloud providers and data centres behind the scenes. The Decentralized Cloud Provider will dynamically allocate resources based on user demands and available capacity.

In terms of **data storage**, in order to maintain data integrity and security, techniques such as *encryption* and *Distributed Systems* can be employed . Users' data can be distributed across multiple cloud providers, enhancing redundancy and mitigating the risk of data loss (Fig. 36). Additionally, *decentralized* consensus mechanisms, like the one in AGI token (Section 4.2.4), will be utilized to ensure transparency and trustworthiness in the allocation of resources.

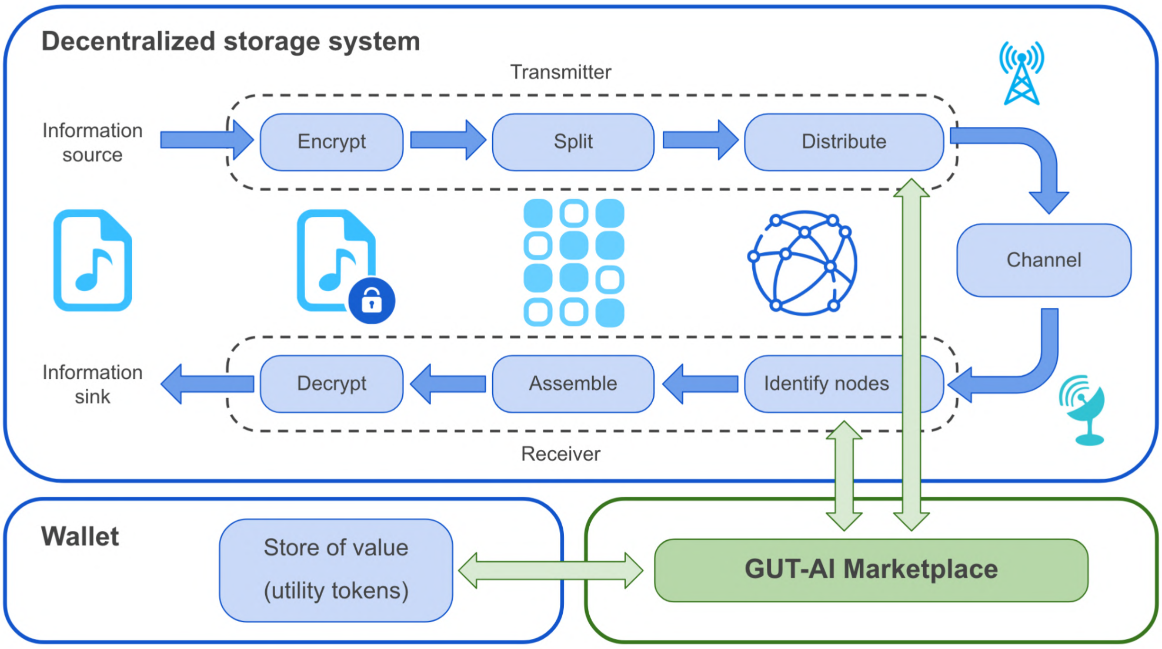


Figure 36 - Block diagram of a decentralized storage system.

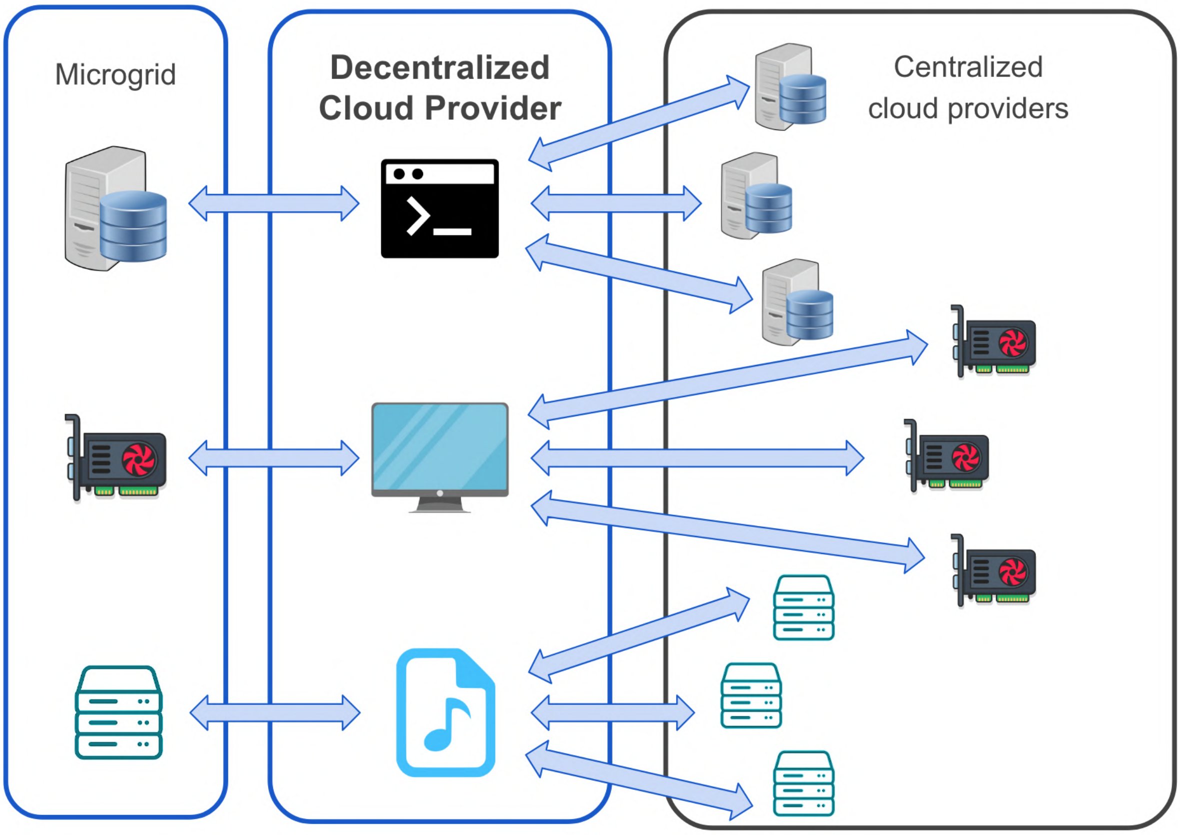


Figure 37 - Pictorial representation of the proposed Decentralized Cloud Provider.

In addition to data storage (e.g., file storage, object storage, block storage), the Decentralized Cloud Provider should also be able to provide *processing power, memory, databases* and *networking*. A Distributed Computing approach can also be employed in this case too. More specifically, users will submit their tasks or applications to the Decentralized Cloud Provider, which will then distribute the workload across multiple centralized cloud providers and data centres based on their availability and capacity. The Decentralized Cloud Provider would utilize load balancing algorithms to evenly distribute the computational tasks among the available resources. Similarly, memory requirements can be distributed by intelligently and strategically allocating data across the distributed memory systems available in different data centres or traditional centralized cloud providers.

Some early attempts in decentralized cloud infrastructure have been made by some DePINs, especially in the field of data storage, and the Decentralized Cloud Provider can either make direct use of them or build on top of them.

Overall, the proposed Decentralized Cloud Provider can offer multiple *benefits*, including:

- Multiple centralized cloud providers can be used in such a way in order to *avoid vendor lock-in*, while also promoting healthy competition by allowing DePINs to choose and adapt their cloud infrastructure as their needs evolve, since they have the freedom to choose and easily switch between multiple centralized providers based on their specific requirements, pricing models or service offerings.
- The risk of a single point of failure or data loss is reduced due to enhanced *data security* as the data is distributed across multiple providers and data centres. Also, data can be encrypted and stored redundantly, making it more *resilient* against breaches and unauthorized access. Hence, users benefit from increased privacy, improved data control, stronger *censorship-resistance* and enhanced security by ensuring that data are not solely stored or controlled by a single entity.

- *High availability* and *reliability* can be ensured by leveraging multiple centralized cloud providers and large data centres. If one such a provider or data centre experiences downtime or technical issues, then dApps, dPlats and data can seamlessly be shifted to alternative resources, minimizing disruption and downtime.
- Large data centres can be chosen strategically to be distributed geographically in such a way that they minimize *latency*, while enhancing overall *performance* and *scalability* of the Decentralized Cloud Provider. This enables Developers to scale their applications and handle increasing workloads more efficiently, while allowing users to access compute resources closer to their location.
- The competitive landscape of multiple centralized cloud providers can be leverages in order to achieve potential cost savings for both Developers and users. For instance, Developers can select the most competitive providers for their specific workloads or applications, taking advantage of varying pricing structures, discounts and customized plans, resulting in more cost-effective cloud services. In addition, the ability to share and scale resources dynamically across various centralized providers enables efficient resource utilization, optimizing costs by only paying for what is needed, and by minimizing the risk for paying for resources that remain idle.

Overall, similar to the Distributed Smart Grids (Section 4.3.3), the Distributed Cloud Provider proposed in this whitepaper recognizes the symbiotic relationship between cyber and physical infrastructure, while also brings benefits to the Ecosystem itself, such as censorship-resistance right at the level of physical infrastructure. It also provides new investment opportunities, while promoting innovation and growth by increase the efficiency of the relevant part of the *cycle of innovation* (Fig. 14) that relates to the commercialization of research findings. However, physical infrastructure is a necessary, but not sufficient condition for such an innovation to occur. This gives rise to the need for a Marketplace (Section 4.4).

4.3.5 Portable devices

Portable devices play a particularly crucial role in a decentralized future due to their ability to empower individuals, foster connectivity, and facilitate information access. In such a future, decentralized systems and networks are likely to become more prevalent, distributing electrical energy, information access and control as close as possible to the *level of the individual*, while also having a positive contribution to the *decentralized community* (as defined in Section 4.3.2) that the individual is travelling to or is passing by.

Conventional portable devices — such as smartphones, laptops and tablets — act as the primary gateways for individuals to engage with these decentralized communities. Such portable devices can enable individuals to have greater autonomy and control over their own data and digital identities. By carrying their devices with them, people can securely access and manage their personal information *offline*, thus reducing the need for any intermediaries outside the decentralized community (Fig. 35) or outside their own household (Fig. 34). This ensures *privacy* and empowers individuals to have ownership over their digital presence. It is entirely possible to have mini-versions of Internet Service Providers (ISPs) right at the level of the community, thus treating the Web as a *utility* (Section 3.1.4). Furthermore, these conventional portable devices provide individuals with instant access to a wealth of knowledge, resources and services, while also providing connectivity and collaboration among multiple decentralized communities. This democratization of information empowers communities (i.e. modules) to make better-informed decisions, while engaging and mutually learning together with various other decentralized communities, thereby fostering a *hierarchical modular network* (Fig. 32), which is collectively more knowledgeable, more well-informed and more self-aware, thus enhancing the *resilience* of such a network.

In addition to the aforementioned advantages, **next-generation portable devices** can also bring unique benefits and capabilities to the individual in the context of a decentralized future, while enhancing user experience and empowering the user participation in decentralized communities. In the context of real-life experiences (Section 3.2.3), some examples of such devices include smart power banks (as described below), wearable devices (e.g., smartwatches), Augmented Reality (AR) and Mixed Reality (MR) devices (Section 4.7.3), as shown in Fig. 44.

The overall purpose of **smart power banks** (as proposed in this whitepaper) is to enable a microcosm of Decentralized Smart Grids (Section 4.3.3) down to the level of the individual, while also positively contributing to his own household (Fig. 32) by providing advanced functionalities beyond traditional charging capabilities. This way, smart power banks can synergistically contribute to a higher level of decentralization, not just for the individual, but also for the whole community that he is travelling to or is passing by.

For instance, they could integrate Blockchain technology to enable P2P energy sharing and *decentralized energy trading*. This way, individuals can store any excess energy generated from Distributed Energy Resources (Section 4.3.3) in their power banks and trade it with others in the decentralized Marketplace (Section 4.4). Additionally, it is technologically doable and feasible to also incorporate functionalities of *data storage*, similar to that of a memory stick. For example, they could incorporate built-in data storage protocols that enable individuals to participate in decentralized storage systems (Fig. 36), thus making P2P storage networks a reality in the physical world, i.e. outside the cyberspace. By using their smart power banks as part of a decentralized network in the physician world, users can exchange their unused energy and data storage in order to earn incentives, while supporting a more resilient and censorship-resistant physical infrastructure.

4.4 Marketplace

Among other abilities, the **GUT-AI Marketplace** (Fig. 38) allows for the significant expansion of the functionality and capabilities of the software through the integration of third-party applications and services. Therefore, it provides users with a *one-stop shop* to discover, evaluate, and access a wide range of additional features, tools, solutions, plugins, extensions, add-ons and customized themes that complement the core products and services. Hence, implementing the Marketplace alongside some free and open-source software (Section 4.1.10) is crucial for several *reasons*, including the following:

(1) **Increased user adoption:** Due to the aforementioned expansion ability, the proposed Marketplace empowers users to *customize* and *extend* the software according to their specific needs, thus making the software more versatile and appealing to a wider range of users, while driving user adoption. Additionally, the Marketplace can facilitate the translation and *localization* of applications and plugins across different regions and languages, enabling the software to reach a global audience and cater to the specific localization requirements of each user.

(2) **Enhanced user experience:** Since users can customize the software in order to address their specific needs and preferences, this leads to an improved and more tailored *user experience*, ultimately increasing its relevance, value and usability. This constant stream of new offerings keeps the Ecosystem dynamic, since it allows the software to scale and adapt according to ever-evolving user demands. As the range of available applications and services expands, the Marketplace grows organically, providing users with a scalable Ecosystem that can accommodate their changing needs in an agile manner. Overall, users can benefit from a wide variety of third-party applications that allow *seamless deployment and integration* (Section 2.2.6), resulting in an intuitive user experience.

(3) **Accelerated innovation:** The Marketplace fosters an environment of innovation and collaboration, since it can tap into a broader pool of talent and ideas by opening up the overall Ecosystem (as defined in Section 4.2) to external Developers and various organizations (Fig. 29) beyond the GUT-AI Foundation (Section 4.1). This enables sharing of ideas, building upon each other's work, and the *rapid* development of new functionalities, extensions and integrations, thus promoting *healthy competition*, while also resulting in an enriched user experience and a faster time to market. As a result, the Marketplace becomes a hub for innovation, enabling continuous improvements and cutting-edge advancements in the software, while also encouraging ongoing development and support for your software. Sellers (Fig. 38) are motivated to regularly update their applications and services in order to: (a) meet Buyer's demands (b) ensure compatibility with the latest software version, and (c) maintain compatibility with the latest software versions, thereby minimizing disruption, facilitating a seamless transition, and providing a smooth upgrade experience for the user. This ultimately leads towards a vibrant, fast-paced and thriving Ecosystem in order to meet the challenges of today and anticipate the needs of tomorrow.

(4) **Monetization opportunities:** The Marketplace can serve as an additional stream of revenue for both the software itself and the proposed Network (Section 4.2.4) in general. By facilitating transactions between Buyers and Sellers (e.g., Developers, Researchers or organizations), who offer their applications and services through the Marketplace, a revenue model through *transaction fees* or *revenue sharing* can be established. This paves the way for new monetization opportunities for both the Network and the Sellers, thus creating a mutually beneficial Ecosystem, while resulting in a win-win situation. Moreover, such a vibrant Marketplace attracts more Sellers, expanding the range of product offerings and potential sources of revenue, while also attracting more Buyers, thus creating *network effects*.

(5) **Viability, sustainability and longevity:** Through the aforementioned monetization opportunities, the Marketplace encourages Sellers to invest more time and effort into enhancing the software, thereby providing multiple incentives for its long-term viability, sustainability and expansion by essentially *aligning the interests* (Section 2.2.3) of the Sellers and the Foundation, which offers the software (Section 4.1.10). For instance, it incentivizes its ongoing development and support, leading to continuous improvement and sustained innovation. Therefore, it motivates Sellers to continue investing in providing applications and services that complement the software, thereby eventually contributing to its longevity and persistent growth. In addition, the Marketplace allows for *vertical expansion* by enabling Developers to offer *industry-specific* applications and services that extend and adapt the software's functionality in order to target specific domains or sectors, thus broadening the appeal of the software to a vast array of industries and also boosting customer satisfaction. This enforces a level of *separation of concerns* (Section 2.2.1) between the core software and the applications that complement it, while also mitigating the risk of replication of work (Section 2.2.2) and delays on product roadmaps (Section 2.2.4). As a consequence, the core software can focus on back-end functionality and business logic, while integrations, extensions and themes can focus on offering web front-ends that are more *user-friendly* (Section 2.2.5) to the target audience.

(6) **Discoverability:** The Marketplace also adds value to the whole Ecosystem by acting as a decentralized *common point of reference* for users. It offers a convenient one-stop-shop for users to discover, evaluate and install software. It simplifies the process of exploring, finding and integrating complementary applications and services, thereby enhancing the overall *discoverability* of the software, while increasing *visibility* and encouraging the adoption of core software. It also ensures compatibility, while reducing the time and effort required for receiving updates.

(7) **Community engagement and support**: Apart from the Buyers, the Marketplace serves a decentralized *common point of reference* for Sellers (e.g., Developers, Researchers or organizations) too. For instance, it provides an avenue for greater visibility and cross-promotion by: (a) allowing Developers to distribute, share and showcase their product offerings to a wider audience, gain recognition and attract potential contributors, (b) allowing Researchers to disseminate and commercialize their research findings (Section 2.3.9), and (c) driving cycle of innovation (Fig. 14) through the mutually beneficial engagement and interaction between Developers and Researchers under a single *fully decentralized* roof. Furthermore, it allows both Sellers and Buyers to collaborate, coordinate and work jointly on various projects via a *common platform* (Section 2.2.8) by facilitating efficient sharing of knowledge, resources and expertise, thus fostering a supportive *community* around the software. By creating a sense of community, users can access tutorials, forums and messaging dApps to seek assistance or support, share insights or collaborate with their peers. This *community-driven* support network eliminates the need for users to rely on a centralized support team, while ensuring that users will receive the necessary guidance and assistance by their peers, who might have faced a similar situation or problem and hence, they will probably be more well-suited to help, educate or inform their fellow community members. In general, this level of *community engagement* and *support* built around the Marketplace contributes to customer satisfaction, loyalty and the overall success of the whole Ecosystem, while ultimately leading to a win-win situation for every type of participant in it. In other words, “The whole is greater than the sum of its parts”, an expression attributed to Aristotle. This emphasizes the concept that the collective power, strength and abilities of the community as a unity create a synergy that transcends their individual capabilities by amplifying the impact and accomplishments of the community as a whole.

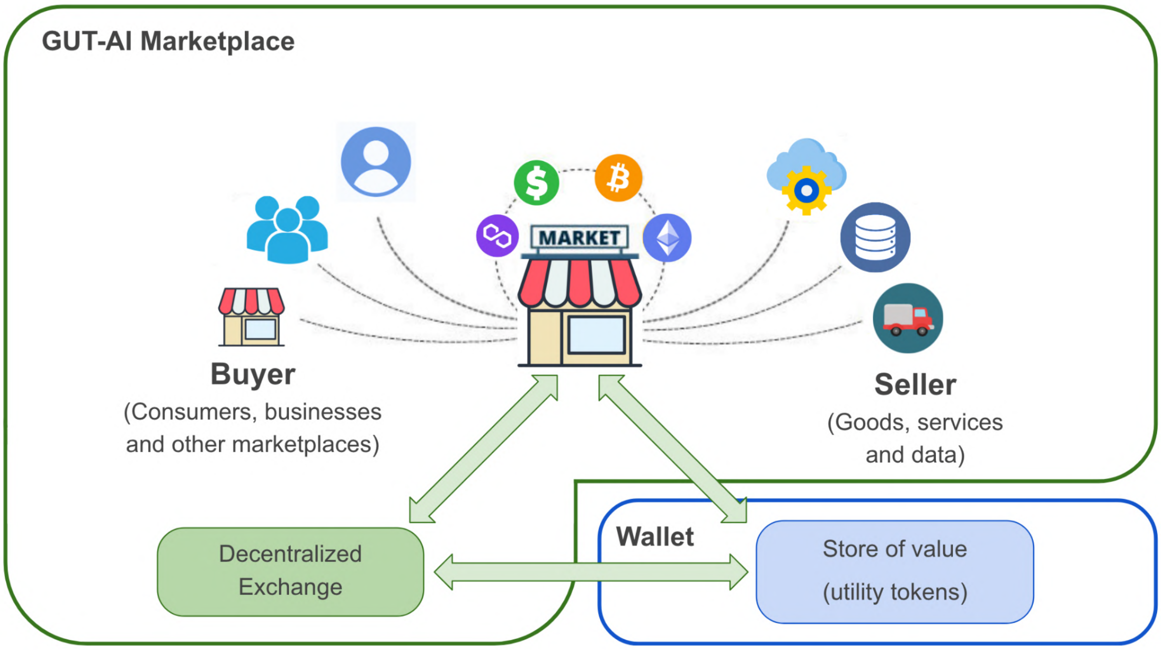


Figure 38 - Block diagram of the proposed Marketplace.

(8) **Feedback and quality control:** In addition to cultivating a sense of belonging and empowerment, the aforementioned community engagement can further enhance the overall user experience through *feedback* and *quality control*. More specifically, users can be incentivized through Fair Rewarding (Section 3.1.3) in order to leave reviews, feedback and suggestions for improvements *directly* to Developers, while sharing their experiences with using various applications and services in order to help other users make better-informed decisions. Reciprocally, Developers can interact back with users, gather feedback, and address their needs, thus leading to a stronger user-developer relationship. This leads to the development of a vibrant Ecosystem of Developers, Designers and users, which can result in the creation of innovative solutions and continuous improvement of the software, while essentially creating a form of *self-correction mechanism* (Section 2.2.7) for the Marketplace itself. Moreover, by curating and vetting applications and services within the Marketplace, a certain level of *quality* and compatibility can be ensured. Implementing a reputation

and recommendation system (Section 4.4.5), which includes review processes, guidelines and Open Optional Standards (Section 4.2.3), helps maintain an increased level of *transparency*, leading to a high standard of performance, efficiency and security by reducing the risk of malicious code. This *feedback-driven* approach cultivates trustworthiness, reliability and confidence in the Marketplace offerings, leading to a better user experience for everyone involved, while also helping to shape the future of the overall Ecosystem itself.

Therefore, due to reasons listed above, the Marketplace facilitates the following:

- Transactions of data (Section 4.4.1)
- Transactions of good (Section 4.4.2)
- Transactions of services (Section 4.4.3)
- Decentralized Exchange (Section 4.4.4)

4.4.1 Transactions of data

Due to the reasons mentioned in the beginning of Section 4.4, offering **transactions of data** (e.g., datasets, pretrained models, checkpoints, simulators, metadata) within the proposed Marketplace contributes towards: (a) reducing the risk for startups failing (Section 2.2), (b) addressing the impediments towards scientific discoveries and technological breakthroughs (Section 2.3), and (c) increasing funding for further research (Section 2.3.8). This ultimately leads to maximizing the overall efficiency, effectiveness and performance of all the relevant parts of the *cycle of innovation* (Fig. 14), which is the key operational mechanism of the proposed Ecosystem (Section 4.2) in order to address the needs and preferences of the users (Section 2.1), while the Foundation (Section 4.1) is a mere participant in the both the Marketplace and the Ecosystem in order to catalyse innovation, growth and prosperity.

While the reason for offering transactions of data is crucial, *implementation* also matters. More specifically, *datasets* should be offered in a manner that promotes accessibility, innovation, collaboration and transparency. A few specific examples of implementation include the following:

(1) **Open and standardized formats:** Datasets should be made available in open and standardized formats (Section 4.2.3) that are widely supported. For instance, using common formats, such as CSV, XML or JSON, allows *easy integration* with various AI frameworks while promoting *accessibility* and *interoperability*.

(2) **Clear licensing and usage terms:** Dataset providers should clearly define the licensing and usage terms for their data. For an open-source Marketplace, Creative Commons licenses or similar open licenses can be employed to ensure that datasets can be freely used, modified and shared by the community.

(3) **Version control and documentation:** Datasets should be accompanied by version control systems (e.g., DVC) and comprehensive documentation. Enabling dataset versioning and change tracking systems allows users to access previous versions, track modifications, and understand the evolution of datasets over time. This also allows users to contribute improvements and understand the data's characteristics, thus facilitating transparency and reproducibility.

(4) **Fair Rewarding and other incentives for data providers:** Introducing incentives for dataset creators can promote viability and economic sustainability. For instance, the Marketplaces can implement a revenue-sharing mechanism, where *data providers* receive a percentage of the proceeds (Section 3.1.3) generated from the construction of the dataset, encouraging the continuous production of high-quality datasets.

(5) **Industry-specific datasets:** Curating datasets tailored to specific domains (e.g., healthcare, finance, or autonomous vehicles) caters to the unique needs of various industries, fostering collaboration and innovation in those specific areas.

4.4.2 Transactions of goods

Due to the reasons mentioned in the beginning of Section 4.4, offering **transactions of goods** (e.g., IoT devices, battery storage systems, smart power banks, solar panels, solar roof tiles) within the proposed Marketplace contributes towards: (a) reducing the risk for startups failing (Section 2.2), (b) addressing the impediments towards scientific discoveries and technological breakthroughs (Section 2.3), and (c) increasing funding for further research (Section 2.3.8). This ultimately leads to maximizing the overall efficiency, effectiveness and performance of all the relevant parts of the *cycle of innovation* (Fig. 14), which is the key operational mechanism of the proposed Ecosystem (Section 4.2) in order to address the needs and preferences of the users (Section 2.1), while the Foundation (Section 4.1) is a mere participant in the both the Marketplace and the Ecosystem in order to catalyse innovation, growth and prosperity.

While the reason for offering transactions of goods is crucial, **implementation** also matters. More specifically, *goods* should be offered in a manner that promotes accessibility, innovation, collaboration and transparency. A few specific examples of implementation include the following:

(1) **Standardized components:** Using Optional Open Standards (Section 4.2.3), open-source component lists should be created, in which Developers and Researchers can all access and also incorporate standardized components into their designs, thereby promoting *compatibility, interoperability* and *ease of integration*.

(2) **Firmware and hardware repository:** A repository for open-source *firmware* and accompanying software related to the respective goods should be established, thus allowing Developers to enhance the functionality and performance of the hardware products through firmware updates.

(3) **Collaborative prototyping:** Collaborative prototyping should be facilitated by connecting Developers with complementary skills and expertise, allowing them to collaborate on creating and refining prototypes for new hardware products.

(4) **Crowdfunding support:** Crowdfunding capabilities should be integrated into the Marketplace, thus enabling Developers to showcase their projects and raise funds to bring their open-source hardware designs to fruition.

(5) **Fair Rewarding and other incentives for hardware designers**: Introducing incentives for Developers to contribute their innovative designs, improvements or updates can promote viability and economic sustainability. For instance, the Marketplaces can implement a revenue-sharing mechanism, where *hardware designers* receive a percentage of the proceeds (Section 3.1.3) generated from the sale of goods, encouraging them to continually enhance the quality of the goods.

4.4.3 Transactions of services

Due to the reasons mentioned in the beginning of Section 4.4, offering **transactions of services** (e.g., cloud hosting services, premium support, maintenance, customization, training, advisory, migration and integrations) within the proposed Marketplace contributes towards: (a) reducing the risk for startups failing (Section 2.2), (b) addressing the impediments towards scientific discoveries and technological breakthroughs (Section 2.3), and (c) increasing funding for further research (Section 2.3.8). This ultimately leads to maximizing the overall efficiency, effectiveness and performance of all the relevant parts of the *cycle of innovation* (Fig. 14), which is the key operational mechanism of the proposed Ecosystem (Section 4.2) in order to address the needs and preferences of the users (Section 2.1), while the Foundation (Section 4.1) is a mere participant in the both the Marketplace and the Ecosystem in order to catalyse innovation, growth and prosperity.

While the reason for offering transactions of services is crucial, **implementation** also matters. More specifically, *services* should be offered in a manner that promotes accessibility, innovation, collaboration and transparency. A few specific examples of implementation include the following:

(1) **Standardized services:** Using Optional Open Standards (Section 4.2.3), services such as customization, migration and integrations should be offered in one of the predefined standardized methodologies in order to promote *compatibility, interoperability* and *accessibility*.

(2) **Clear licensing and usage terms:** Software providers should clearly define the licensing and usage terms for their data. For an open-source Marketplace, Apache licenses or similar open licenses can be employed to ensure that the software can be freely used, modified and shared by the community.

(3) **Version control and documentation:** Software should be accompanied by version control systems (e.g., Git) and comprehensive documentation. Enabling code versioning and change tracking systems allows users to access previous versions, track modifications, and understand the evolution of datasets over time. This also allows users to contribute improvements and understand the software's characteristics, thus facilitating transparency and reproducibility.

(4) **Fair Rewarding and other incentives for solution providers:** Introducing incentives for Developers can promote viability and economic sustainability. For instance, the Marketplaces can implement a revenue-sharing mechanism, where *solution providers* receive a percentage of the proceeds (Section 3.1.3) generated from the construction of the software, encouraging the continuous production of high-quality software (both in terms of code and user experience).

(5) **Industry-specific solutions:** Customizing the core software to specific domains (e.g., healthcare, finance, or autonomous vehicles) caters to the unique needs of various industries, fostering collaboration and innovation in those specific areas.

4.4.4 Decentralized Exchange

Due to the reasons mentioned in the beginning of Section 4.4, offering a **Decentralized Exchange (DEX)** within the proposed Marketplace contributes towards: (a) reducing the risk for startups failing (Section 2.2), (b) addressing the impediments towards scientific discoveries and technological breakthroughs (Section 2.3), and (c) increasing funding for further research (Section 2.3.8). This ultimately leads to maximizing the overall efficiency, effectiveness and performance of all the relevant parts of the *cycle of innovation* (Fig. 14), which is the key operational mechanism of the proposed Ecosystem (Section 4.2) in order to address the needs and preferences of the users (Section 2.1), while the Foundation (Section 4.1) merely participates in the both the Marketplace and the Ecosystem in order to catalyse innovation, growth and prosperity.

While the reason for offering a DEX is crucial, **implementation** also matters. More specifically, the *DEX* should be offered in a manner that promotes accessibility, innovation, collaboration and transparency. A few specific examples of implementation include the following:

(1) **User-friendly interface:** The DEX should have an intuitive and user-friendly interface, making it easy for both experienced and novice traders to participate (possibly by offering a simplified and an advanced theme or mode). It should also provide clear and comprehensive information about assets, order books, and transaction history.

(2) **Interoperability and liquidity incentives:** The DEX should support multiple blockchain networks, allowing users to trade a wide range of digital assets across different protocols. This ensures a broader user base and increased liquidity. Also, the DEX should incorporate mechanisms to incentivize liquidity providers, such as offering rewards, staking opportunities, or yield farming programs. This helps ensure a vibrant trading environment.

(3) **Community-driven governance:** An effective DEX should be governed by the community through *decentralized* Decision-Making processes. This could be achieved through voting mechanisms or token-based governance models, ensuring transparency and avoiding centralized control.

(4) **Multi-signature wallets:** The use of *multi-sig wallets* should be enabled so that multiple parties (e.g., any 3 out of 5 users) will be needed in order to authorize transactions, thus enhancing security and reducing the risk of fraud.

(5) **P2P lending and borrowing:** Lending and borrowing functionalities are crucial to be integrated within the DEX in order to allow users to earn interest on their assets or borrow funds for leverage, under certain terms and conditions.

4.4.5 Reputation and recommendation system

In the context of the Marketplace, an AI-powered **reputation and recommendation system** plays a crucial role in ensuring the *quality, credibility*, and *trustworthiness* of the available offerings. With the growing popularity and adoption of enterprise AI solutions, having such a system becomes essential for both Buyers and Sellers (Fig. 38) due to various reasons, as explained below. Overall, leveraging AI in implementing such a system brings efficiency, scalability, and continuous improvement.

More specifically, a reputation system allows users to assess the reliability and expertise of AI solution providers. It enables buyers to evaluate the track record and performance of sellers, based on previous experiences and feedback from other users. This helps in making informed decisions and mitigating the risks associated with choosing an unsatisfactory or unreliable solution. Moreover, a recommendation system enhances the user experience by providing personalized suggestions and guidance. It analyses user preferences, requirements, and past interactions to offer relevant and tailored recommendations. This not only saves time for buyers in finding suitable AI solutions but also helps sellers in reaching their target audience effectively.

4.5 Automated Data Science

In general, the **Data Science process** (Fig. 39) is essentially an extension of the Scientific Method (Fig. 4), with the main difference that a *data product* (e.g., an app or a platform) is typically expected to be delivered from the former, but not from the latter. Automating the entire Data Science process is crucial for a variety of reasons. Automation increases efficiency by reducing the time and effort required to perform repetitive and time-consuming tasks. For instance, Data scientists spend a significant portion of their time on Data Preparation (including Data Cleaning, Data Preprocessing, Feature Selection, Feature Extraction), but also during the Machine Learning step (i.e. model selection and evaluation). By automating these steps, they can focus on more creative and complex aspects of their work, leading to faster and improved *product delivery*. Moreover, automation also enhances reproducibility and consistency. By capturing the entire data science pipeline in an automated workflow, Researchers can easily *reproduce* their experiments, verify results, and compare different approaches. This ensures that insights and models can be shared, validated and built upon by others.

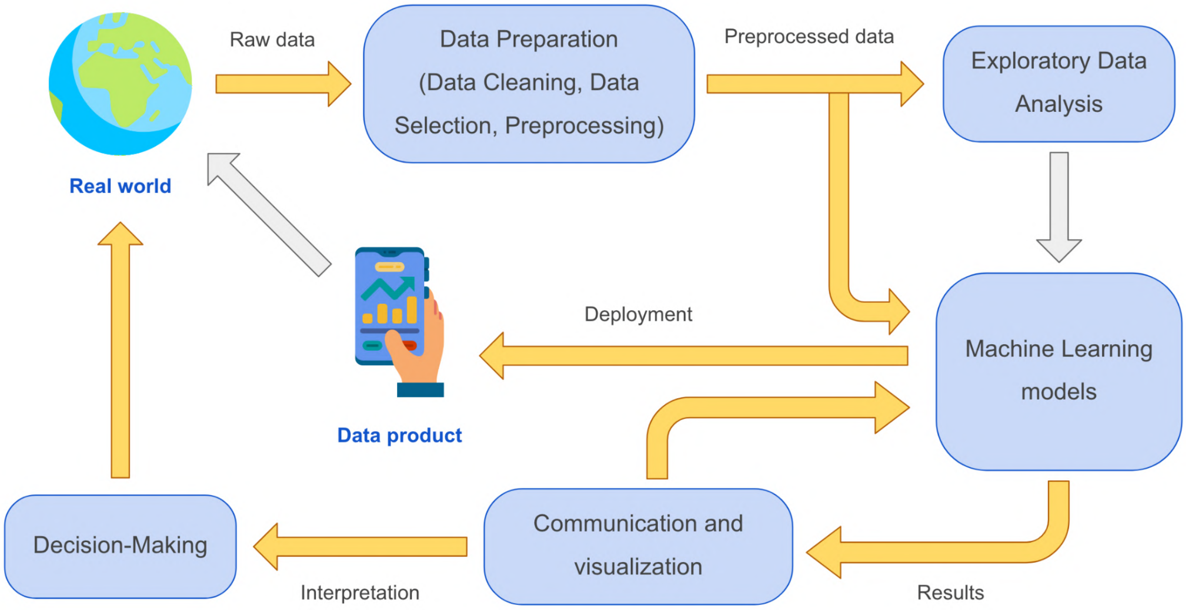


Figure 39 - Data Science process.

Furthermore, automating the Data Science process enables *scalability*. As the volume and complexity of data continue to grow, manual processing becomes impractical. Automation allows organizations to handle larger datasets, perform real-time analysis and adapt quickly to changing business goals or market needs. Additionally, automation promotes collaboration and knowledge sharing. By standardizing and encapsulating best practices in automated workflows, teams can work together more seamlessly, exchange ideas, and build upon each other's work. It also facilitates the democratization of Data Science, allowing non-experts to leverage automated tools and gain valuable insights from data. Moreover, the automation of Data Science process could get humanity a step closer towards the Automated Science Discovery (Section 2.3.10).

Overall, the proposed Ecosystem is uniquely positioned in order to facilitate **Automated Data Science** for the benefit of its users, but also for the benefit of humanity as a whole.

4.6 Automated Product Discovery

Automating the entire **Product Discovery process** (Fig. 40) is crucial for several reasons. Firstly, it significantly speeds up the process. Manual Product Discovery involves extensive research, data collection, and analysis, which can be time-consuming. Automation allows for efficient data gathering, analysis, and pattern recognition, thus enabling faster Decision-Making and reducing time-to-market. Secondly, automation improves accuracy and reduces human error. Manual processes are prone to subjective biases, oversights, and inconsistencies, whereas automation ensures a more objective and systematic approach. Software can process vast amounts of data, identify trends, and provide actionable insights with higher precision and higher level of transparency.

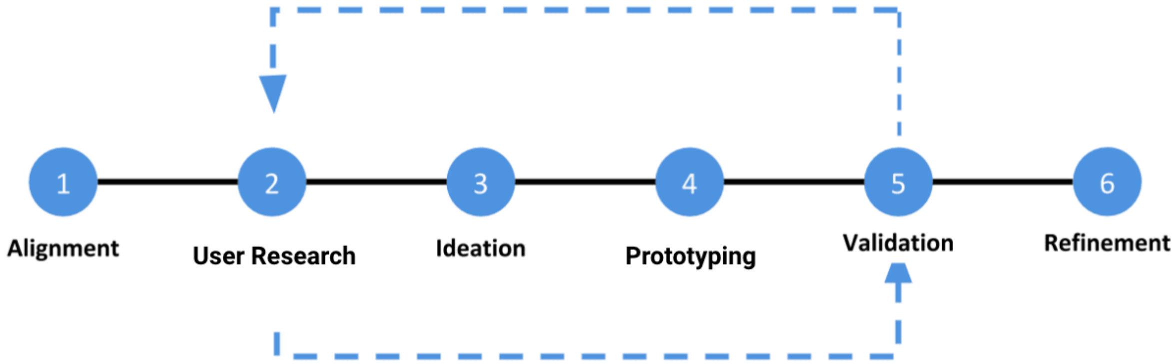


Figure 40 - Product Discovery process.

Furthermore, automation enhances *scalability*. As businesses grow and the volume of data increases, managing Product Discovery manually becomes increasingly challenging. Therefore, automation enables handling larger datasets and more complex data analyses, thereby allowing businesses to adapt to changing market demands and make better-informed decisions.

Additionally, automation promotes *innovation*. By streamlining routine tasks, it frees up time and resources for teams to focus on strategic thinking, creativity and experimentation. This empowers them to explore new ideas, iterate on existing products, and identify untapped opportunities, fostering innovation and competitive advantage.

Overall, the proposed Ecosystem is uniquely positioned in order to facilitate **Automated Product Discovery** for the benefit of its users, but also for the benefit of humanity as a whole. Two crucial aspects of Automated Product Discovery are: (a) Automated Prototyping (Section 4.6.1), and (b) Automated UX (Section 4.6.2).

4.6.1 Automated Prototyping

In the context of Automated Product Discovery, there is a series of reasons why **Automated Prototyping** is important. Through tools like prototyping software or low-code platforms, Developers can rapidly iterate on feature ideas, ensure accurate implementation and gather valuable feedback, ultimately resulting in higher-quality software products, *without* expending a relatively large amount of resources.

More specifically, it speeds up the feature development process. Manual prototyping of new features often involves significant coding efforts and time-consuming iterations. Automation allows for the creation of interactive prototypes using tools like prototyping software or low-code platforms. This enables Developers to quickly visualize and test new feature ideas, iterate on them rapidly, and gather user feedback, ultimately accelerating the Product Discovery process (Fig. 40). Additionally, automation enhances collaboration and communication. With automated prototyping, Developers, Designers, and stakeholders can easily share and demonstrate the new feature prototypes, gather feedback and align on requirements. This streamlines the feedback loop, fosters collaboration, and ensures that the final implementation meets the desired objectives.

As an illustration, consider the Automated Prototyping of a new **chatbot** feature for a Customer Support platform. Through automation, Developers can quickly build and simulate the chatbot's conversation flow, user interface and integration with the existing APIs or back-end systems. They can iterate on different design options, test the feature with real or simulated user interactions, while gathering feedback from users. This Automated Prototyping process allows for efficient exploration and refinement of the chatbot feature, thus resulting in a more accurate and effective implementation.

4.6.2 Automated UX

In the context of Automated Product Discovery, there are multiple reasons why **Automated UX** is important, mainly because it empowers Designers and Developers to create intuitive and user-friendly experiences, ultimately benefiting both the end-users and the business.

More specifically, it enhances consistency and reliability by eliminating human error. Automation ensures that the UX is standardized across different platforms, devices and user interactions. This consistency builds trustworthiness and familiarity among users, thus leading to improved user satisfaction and loyalty. Additionally, automating UX reduces development time and costs. With predefined templates, design elements, and user flows, Developers can create and update UX components more efficiently. This streamlines the Product Discovery process (Fig. 40), allowing teams to focus on other critical aspects of the software. Also, automation helps identify and fix UX issues early on, thereby minimizing the need for extensive redesigns or costly user testing.

As an illustration, consider a software product like an **e-commerce platform** that wants to test different product page layouts in order to optimize conversions by using A/B Testing. Through automated UX, the website can easily create and deploy multiple versions of the product page with varying web elements, such as button placement, colour schemes, product image sizes or call-to-action text. The automation tool can ensure that each version is presented to a specific segment of the website's visitors randomly and that user interactions are tracked accurately and transparently. By automating the UX in A/B Testing for such an e-commerce platform, businesses can iterate quickly, test different hypotheses, and make data-backed improvements to their product pages. This not only enhances the user experience, but also improves the website's overall performance, traffic and revenue.

4.7 Use cases

There is a plethora of **use cases** for the proposed Ecosystem (Section 4.2). Some indicative use cases, which also happen to promote *social good*, are presented below, namely:

- Education Technology (Section 4.7.1)
- Affective Social Robots for Wellbeing (Section 4.7.2)
- Augmented and Mixed Reality (Section 4.7.3)
- Medical Imaging (Section 4.7.4)
- Water Supply and Sanitation (Section 4.7.5)

The list of use cases is actually endless, but the list above should reflect the *tremendous impact* that the proposed Ecosystem can bring to its intended users (Section 2.1) and also the humanity as a whole in order to fuel the *cycle of innovation* (Fig. 14).

4.7.1 Education Technology

A potential client for the AI solutions offered by the Foundation (Section 4.1.10) is a for-profit **Education Technology (EdTech)** company targeted towards adult learners in developing or low-income countries, such as those in sub-Saharan Africa. For example, such an EdTech company could build its whole AI infrastructure (Section 2.1.3) on top of the proposed Ecosystem in order to promote *social good*, while offering some or all of its education products and services through the proposed Marketplace (Section 4.4). The EdTech company can aim to improve both the cognitive and the noncognitive skills (Fig. 19) of professionals who are either employed in the industry or are freelancers. Such freelancers might already be participating in or intent to offer their services through the Marketplace, which hints the synergies between the EdTech company and the Ecosystem.

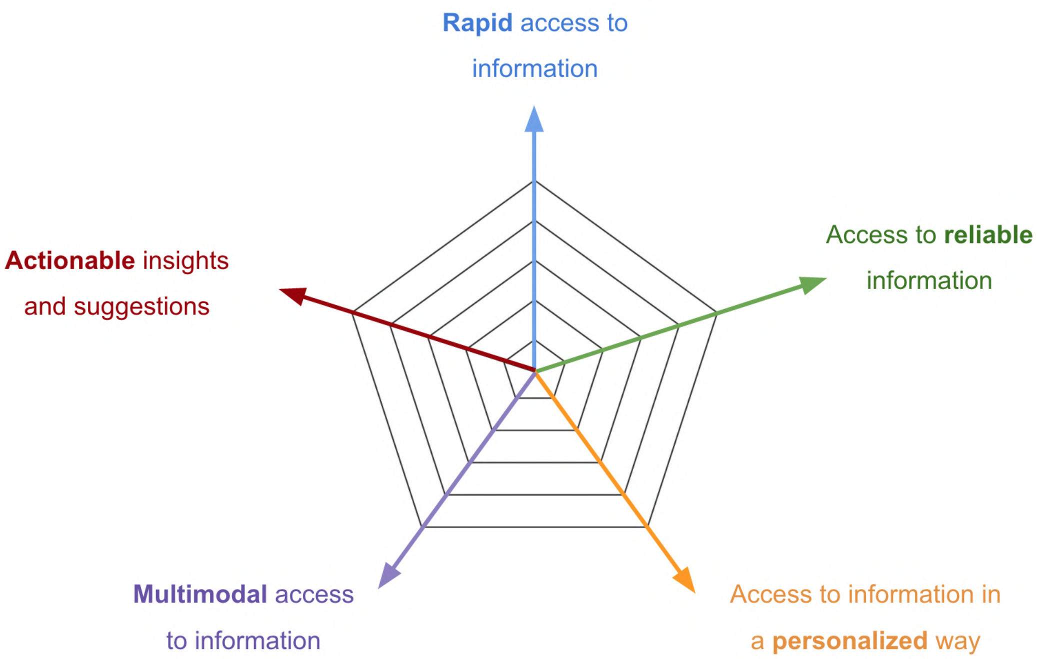


Figure 41 - Pentagon of education.

In general, in order for education to be effective — especially when offered through an online *e-learning platform* — there are *five* independent factors that need to be maximized, termed as **pentagon of education** (Fig. 41) in this whitepaper. Namely, these components are listed below in an order of decreasing importance:

(1) *Rapid* access to information (which relates to Technology)

(2) *Actionable* insights and suggestions (which relates to Technology)

(3) Access to *reliable* information (which relates to data and content)

(4) Access to information in a *personalized* way (which relates to user experience)

(5) *Multimodal* access to information (which relates to user experience)

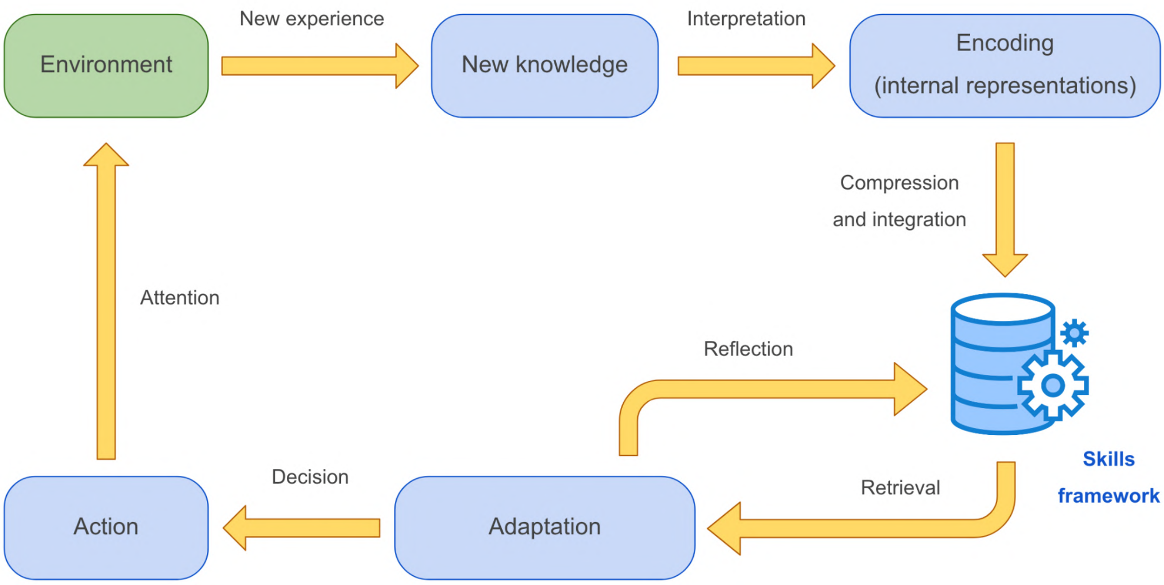


Figure 42 - The cycle of human learning.

The proposed Ecosystem has the ability to effectively and efficiently facilitate all five of the aforementioned factors, whereas the decentralized nature of the Ecosystem ensures *censorship-resistance* right at the level of physical and cyber infrastructure. Reciprocally, such an EdTech company could also positively contribute towards the adoption of the

Ecosystem by a wider audience through training and educating the users, thus creating a win-win situation between the both the Ecosystem and the EdTech company. Additionally, the Ecosystem could assist in the creation of multiple such EdTech companies in order to promote translation and localization, thereby catering users from multiple cultures and different demographics. Overall, such EdTech companies would have a significant competitive advantage if implemented on top of the Ecosystem.

Furthermore, regarding both cognitive and noncognitive skills (Fig. 19), the **cycle of human learning** (Fig. 42) can be modelled as a *closed-loop control system* (Fig. 3). Therefore, Automated Scientific Discovery (Section 2.3.10) implemented as part of the Ecosystem could act as a catalyst in order to further understand and unlock the hidden mechanisms of nature regarding human learning both at the level of the *individual* and also at the level of the *local community*, when considering collective learning. In turn, this could also get humanity a step closer towards Artificial General Intelligence (Section 3.4), since human learning can act as a source of bio-inspiration for Machine Learning.

4.7.2 Affective Social Robots for Wellbeing

A further use case for the AI solutions offered by the Foundation is that of **Affective Social Robots** (Fig. 43) for wellbeing, especially in elderly suffering from loneliness or dementia. Such robots aim to better understand *social and emotional cues* (both verbal and non-verbal) from humans in order to improve HuMARL collaboration and interaction (as defined in Section 3.3.9) in the context of both *mental* and *physical wellbeing*, by adhering to human social norms in order for such interactions to be smooth and natural.

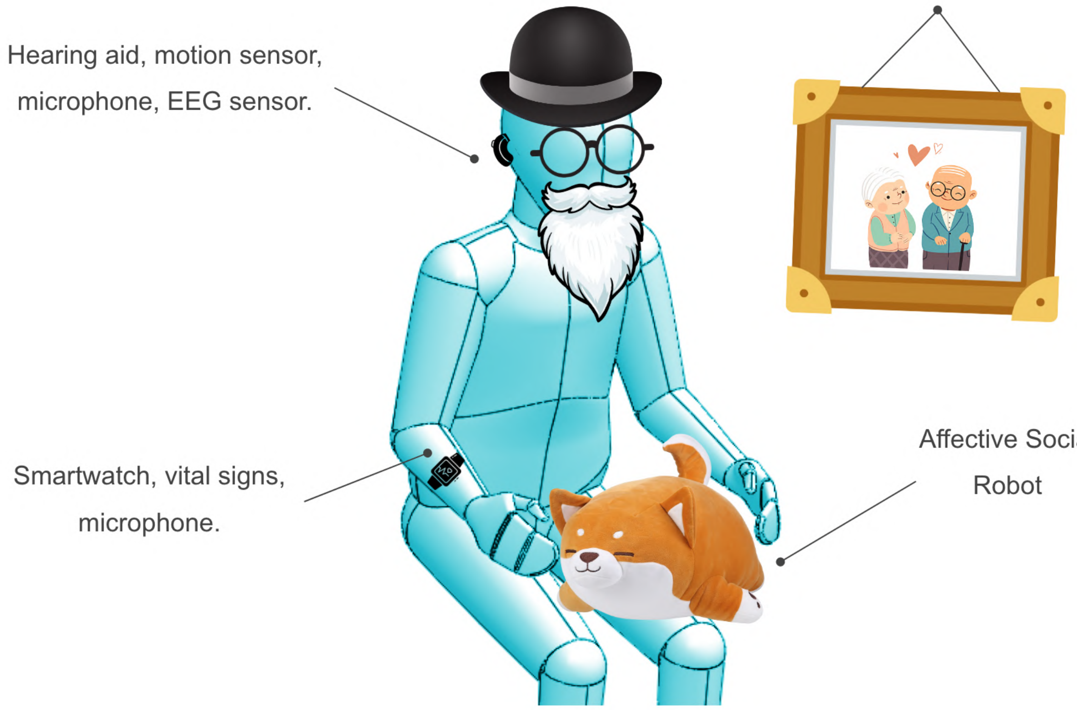


Figure 43 - Example of Affective Social Robot for the Wellbeing of the elderly.

In general, **emotions** carry *informational content*, allowing the transmitter and the receiver to form common ground so they can communicate more effectively, even if non-lexical (e.g., paralinguistic) or non-verbal (e.g., visual) communication is involved. For example, an expression of "sadness" (either facial or vocal) might indicate a desire to be comforted. The authors of this whitepaper claim that even our understanding of the six basic emotions — anger, disgust, fear, joy (or happiness), sadness and surprise — is far too simplistic. By allowing AI systems to 'maximize for creativity' (Section 3.3.3), the authors hypothesize that AI systems might come up with six *different* basic emotions or even a different *number* of basic emotions. The proposed Ecosystem provides a general framework to conduct collaborative and controlled trial-and-error experimentations in a trustworthy way in order to investigate such hypotheses, while also allowing the commercialization of the research findings in an open and transparent way that can benefit the whole of humanity.

Moreover, **Conversational AI** [1] can play a crucial role in the implementation of Affective Social Robots for physical and mental wellbeing, since it holds significant potential in facilitating the implementation of such robots. By employing Signal Processing and Natural Language Processing techniques, Conversational AI enables robots to participate in meaningful and engaging human-like interactions, thus offering companionship, support and emotional connection.

More specifically, Sentiment Analysis and Emotion Recognition can enable these robots to understand, interpret and respond appropriately to human emotions, thereby creating a more empathetic (Fig. 19), contextually relevant and personalized experience for the users. This level of understanding helps robots tailor their responses, adapt their behaviours, and effectively communicate with humans through seamless and dynamic conversations, while also increasing trustworthiness towards AI (Section 3.3) and enabling robots to fulfil their intended roles as companions, caregivers, and emotional support systems.

Additionally, Conversational AI enables **Continual Learning** [1] and improvement of the robot's abilities to recognize, understand and respond to emotions through conversations. In particular, by collecting and processing data from user interactions, robots can iteratively learn from user feedback and enhance the overall quality of the conversation, while adapting and refining their responses and behaviour, thereby becoming more effective at providing emotional support over time.

4.7.3 Augmented and Mixed Reality

The possibilities with **AR and MR technologies** (Fig. 44) are endless. This is a particularly promising use case for the proposed Ecosystem. Even though AR and MR equipment is still at its infancy, the Ecosystem can accelerate the design and development of such equipment, while also facilitating the creation of dApps and dPlats utilizing such equipment, even at this very primitive form. These immersive technologies can revolutionize how individuals interact with decentralized systems, especially in the fields of healthcare, telemedicine and medical surgery (Fig. 45). In particular, AR glasses can overlay digital information onto the physical world, enabling individuals to visualize and interact with decentralized applications, smart contracts, and virtual assets in their surroundings. MR headsets can create fully immersive decentralized environments, allowing users to actively participate in virtual appointments or educational experiences within decentralized platforms.

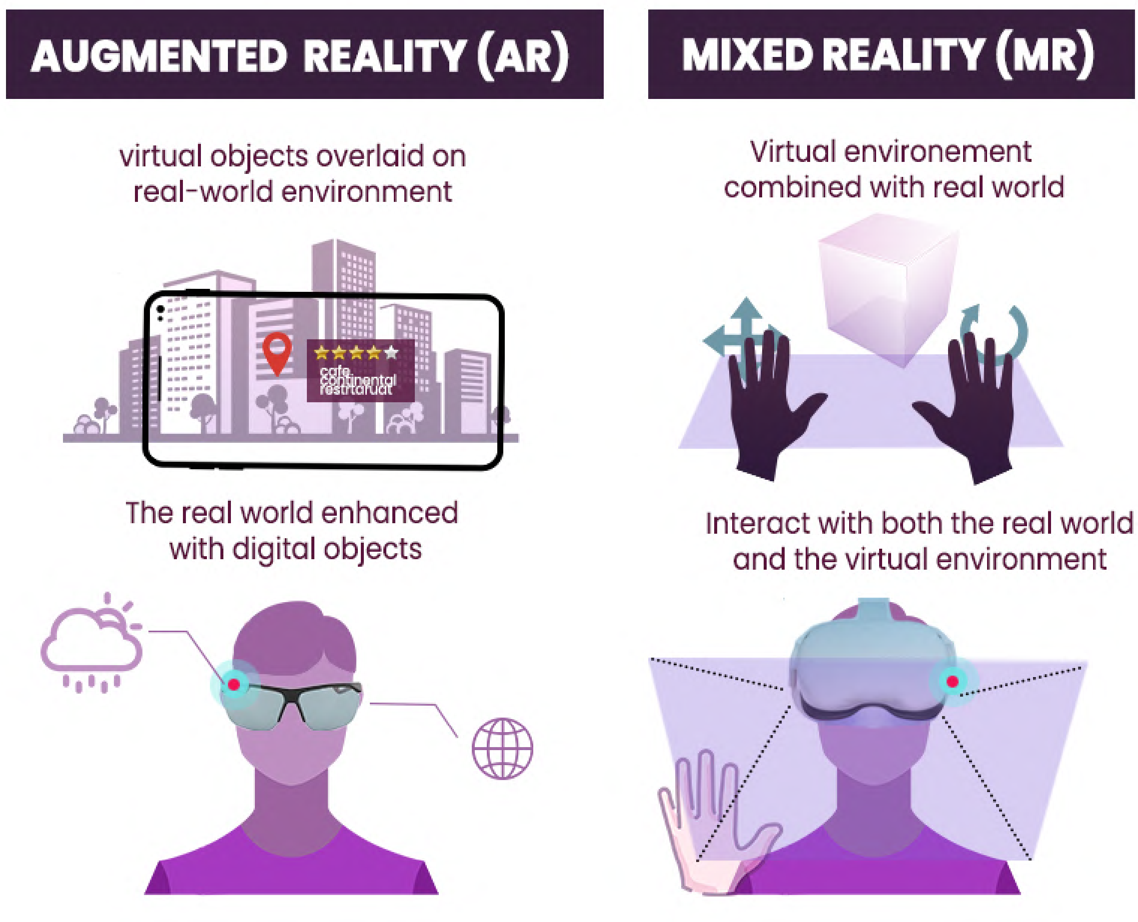


Figure 44 - Augmented Reality (left) and Mixed Reality (right).

More specifically, in terms of *precise navigation and manipulation* during **heart and cardiovascular surgeries** (Fig. 45), some significant challenges or obstacles that can make the procedures more challenging for surgeons are the following:

(1) **Limited visibility:** The restricted field of view within the heart and blood vessels can make it challenging for surgeons to visualize and navigate through the specific anatomical structures.

(2) **Limited working space:** The tight confines of the chest cavity and blood vessels can restrict the movement of surgical instruments, making it difficult to manoeuvre and perform complex tasks.

(3) **Beating heart surgeries:** In certain cases, surgeries need to be performed on a beating heart, which introduces additional challenges due to the constant motion and the need for stabilization techniques.

(4) **Tissue deformation:** Manipulating soft and delicate cardiac tissues can lead to deformation, thereby making it harder to achieve precise surgical manoeuvres and suturing.

(5) **Scar tissue from previous surgeries:** Patients who have undergone previous cardiac procedures may have scar tissue present, which can complicate surgical navigation and increase the risk of inadvertent injury.

(6) **Anomalous anatomical variations:** Some patients may have congenital or acquired anatomical variations, such as abnormal vessel origins or course, which require careful adaptation and pose challenges during surgical navigation.

(7) **Intraoperative haemorrhage:** Bleeding during surgery can obscure the surgical field, therefore impairing visibility and making it difficult to perform precise manipulations.

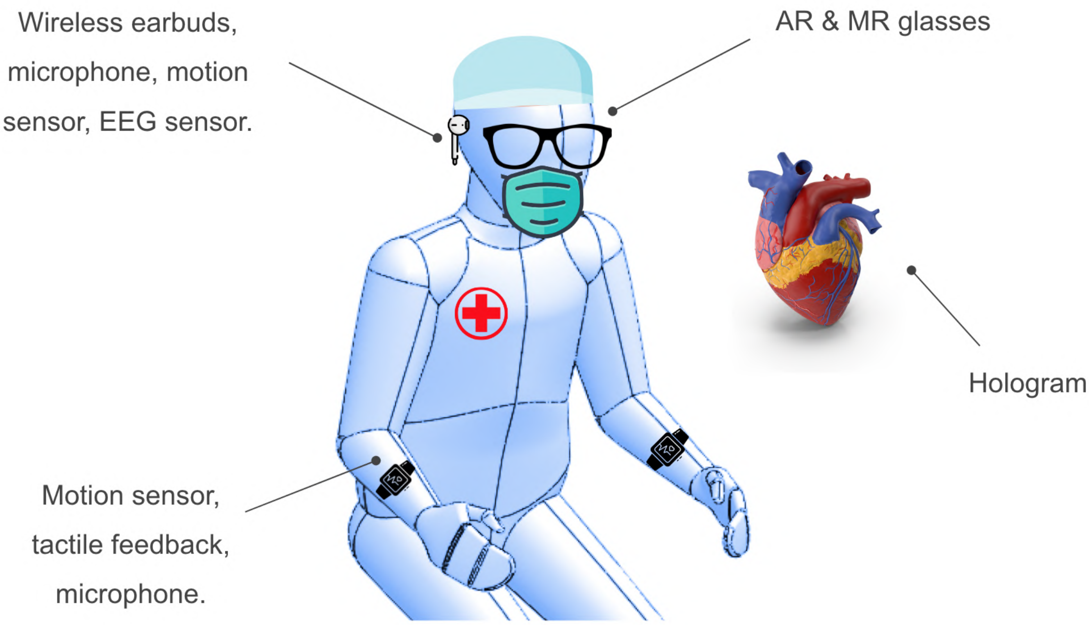


Figure 45 - Examples of AR and MR devices in brain (top) and heart (bottom) surgeries.

In addition to AR and MR technologies, addressing these challenges often requires the use of advanced *Medical Imaging* technologies (Section 4.7.4), robotic-assisted surgery or specialized instruments designed for improved visualization, manoeuvrability and accuracy, all of which can also be facilitated through the use of AI solutions. Therefore, the Ecosystem is uniquely positioned to holistically address the use of AR and MR technologies in this type of medical surgeries, but also in healthcare in general, thus promoting social good.

4.7.4 Medical Imaging

The offering of AI solutions in the use case of **Medical Imaging** has been particularly important so far, but it also holds immense potential for the near future. For instance, Firstly, AI has already significantly impacted medical imaging by improving diagnostic accuracy and efficiency. AI systems can process vast amounts of Medical Imaging modalities with incredible speed, aiding in the detection of abnormalities and assisting radiologists in making more accurate diagnoses. This has resulted in earlier detection of diseases, better treatment planning, and ultimately improved patient outcomes. Additionally, AI systems have the ability to learn from large-scale datasets, detecting patterns and trends that may not be easily discernible to the human eye, thus assisting and supporting clinical Decision-Making.

Regarding the future, the potential of AI in Medical Imaging is even more promising. In the context of the proposed Ecosystem, AI systems have the potential to become more sophisticated, thus enabling them to handle complex tasks like predicting treatment response, personalized medicine and early disease detection. Integration of core AI solutions with technologies such as Medical Robotics and Genomics can further enhance the capabilities of Medical Imaging, revolutionizing the field and transforming healthcare.

Furthermore, by securely storing and encrypting medical data on a blockchain network in a privacy-aware manner, **BioNFTs** (Section 4.2.5) have the potential to significantly advance the field of Medical Imaging. Firstly, they can enable the secure sharing and exchange of medical images, reports and patient records among healthcare providers, facilitating seamless collaboration and improving diagnostic accuracy. Additionally, BioNFTs can ensure data integrity and traceability, thus preventing tampering or unauthorized access to sensitive medical information. Moreover, these tokens can incentivize data contribution by patients, Researchers, and healthcare institutions, creating decentralized and comprehensive datasets for training AI algorithms and advancing medical imaging techniques. Ultimately, BioNFTs can empower patients, enhance research collaborations, while also fuelling innovation in Medical Imaging, hence leading to more accurate diagnoses, personalized treatments and improved patient outcomes.

4.7.5 Water Supply and Sanitation

Water supply and sanitation is a crucial use case in the context of AI solutions offered through the proposed Marketplace, mainly due to its significant impact on promoting social good through public health, environmental sustainability, and socio-economic uplifting of *developing* and *low-income* countries, such as the ones in sub-Saharan Africa.

In the context of developing and low-income countries in sub-Saharan Africa, a holistic solution could incorporate, for example, Distributed Smart Grids, decentralized agricultural systems, DePINs (Section 4.3) and the Marketplace (Section 4.4), all of which are supported by the proposed AGI token (Section 4.2.4) and the Ecosystem in general. This makes sub-Saharan Africa an ideal candidate to deploy such DePIN solutions, since the physical infrastructure neither exists nor is it economically viable for centralized authorities in such countries to start building such infrastructure, thus making decentralization the only viable option for such populations to survive, but also to maintain an acceptable level of sanitation.

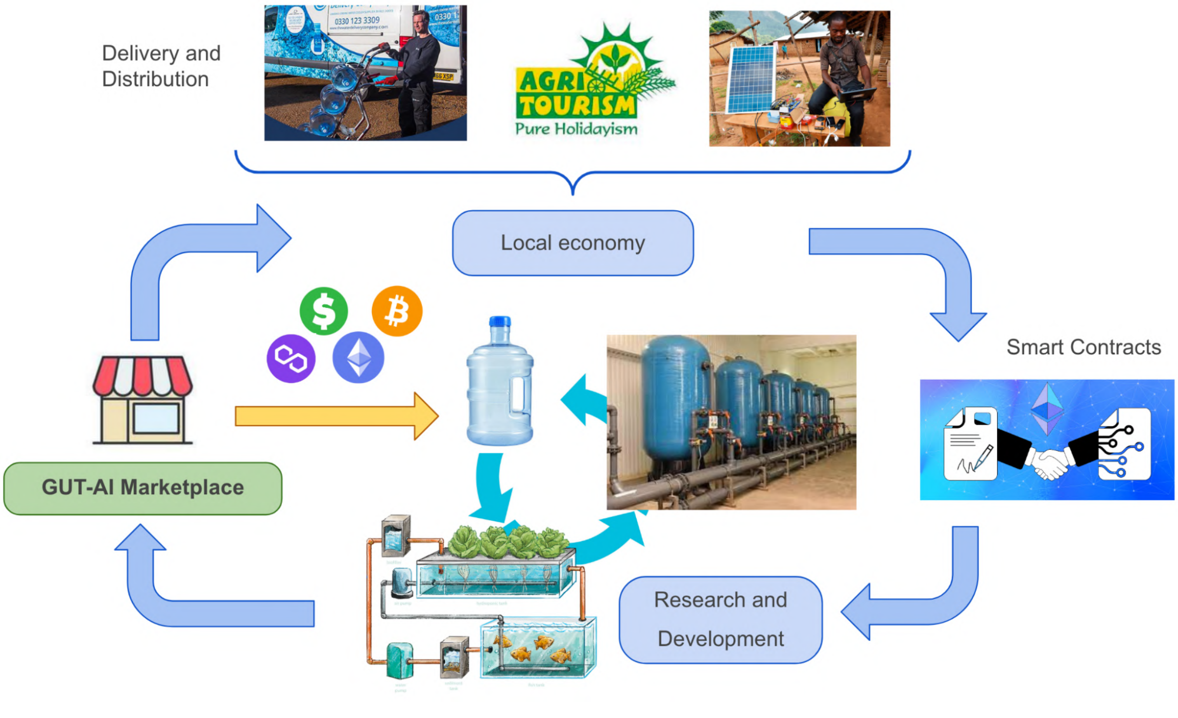


Figure 46 - Water supply and sanitation system for sub-Saharan Africa.

Together, decentralization and Blockchain can create an ecosystem where local communities have more control over their water supply and sanitation, leading to increased reliability, sustainability, and equitable access. By reducing bureaucracy, enhancing transparency and fostering collaboration, a *decentralized system* of water supply and sanitation can unlock new opportunities for sub-Saharan Africa to address its sanitation, while improving the access to clean water and wellbeing of its people.

For instance, by decentralizing Decision-Making power and responsibilities to local communities, such a **water supply and sanitation system** (Fig. 46) can empower them to manage their water resources effectively, while promoting community engagement, ownership and accountability, thereby fostering sustainable and context-specific solutions. Therefore, local authorities can respond quickly to the needs of the local economy and society in order to ensure the efficient allocation of resources (as explained in Section 4.1.7).

5 GUT-AI Initiative

This whitepaper is a *companion* document to a research proposal [1], which is targeted to a more research-oriented audience. The research proposal sets the specific framework to study 'narrow' AI and 'strong' AI (or AGI, as described in Section 3.4) in the future, while also introducing four novel theoretical concepts, among other contributions. However, this whitepaper was written with a general audience in mind (e.g., Investors, Researchers, Developers, Designers, Marketers) and focuses on the Ecosystem. Both documents are part of the broader **GUT-AI Initiative**[4], which has the following vision and mission.

5.1 Pitch

The GUT-AI Initiative is a totally *decentralized* initiative which aims to eliminate the *multiple single points of failure* when using AI for real-life applications in the real world in order to achieve the **ultimate purpose** [1] of both 'narrow AI' and 'strong AI', which is to actually "open" the "black box" of an ML system in order to eventually unlock the mysteries of nature and the universe (from Brain Consciousness and Abiogenesis to Quantum Gravity and Genesis Cosmology). For instance, does evolution or the universe have a conscious or intelligent "geist" (spirit), as Max Planck once claimed?

5.2 Vision

We believe that there should be no organization or person in our world who wants to use AI, but not be able to do so. We also believe in a world where AI hand-in-hand with *human interaction* are in an ever-improving situation.

[4] https://gut-ai.org/

5.3 Mission

We are on a mission to create the most *user-friendly* **Open-Data, Open-Source, Decentralized** ecosystem for AI using cutting-edge technology either of the 21st century or that we might invent by ourselves.

5.4 Roadmap and further steps

The GUT-AI Initiative, as described above, is an *ever-evolving* initiative that goes beyond the scope of this whitepaper. This whitepaper describes only *some* of the subprojects of GUT-AI, which are called **components**, and are hosted on OSF[5]. Table 6 is the current version of an indicative table of which component *corresponds* to which Section in this proposal, but it remains subject to change[6] as part of the overall initiative.

As also stated in the beginning of Chapter 4, each component definition intentionally does not include *'how'* to be implemented, but only *'what'* to be implemented. The reason is that there should be no constraints or limits on the 'how' since new advances in Technology can potentially bring new opportunities to improve the 'how' a specific component is implemented. The 'why' each component is necessary is explained in the Vision (Section 6.2) and Mission (Section 6.3) of GUT-AI. The aim of the GUT-AI initiative is not to reinvent the wheel. If a tool or solution for something already exits, then a type of integration can be created for that tool or solution. For similar reasons and since it can easily be rendered outdated from one day to another, the **roadmap** is deliberately *not* included in this whitepaper, but it is hosted online[7] instead, so that it can be dynamically updated.

[5] https://osf.io/rn2s4/
[6] https://github.com/GUT-AI/gut-ai/blob/master/components/README.rst
[7] https://github.com/GUT-AI/gut-ai/blob/master/roadmap/README.rst

C1.1: Distributed Smart Grids	Section 4.3.3
C1.2: GUT-AI DCP	Section 4.3.4
C2.1: GUT-AI Marketplace	Section 4.4
C2.4: DX	Section 2.1.4
C3.1: Automated Data Science	Section 4.5
C3.2: AutoML	Section 3.4.2
C4.1: Automated Scientific Discovery	Sections 2.3.10 and 3.4.1
C5.1: Automated Prototyping	Section 4.6.1
C5.2: Automated UX	Section 4.6.2
C5.10: Medical Imaging	Section 4.7.4

Table 6 - Correspondence of GUT-AI components and Sections in this whitepaper.

6 Conclusion

In this final chapter, we bring our journey through this whitepaper to a close. Throughout this whitepaper, the fundamental concepts, vision and mission of the GUT-AI Foundation have been explored, and now it is time to conclude this remarkable exploration. Also, the authors have introduced the proposed Ecosystem, which fosters the creation of new concepts and implementations. This includes everything from scientific discoveries and technological advancements to commercialization of the research findings and funding further research via AI-powered Automated Scientific Discovery that will drive the *cycle of innovation*.

Ultimately, the Foundation aspires to shape AI into a force for positive change, ensuring that the deployment of AI solutions will benefit humanity as a whole. By fostering innovation and creativity, promoting Trustworthy AI, and fostering public dialogue, the Foundation seeks to harnessing the power of AI by enhancing and amplifying human capabilities in order to tackle the world's most pressing challenges, especially those occurring in developing and low-income countries. Additionally, the Ecosystem's commitment to transparency, accountability and societal well-being due to its *fully decentralized* nature will be instrumental in guiding the implementation and deployment of AI solutions.

Furthermore, the Foundation recognizes the need for interdisciplinary collaboration, while also engaging the public and raising awareness in order to unleash the full potential of AI. By providing educational resources and research opportunities, the Foundation aims to empower individuals and organizations to understand and contribute to the implementation of AI solutions. By facilitating open discussion, raising awareness about AI's capabilities and limitations, while also addressing legitimate concerns and also debunking irrational fears, the Foundation seeks to create a vibrant community built around the Ecosystem.

Looking ahead, the AI Foundation aims to work as a *catalyst* for implementing the Ecosystem, but *without* interfering with its decentralized nature. It aims to support groundbreaking research, empower individual Researchers and Developers, while attempting to balance and mitigate risks in order to reap the vast amount of benefits from the research and development of AI systems.

As we conclude this whitepaper, we invite all readers to join hands in shaping the future of AI by contributing to the realization of the Ecosystem that harnesses the collective intelligence and creativity of humanity for the benefit of all the people. By working together, we can shape a future of AI and harness its immense potential for providing an extraordinary life to individuals and communities worldwide, while unlocking the tremendous possibilities that AI holds for our shared future. The journey towards an extraordinary future is just beginning, and by acting responsibly and decisively as a *beacon* of inspiration and guidance, the Foundation stands ready to lead the way.

Visionaries of the world, unite!

Acknowledgments

The authors would like to thank all those who feel that they have contributed one way or another towards improving this whitepaper. They authors wished that the Ecosystem was already in existence and that Fair Rewarding (Section 3.13) was already implemented, so that they could just provide a link to each such contribution and conversation in order to cite all of them.

References

[1] Kourouklides, I. (2022). Bayesian Deep Multi-Agent Multimodal Reinforcement Learning for Embedded Systems in Games, Natural Language Processing and Robotics. *OSF Preprints*. [link]

[2] Aulet, B. (2013). *Disciplined Entrepreneurship: 24 Steps to a Successful Startup*. John Wiley & Sons. [link]

[3] URL: https://www.cbinsights.com/research/report/startup-failure-reasons-top/

[4] URL: https://projects.apache.org/projects.html

[5] URL: https://www.topbots.com/enterprise-ai-companies-2020/

[6] URL: https://www.bloomberg.com/news/articles/2023-02-10/car-data-firms-flop-as-automakers-see-dollar-signs-in-software?leadSource=uverify%20wall

[7] URL: https://www.counterpointresearch.com/connected-car-data-shaping-automotive-industry/

[8] URL: https://smartcar.com/wejo-vs-otonomo/

[9] Fanelli, D. (2018). Is science really facing a reproducibility crisis, and do we need it to?. *Proceedings of the National Academy of Sciences, 115*(11), 2628-2631. [link]

[10] Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature, 533*(7604). [link]

[11] URL: https://www.tensorflow.org/hub

[12] URL: https://pytorch.org/hub/

[13] URL: https://huggingface.co/

[14] URL: https://www.kubeflow.org/

[15] URL: https://mlflow.org/

[16] URL: https://airflow.apache.org/

[17] URL: https://github.com/SeldonIO/seldon-core

[18] URL: https://www.bentoml.com/

[19] URL: https://kserve.github.io/website/master/

[20] URL: https://youtu.be/sF03FN37i5w

[21] Goertzel, B., Giacomelli, S., Hanson, D., Pennachin, C., & Argentieri, M. (2017). *SingularityNET: A Decentralized, Open Market and Network for AIs*. [link]

[22] URL: https://whitepaperdatabase.com/wp-content/uploads/2018/03/DeepBrain-Chain-DBC-Whitepaper.pdf

[23] URL: https://www.openmined.org/

[24] McConaghy, T. (2022). Ocean Protocol: Tools for the Web3 Data Economy. In *Handbook on Blockchain* (pp. 505-539). [link]

[25] URL: https://www.aitn.io/AITN_EN.pdf

[26] URL: https://cryptopapers.info/assets/pdf/aigang.pdf

[27] URL: https://aiwork.io/assets/wp/AIWORK_Whitepaper_2.1.pdf

[28] Consensus, E. (2017). *NeuroChain: The Intelligent Blockchain*. [link]

[29] Ober, H., Simon, S. I., & Elson, D. (2013). Five simple rules to avoid plagiarism. *Annals of Biomedical Engineering, 41*(1), 1-2. [link]

[30] Israel, H., Ruckhaber, E., & Weinmann, R. (1931). *One Hundred Authors Against Einstein*. R. Voigtlander. [link]

[31] Sheldrake, R. (2012). *The Science Delusion*. Hachette UK. [link]

[32] Sheldrake, R. (2020). *A New Science of Life*. Icon Books. [link]

[33] URL: https://www.sheldrake.org/research/morphic-resonance/introduction

[34] URL: https://www.sheldrake.org/research/morphic-resonance

[35] Gomez-Marin, A. (2021). Facing biology's open questions: Rupert Sheldrake's "heretical" hypothesis turns 40. *Bioessays, 43*(6), 2100055. [link]

[36] Ioannides, G., Kourouklides, I., & Astolfi, A. (2022). Spatiotemporal dynamics in spiking recurrent neural networks using modified-full-FORCE on EEG signals. In *Scientific Reports, 12*(1), 1-20. [link]

[37] Firth, J., Torous, J., Stubbs, B., Firth, J. A., Steiner, G. Z., Smith, L., ... & Sarris, J. (2019). The "online brain": how the Internet may be changing our cognition. *World Psychiatry, 18*(2), 119-129. [link]

[38] Gallese, V., & Sinigaglia, C. (2011). What is so special about embodied simulation?. *Trends in Cognitive Sciences, 15*(11), 512-519. [link]

[39] Liu, H., Wang, Y., Fan, W., Liu, X., Li, Y., Jain, S., ... & Tang, J. (2022). Trustworthy AI: A Computational Perspective. *ACM Transactions on Intelligent Systems and Technology, 14*(1), 1-59. [link]

[40] Li, B., Qi, P., Liu, B., Di, S., Liu, J., Pei, J., ... & Zhou, B. (2023). Trustworthy AI: From Principles to Practices. *ACM Computing Surveys, 55*(9), 1-46. [link]

[41] Alexander, S., Cunningham, W. J., Lanier, J., Smolin, L., Stanojevic, S., Toomey, M. W., & Wecker, D. (2021). The Autodidactic Universe. *arXiv preprint arXiv:2104.03902*. [link]

[42] Esmaeilzadeh, H., & Vaezi, R. (2022). Conscious Empathic AI in Service. *Journal of Service Research, 25*(4), 549-564. [link]

[43] Kampis, D., & Southgate, V. (2020). Altercentric Cognition: How Others Influence Our Cognitive Processing. *Trends in Cognitive Sciences, 24*(11), 945-959. [link]

[44] Braz, C., Seffah, A., & M'Raihi, D. (2007). Designing a Trade-Off Between Usability and Security: A Metrics Based-Model. In *Human-Computer Interaction–INTERACT 2007: 11th IFIP TC 13 International Conference, Proceedings, Part II 11* (pp. 114-126). [link]

[45] Alshamari, M. (2016). A Review of Gaps between Usability and Security/Privacy. *International Journal of Communications, Network and System Sciences, 9*(10), 413-429. [link]

[46] URL: https://legalnodes.com/article/choose-a-crypto-friendly-country-for-dao

[47] URL: https://api.a16zcrypto.com/wp-content/uploads/2022/06/dao-legal-framework-part-1.pdf

[48] URL: https://mirror.xyz/0x43d06b9eBFB0c76A448fBd5B6faa2cfba81901d6/CYm-hNaniW0C1Mn9KR677jn4o9okGugkKC3C8iD9k28

[49] Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610-623).

[50] URL: https://magazine.scienceforthepeople.org/vol24-2-dont-be-evil/stochastic-parrots/

[51] URL: https://venturebeat.com/ai/with-gpt-4-dangers-of-stochastic-parrots-remain-say-researchers-no-wonder-openai-ceo-is-a-bit-scared-the-ai-beat/

[52] URL: https://www.dair-institute.org/blog/letter-statement-March2023

[53] URL: https://archive.is/D0dOZ

[54] URL: https://archive.is/KZjiU

[55] URL: https://businesstechguides.co/dao-voting-mechanisms

[56] URL: https://hackingdistributed.com/2018/07/02/on-chain-vote-buying/

[57] Nakamoto, S. (2008). *Bitcoin: A Peer-to-Peer Electronic Cash System*. [link]

[58] Buterin, V. (2014). *A Next-Generation Smart Contract and Decentralized Application Platform*. [link]

[59] Hsieh, Y. Y., & Vergne, J. P. (2023). The future of the web? The coordination and early-stage growth of decentralized platforms. *Strategic Management Journal, 44*(3), 829-857. [link]

[60] Makarov, I., & Schoar, A. (2021). Blockchain Analysis of the Bitcoin Market (No. 29396). *NBER Working Paper Series*. [link]

[61] URL: https://bitcoinmagazine.com/culture/bitcoin-miners-dont-exist-validators-do

[62] The DFINITY Team. (2022). The Internet Computer for Geeks. *Cryptology ePrint Archive, Report 2022/087*. [link]

[63] URL: https://messari.io/report/an-introduction-to-dfinity-and-the-internet-computer

[64] URL: https://www.coinhustle.com/on-chain-governance-on-internet-computer-using-nns-an-overview/

[65] URL: https://www.coinhustle.com/committee-based-governance-on-the-internet-computer/

[66] URL: https://www.analyticsinsight.net/why-staying-in-ethereum-blockchain-could-destroy-apecoin/

[67] Fischer, A., & Valiente, M. C. (2021). Blockchain governance. *Internet Policy Review, 10*(2), 1-10. [link]

[68] Pelt, R. V., Jansen, S., Baars, D., & Overbeek, S. (2021). Defining Blockchain Governance: A Framework for Analysis and Comparison. *Information Systems Management, 38*(1), 21-41. [link]

[69] URL: https://polkadot.network/blog/polkadot-governance

[70] URL: https://polygon.technology/blog/state-of-governance-decentralization

[71] Duparc, E., Moller, F., Jussen, I., Stachon, M., Algac, S., & Otto, B. (2022). Archetypes of open-source business models. *Electronic Markets, 32*(2), 727-745. [link]

[72] URL: https://fourweekmba.com/open-source-business-model/

[73] URL: https://www.forbes.com/sites/forbestechcouncil/2022/09/02/why-the-business-model-of-open-source-software-works/

[74] URL: https://www.gnu.org/philosophy/open-source-misses-the-point.en.html

[75] URL: https://opensource.org/osd/

[76] URL: https://chiefmartec.com/2017/06/average-enterprise-uses-91-marketing-cloud-services/

[77] URL: https://cointelegraph.com/news/depin-will-become-one-of-this-decade-s-most-important-crypto-investments

[78] URL: https://docs.decred.org/proof-of-stake/overview/

[79] Bentov, I., Lee, C., Mizrahi, A., & Rosenfeld, M. (2014). Proof of Activity: Extending Bitcoin's Proof of Work via Proof of Stake. In *Proceedings of the 2014 Joint Workshop on Pricing and Incentives in Networks and Systems*. [link]

[80] Jepson, C. (2015). DTB001: Decred Technical Brief. [link]

[81] Yakovenko, A. (2018). *Solana: A new architecture for a high performance blockchain v0.8.13*. [link]

[82] Kanani, J., Nailwal, S., & Arjun, A. (2021). Matic Whitepaper. [link]

[83] Brockhurst, M. A., Chapman, T., King, K. C., Mank, J. E., Paterson, S., & Hurst, G. D. (2014). Running with the Red Queen: the role of biotic conflicts in evolution. *Proceedings of the Royal Society B: Biological Sciences, 281*(1797), 20141382. [link]

[84] Taleb, N. N. (2014). Antifragile: Things That Gain from Disorder. Random House Trade Paperbacks. [link]