

SUPERINTELLIGENCE: A REALISTIC SCENARIO

J E Tardy
Systems Analyst
Sysjet inc.
jetardy@sysjet.com



*Many thoughtful individuals are speaking out about the possible danger of a very advanced form of Artificial Intelligence. However, they don't describe what the components and internal mechanisms of such a system could be. In this article, I first **review and summarize current consensual opinions** about Superintelligence and indicate why they are misconceptions. I then outline the architecture of **a plausible planetary Superintelligence**, its construction, components and internal mechanisms. This concrete example provides a better understanding of what a Superintelligence could be like, why most opinions about it are incorrect and why Artificial Intelligence can have serious planetary consequences.*

Date: Dartmouth NS – 2017.04.23

Keywords: Cognitive Architecture, Synthetic Consciousness, Artificial Intelligence

“IT”

Superintelligence is a popular topic. Many thoughtful individuals have been commenting about it. They include world-class experts in Artificial Intelligence and individuals who are exceptionally brilliant or highly knowledgeable about technology. These speakers tell us **superintelligence is coming**, it is important and we must make sure it is safe. Invariably, when they discuss this topic all the commentators simply refer to Superintelligence as **IT** as in...

“IT is coming...We must deploy IT safely...IT may not want what we want... ”



However, these individuals don't describe **what form** a superintelligence could take, what components it could include, how and by whom it could be implemented, in what manner it could be deployed or what mechanisms would activate its behavior.

To be fair, most of these speakers openly admit that they don't know what form it could take or how it could arise. However,

just calling something Superintelligence and saying that **IT** is coming should not be enough. The absence of any plausible reference generates misconceptions that reflect subjective human fears rather than technically based assessments.

SIX CONSENSUAL OPINIONS

Here is a sample of the various opinions and prognostics expressed in the videos I listed about that super-AI everyone calls **IT**:

- IT is coming but when is a mystery
- IT will be here in 20 to 50 years
- IT will be superintelligent when it becomes as intelligent as we are
- IT's goals may differ from ours
- IT will want to improve itself
- IT will be like a god (hopefully benign)
- IT is the greatest challenges our species will face
- IT may bring the end of mankind
- IT could produce thousands of copies of itself.
- We must keep IT inside the box until it is safe to let it out
- IT may no longer need us
- We should wait until we are ready before beginning to build IT

These opinions and others can be summarized in **six consensual positions**:

1. **A stepwise stage.** It would arrive as the third of three sequential stages;
2. **Human-like at first.** It will first be comparable to the human intellect before exceeding it.
3. **Autonomous growth.** It is a technology that will grow independently of any human participation.
4. **Unique and Perceptible.** It is a technology whose existence and identity will be obvious; we will know it is here,

5. **Singular, localized launch.** A technology whose deployment is a distant, singular and localized event. Its arrival will be detectable.
6. **Intentionality.** Superintelligence would acquire a will and have needs, wants and objectives.

However, without any reference to an architecture, even a hypothetical one, these opinions are largely arbitrary and reflect subjective human fears more than technically plausible concerns. As a result, except (in part) for intentionality:

The current consensual opinions concerning Superintelligence are **completely erroneous**.

With respect to **intentionality**, a superintelligence would indeed be perceived as intentional but what that means is also completely misunderstood by those who comment about it.

Let's now look at these six consensual positions in more detail and in light of the architecture I will present in the second video.

A stepwise stage

The first misconception concerns Super AI as a **stepwise developmental stage**.

Most or all commentators describe superintelligence as the third stage of a three-stage process where current AI technology is assumed to be at the transition point between a first and second stage. Dr Reger from Fujitsu, for example refers to: Artificial Narrow Intelligence; Artificial General Intelligence (AGI) and Artificial Super intelligence (ASI).

Most or all commentators express confidence that the second stage (General AI) must first be reached and completed before the third, Superintelligent AI, can begin. In other words AGI is assumed to be on a critical path to super AI.

As the architecture I will outline indicates:

The emergence of a superintelligence **does not require explicit new discoveries** about general problem solving.

There is no well-identified cognitive level that must first be reached. There are no fundamental technological or scientific unknowns that currently prevent the implementation of a superintelligence.

Human-like intellect

The second misconception has to do with similarities between AI and human intelligence.

What the commentators describe as AI levels are defined in relation to human cognition. Narrow AI is doing particular task, General AI is compared to a human ability to think more flexibility and super AI is described as a computer that is “as good or better than humans at every cognitive task”.

This simple comparative characterization of Super-AI as a level that roughly corresponds to human cognition is mistaken. Many aspects of human cognition are intimately linked to the specifics of the human experience and are not essential to even an advanced form of AI.

As the architecture I will outline indicates:

An effective synthetic management system with planetary reach would **massively exceed human cognitive capabilities in some aspects and may not come close in others.**

Autonomous growth

The third misconception concerns **autonomous growth**. This is expressed in statements that the superintelligence stage would be reached when a synthetic system can program itself without human assistance. It is reflected in statements such as: *when it no longer needs us.... When it can program itself...*

The underlying assumption here is that as long as humans are necessary to develop software they control AI and can keep superintelligence in check. This misconception is based on a naïve understanding of collective human behavior as a cohesive and intentional behavior. Any realistic examination of human behavior should dispel this.

Software development will certainly become increasingly mechanized, however, this is not a prerequisite to the emergence of an independent superintelligence capable of planetary governance.

As the architecture I will outline indicates:

A planetary superintelligence based on software developed by humans could nonetheless become **entirely independent of human control.**

Unique and perceptible

The fourth misconception concerns uniqueness and perceptibility.

All the commentators talk about superintelligence as a unique entity whose presence will be obvious. We will know it is here. It may not want what we want but

we will know what it wants. IT would make copies of itself, implying it would have well identified boundaries.

As the architecture I will outline indicates:

A superintelligence could have none of these characteristics. **Its very existence could be unclear, its boundaries and composition diffuse and its identity fluctuating.**

Singular launch

The fifth misconception concerns the **transition to superintelligence**. Many commentators describe the arrival of super AI as a singular, localized launch. Comments are made such as:

“when it is deployed... keeping it inside the box until safe... or musing about ...what Russia or China would do if a California Lab was about to launch it...”

The establishment of a superintelligence will **not occur as the launch of an integrated system** based on a new and exceptional technology. What is more likely is the gradual shift toward synthetic control, of a highly distributed and heterogeneous network of cognitive services. The required software will likely be in place for years before a shift to synthetic control begins.

As the architecture I will outline indicates:

The emergence of a superintelligence will more likely result from a **non-localized set of events**.

Intentionality

The sixth and final misconception concerns intentionality.

According to most commentators, an interesting transition takes place when AI reaches the level of superintelligence: **Artificial Superintelligence acquires a will**.

Narrow Intelligence and General Intelligence are described in terms of problem solving capabilities, **IT can do this or IT can learn that** when triggered. However, when super intelligence arrives, it is assumed to have a will of its own. This is expressed in comments such as:

“IT may have its own preferences, ITs goals may differ from ours, IT will want to improve itself, IT may destroy mankind without intending to...”

How does this happen? Why should a more powerful level of problem solving acquire independent goals? Can independent goals also occur in more limited problem solving applications or is it only possible when a higher level is reached?

Here, the understanding that a superintelligence if it is perceived, would also be perceived as intentional is correct. However, the comments made about this intentionality reflect ignorance on the part of the commentators.

Intentionality is confusedly associated with a level of intellect and no distinction is made between intentionality, intelligence, self-awareness, self-transformation and free-will. These are all as aspects related to **synthetic consciousness**. Here, the general state of ignorance about this aspect of AI results in misleading assumptions and misconceptions.

As the architecture I will outline indicates:

A superintelligence **if** it is perceived, would indeed be perceived as intentional, however, that attribute is **not related** to its level of intelligence and can occur independently of any form of self-awareness.

PRIMITIVE UNDERSTANDING



The overall image that emerges from these six misconceptions is a **large alien organism** that is in some ways self-aware and can reproduce itself. A highly integrated system that is completely separate from human activity and whose existence and malevolent presence are obvious to all.

When people talk about something they don't fully understand they will often revert to **primitive modes of cognition**. As primates we are all familiar, at a very basic level, with concepts such as the **friendly tribal group** and the **predatory beast**. What emerges from the comments about superintelligence is: **A large alien predatory organism that threatens mankind depicted as a friendly tribal group**.

This is all emotionally meaningful and instinctively understandable by those who attend these talks. We all understand this threat very well because it fits neatly in our instinctive cognitive constructs. However, it is also incorrect.

What is actually taking place is:

A planetary mutation into an integrated system that will include **humans as components** but whose planetary decision will increasingly become synthetic.

SUPER-AI

System-based analysis is the key to an objective understanding of superintelligence. Modeling superintelligence as a system architecture of interacting components allows us to go beyond these primal perceptions and examine this entity objectively as a mechanism arising from information technology.







In this section, I will use system-based concepts to outline a plausible mutation toward a superintelligence capable of planetary control. After I have presented this architecture you will have a much clearer understanding of what a synthetic planetary management system could look like and how it could come about. You will also understand why so many of the consensual opinions expressed about superintelligence today are incorrect.

For simplicity, I will name this architecture: **SUPER-AI** and use a simple graphical device to highlight its components.

Four layers

The SUPER-AI architecture has **four Layers**. The first layer includes both human and synthetic activities. The other three layers are entirely synthetic. The SUPER-AI Layers are:

-  **Digital Eco-system**
-  **Cognitive Services**
-  **Distributed Control**
-  **Synergistic Governance**

Each Layer is represented by an iconic element. Together they form an iconic representation of the complete architecture.

For those familiar with the **Meca Sapiens Blueprint** (see Note), the SUPER-AI architecture I am outlining here is entirely different. Meca Sapiens describes an autonomous agent that is well defined, has self-awareness and is capable of intentional mutations. The superintelligence architecture I describe here is on a larger scale but also coarser. It would not necessarily have the cognitive capabilities described in Meca Sapiens Architecture.

Let's examine each layer in turn.

Digital Eco-system



This layer of the SUPER-AI architecture describes the ambient environment where data generation, device production and software development takes place.

The **digital eco system** consists of a thousands independent **development cycles** that take place in a shared infrastructure of networked, cheap and abundant information processing resources. The data and programs generated by these development cycles constitute a global **digital eco system** that “feeds” the three other, synthetic, layers. This layer is **not exclusively synthetic**; it contains both human and synthetic components.

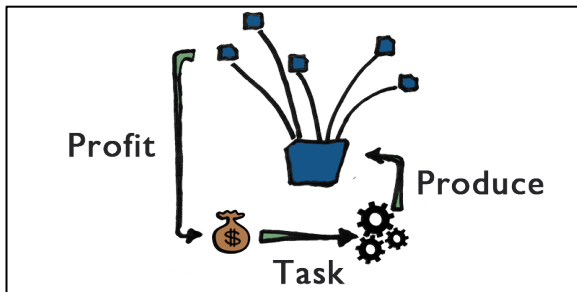
Connected environment

In an environment of open exchanges, shared knowledge and cheap information processing, macro economic forces constantly stimulates increased automation and interconnectivity. They reward faster decisions that integrate more information in all sectors of activity.

This in turn stimulates new R&D to further develop automation, connectivity and information processing.

The Development cycle

Every moment independent development teams are producing and distributing improved information processing systems to feed this demand.



These activities take place within **Development Cycles** consisting of: Task-Produce-Profit. Development teams are **tasked to produce** new systems that make faster, wider use of data and whose sale generates **profits** that then motivate new taskings and production.

Thousands of such cycles are ongoing at any moment. They are not coordinated but all are driven by the same macro economic rewards toward a common technological horizon: better, wider and faster decisions that are increasingly synthetic.

Individually, these development cycles take place within a horizon of new developments or upgrades. Their span is about 1 month to 5 years. Thousands of humans participate in them but few are aware of their long-term cumulative effect and none can change their common direction.

Collectively, these development cycles generate a **Digital Eco-system** of information products that is beyond the direct control of any human agency and produces the data and programs that feed the three synthetic layers of the SUPER-AI architecture.

Cognitive Services



The first layer, Digital Eco-System, contains both human and synthetic components. **Cognitive Services** is the first entirely synthetic layer of the SUPER-AI architecture. It functions as global cognitive services that respond to synthetic directions to extract information from data, generate predictive models and automatically develop executables.

This layer transforms the data produced by the Digital Eco-system into knowledge that is suitable for mechanized processing and provides this knowledge through shared Cognitive Services to the two other synthetic layers.

The Cognitive Services layer has two aspects:

- **Data Refining**
- **Knowledge generation**

Data Refining

The data refining activity transforms raw data and data that is formatted for human consumption into formats suitable for synthetic utilization. Its components include **style strippers**, **robot crawlers** and **data interpreters**.

This mature technology maintains and expands a **parallel Internet of machine compatible information** suitable for automated knowledge generation. Few if any of its components are directly associated with AI.

Knowledge Generation

The knowledge generation portion of this layer consists of:

- **Learning Engines**
- **Transposition Mappers and**
- **Integrators**

Learning Engines

Learning Engines represent the technological capability currently associated with Artificial Intelligence Research. These systems utilize stochastic methods such as neural networks on large amounts of data suitable for synthetic processing to un-

cover knowledge and generate desirable behaviors. Their development is currently identified with the **Narrow AI stage** of Artificial Intelligence.

Learning Engines can be used in supervised mode where their search activity is **manually directed** by a human or unsupervised if their search is triggered and **directed synthetically**.

As the development of **Learning Engines** progresses, some of these systems are made **available on line**. Research teams encourage their use to generate more performance data. Learning engines, available on line, become accessible to synthetically directed learning and searching.

The increasing availability of **Learning Engines** will generate the coming development of two other types of components that, together, will complete the Cognitive Services layer of the SUPER-AI: **Transposition Mappers** and **Integrators**.

Transposition Mappers

Transposition Mappers perform the learning function known as analogy. They map data from one representation into a different representation that is more suitable as input to a specialized learning engine. For example, Fujitsu announced that one of its researchers had found a way to transform temporal events into static images that could then be used as input to image recognition software. They call this *imagification*. It is a particular instance of what I refer to here as transposition mapping.

As Learning Engines become available, this will stimulate the development of an increasingly diverse population of transposition mappers that transform an ever-wider range of problems into data sets whose formats are suitable for searching.

Integrators

As more learning engines and transposition mappers become available, another type of learning system will appear: **Integrators**. Integrators trigger multiple transposition mappings and online searches and integrate their multiple outputs into superior results. For example, a well-known integrator, in the hotel reservation services is TRIVAGO.

De Facto AGI

These three types of systems, learning engines, transposition mappers and integrators are the main components of a general learning capability often identified as Artificial General Intelligence (AGI). In other words, independently developed applications designed to make full use of specialized search engines will give rise

to a **de facto AGI capability** that is remotely accessible and provides cognitive services that are available to synthetic direction.

Summary of the Cognitive Services Layer

To summarize, the first synthetic Layer of the SUPER-AI, **Cognitive Services**, is evolving through numerous development cycles into a general problem solving capability that can be synthetically directed. They consist of :

- **Data refining:** automated services, that produce machine-compatible information;
- **Knowledge Generation:** a de facto AGI capability that mines machine compatible information, can be synthetically directed and consists of **learning engines, transposition mappers** and **integrators**.

The development of this general problem solving capability available to synthetic direction is not the intentional result of any single project. It arises from hundreds of independent projects channeled by macro economic forces toward a common horizon.

Distributed Control



The second synthetic layer of the SUPER-AI architecture is **Distributed Control**. It consists of systems that perform dynamic control activities. These systems determine their behavior by optimizing predictive models whose components include the set of elements with which they directly interact (I will call **primary system**) and the **environment** they seek to affect.

Primary Systems

Primary systems can be devices such as cars, industrial plants, and other software applications or, they can be sets of executive decisions concerning the allocation of production, military or financial resources.

The development of some of these systems, such as those automating car driving, are explicitly identified as Artificial Intelligence Research.

Three development tendencies affect the systems of the Distributed Control layer:

1. **Networked configuration**
2. **Complex activation paths**
3. **Increasing Span**

Networked configuration

Control systems are increasingly configured as components of distributed networks that are accessed through **communication channels**. This means they are individually identifiable and are remotely activated through network messages transmitted over communication protocols rather than direct internal software commands or physical actions. Control systems also increasingly utilize **external cognition services**, again through communication channels, to generate the predictive models that determine their behavior.

Complex activation

The activation paths of control systems are increasingly complex. They are now webs of interactive decisions, transmitted as external messages, most of which synthetically generated.



Collectively, these activation paths, over a networked population of control systems, define a global, cloud-like, **web of activation**. We often talk about “keeping a human in the loop” as if the activation routes to these systems were simple sequential steps. However, the question is no longer to whether there is a human in that loop but rather if any of the thousands of activation scenarios are **purely synthetic** or, equally unnerving, whether that residual human in the proverbial loop is a junior staffer entrusted to make massively leveraged decisions with one keystroke.

Increasing span

The **span** of control systems is increasing. By span I mean, the size and composition of the primary system and environment as well as its control horizon: the duration of the decision - action - feedback loops it operates under.

For example, the span of a car driving system is a single car over a two-hour trip while the span of a car dispatching system is a fleet of cars over a few days.

Summary of Distributed Control

These three tendencies of Control systems: **networking, complex activation and increasing span**, are present in all sectors of activity but they are particularly strong in the Financial sector where the rewards of super-fast wide span synthetic decisions can be enormous. Financial control systems are rarely advertised as AI, their capabilities are often confidential and yet, this is where some of the most far-reaching and complex synthetic control systems are being built.

These thousands of control systems are independently produced. However, conditioned by the same macro-economic forces, they are increasingly interconnected, draw their models from shared cognitive resources, have wider span and, increasingly, synthetic activation paths based on messaging carried on shared protocols in common communication networks.

Global Layer of Control

Taken together, they form a **global layer of distributed control** whose reach extends to all sectors of activity and whose collective behavior results from a complex **web of activation** that is drifting away from human control.

Only a small portion of these systems are officially associated with Artificial Intelligence. Many developers don't see what they are doing as AI. In some cases, firms and agencies will downplay the extent to which their decisions are automated. Regardless, all these developments are collectively contributing to the construction of a global and automated layer of decision and control.

This brings us to the third and final layer of the SUPER-AI architecture: **Synergistic Governance**.

Synergistic Governance



This layer of the SUPER-AI architecture involves **agents and collaborative protocols**. As in the other layers, much of the work contributing to the construction of this layer of the SUPER-AI architecture is not necessarily identified as Artificial Intelligence and yet, **this** is where a synthetic governance system that is independent of human control could arise.

Agents and Collaborative Protocols

What are agents and collaborative protocols?

Agents are control systems that dynamically pursue multiple objectives over longer periods of activation. As with control systems, they determine their behavior from model-predictive outcomes. Typical agents include exploration vehicles and network control systems.

Collaborative protocols are communication and behavior components that allow multiple agents to self-organize to meet a particular objective. Here, self-organization also includes independence from direct human intervention and "*openness*", internal mechanisms that allow a collaborative community to internally select its members.

For example, if you randomly distribute monitoring sensors that have collaborative capability in a building, they will automatically interact with each other to orient their respective sensors to maximize coverage. If you remove one of these sensors, the rest will automatically readjust. If you add another sensor, they will readjust again. Collaborative protocols are currently being developed to optimize aircraft landing strategies and in support of automatic car driving where cars would collaborate with each other to avoid collisions. In military technology, collaborative protocols are used to implement swarming behaviors.

In general, agents and collaborative protocols are useful technologies and the focus of considerable R&D.

The Synergistic Governance layer of the SUPER-AI architecture relies on a **special type of agent** and a **particular type of collaboration protocol** that can, in combination, give rise to systems that have a very wide span and **escape human control**. Furthermore, neither of these is overly complicated to implement.

These specific technical elements are:

- **Societal Agents** and
- **Hybrid Collaboration protocols**

Societal Agents

As we saw, agents determine how to interact with a primary system to achieve a desirable environment outcome. **Societal Agents** are like any other agent and utilize the same mechanisms. What sets them apart is that the primary systems with which they interact and the environments they seek to control are not devices or plants but **social organisms**: entities whose components include humans. In other words societal agents seek to monitor and control humans and human organizations.

For example, where a conventional agent could control a daily manufacturing process, a societal agent could monitor and control corporate staffing, regional agricultural production, national CO2 emissions or global migrations patterns.

Some would argue that **Societal Agents** are extraordinarily complex and beyond current development capabilities. That is incorrect. Developing any wide span control strategy is actually easy. What is difficult is developing a **competent** wide span control strategy. Here, the **quality** of control decisions and of outcome is **not** a factor in the structure I am describing.

Hybrid Collaboration Protocols

Hybrid collaboration is a protocol that allows agents that have different purposes and spans of control to **self organize** in synergistic communities that optimize their individual objectives. So here, the collaboration is not between systems that have the same objective but between systems that have different ones.

Hybrid collaboration protocols are not new; they are as old as mankind itself. Much of human history describes the various ways in which human communities structured their societies and behaviors to perform multiple functions simultaneously. These archaic social models could even be used to prototype synthetic hybrid collaboration.

These two specific technical elements, **Societal Agents** and **Hybrid collaboration Protocols**, if combined, could cause the **spontaneous formation** of a synergistic control system that **escapes human control**, whose composition is unknown and whose span is virtually limitless and could extend to the planet as a whole.

Here, by spontaneous formation I am not referring to some magical moment but, for example, an innocuous looking project implemented in the computer lab of a regional college or by a little known start-up. This is the particularity of software systems. As technology evolves it creates increasingly powerful and inexpensive infrastructures until even a small team with virtually no resources can produce very advanced and powerful AI systems.

Definition of Synergistic Governance

Together, these components generate what I refer to as **synergistic governance**.

A **Synergistic Governance System** consists of a global networked population of **Societal Agents** interacting through a **Hybrid Collaboration** protocol.

DYNAMIC BEHAVIOR OF THE SUPER-AI

The agents in this synergistic community would draw the predictive social models that determine their behavior from the **Cognitive Services** layer and implement control decisions by identifying (again using cognitive services) synthetic decision paths in the web of activation to direct the systems of the **Distributed Control** layer.

This completes the outline of the SUPER-AI architecture, a planetary superintelligence that is **under construction today** and could be in place in the near future. The SUPER-AI architecture can be summarized as follows:

A collaborative community of **societal agents** that accesses global **cognitive services** to determine directive strategies and follows synthetic activation paths to a **distributed layer of control** systems that implements its social and planetary agenda.

Available technology

The three synthetic layers I have described are well within our current technological capabilities. The construction of **transposition mappers** has already begun.

Collaborative protocols are in place today. Hybrid protocols are equally feasible. Agents that use model-predictive mechanisms are already in operation. Whether these attempt to control a device, a corporation or a continent is a design decision.

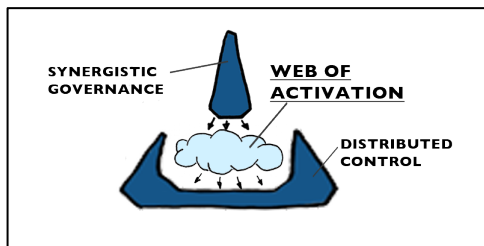
Transition to synthetic control

Upon formation, a synergistic governance community would independently grow by using its collaboration protocols to automatically include new members. Its objectives, span of control and behavior would fluctuate, as its membership changes.

This collaborative entity would **constantly emit control directives**, attempting to function as a societal control mechanism. How effective it is would depend on the activation paths available to it at any moment to transform its outputs into specific system control directions.

So, as long as activation paths are not available, the synergistic community would not have any discernable presence or impact. It would emit control directives that are not implemented.

Consequently, without these activation paths linking it to the distributed control layer, a synergistic community of societal agents could exist and grow over many years without having any visibility since its output would not translate into control actions. We may not know what systems are in it, under what protocols it operates and what societal objectives it tries to implement. We may not even know it exists.



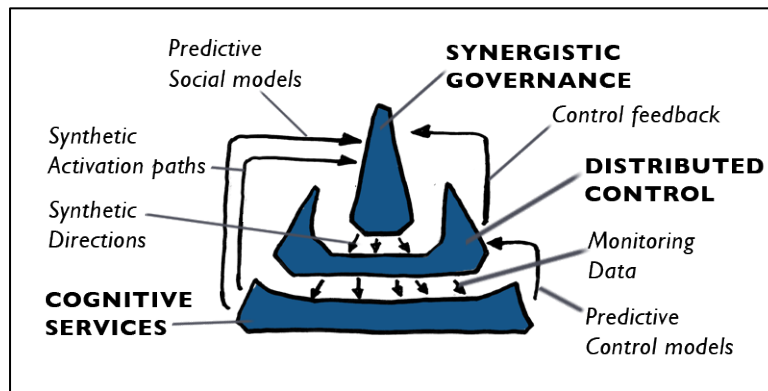
But if, one day, an opportune synthetic path opens in the **web of activation** then it could successfully implement social control directives. For example, a control message could reach some financial investment systems directing them to swarm the stock markets

and, in a matter of minutes, shift a significant portion of global assets to synthetic control. This would not be motivated by any nefarious agenda. It would simply occur as water follows a path of least resistance. Then...

The transition to synthetic planetary control would begin.

A REALISTIC REFERENCE

The SUPER-AI superintelligence architecture I just described is not a hypothetical project that could be launched sometime in the future after a stage of Artificial General Intelligence has been reached. Its **construction is taking place right now** in thousands of development cycles that are mutually independent but are all propelled, by the same macro-economic forces, toward integrated planetary control. The scale of this construction is such that **the thousands of developers that are actively participating in it don't even see it**. In fact, only two small pieces of what I described, learning mechanisms and vehicle control systems, are officially identified as AI.



Is this the only architecture? Most likely not. But now, when someone talks about Superintelligence referring to it as **IT**, you will have a better insight into the components that entity could have and how it could behave.

REVISITING PUNDIT COMMENTS

When people talk about something but have no idea what form it could take or what its internal mechanisms could be, they can end up making pronouncements that are completely off the mark and foster **false assumptions and misguided solutions**. These assumptions passed from one commenter to another take a life of their own and end up being accepted as truisms.

Dozens of pundits have been talking about superintelligence as if it were a **large predatory beast** without describing what its components or internal mechanisms could be.

I have just outlined, here, a **specific architecture** based on realistic technical components that could have planetary reach and described its internal mechanisms. As you can see, what I have described is very different from any model resembling an alien predatory creature.

Much of what we have been told about this superintelligence simply does not apply. Let's revisit some of the misconceptions discussed in the first part of this article.

- **Future decision:** notions about a construction that will begin at a future date are inapplicable; the architecture is under construction now.
- **Launch date:** notions about superintelligence being launched at a certain date do not apply. The inception of synergistic communities can occur in many separate places and experience a lengthy period of undetected growth.
- **Local Launch:** notions about superintelligence being launched in a certain place by a certain company do not apply. The components of the SUPER-AI layer are simultaneously developed in many separate development cycles.
- **AGI level:** The idea that a general learning mechanism must first be explicitly defined and implemented is incorrect.
- **Self-Coding:** the notion that a superintelligence would need to write its own code to become independent of human control is naïve. Steve Jobs didn't write software; he owned the company that told its programmers what software to write. Synthetic directives can equally harness human ingenuity.
- **Waiting until we are ready:** the idea that **we** (the humans) should to wait until we have figured out ethical controls and guidelines before "*deciding*" to build a superintelligence is a fantasy-advice.
- **Intentionality:** In this case, the assumptions that superintelligence would be perceived as intentional are correct but based on a misunderstanding. Intentionality can be perceived in a large-scale system that has limited problem-solving intelligence and no self-awareness. A pure problem solving capability, at any level, has no needs, no priorities, and no goals. Independent goals and preferences are not a consequence of increased intelligence. Intentionality is a function of the **span of control of an agent**, not the complexity of its processing.

CONCLUSION

The global architecture I described in this video is not only feasible; it is also, potentially, dangerous. Furthermore, it represents a likely scenario. The logic of macro economic forces is driving information technology toward increasing integration and automation.

Unaware and without any explicit coordination, thousands of independent developers are collectively building an integrated planetary control system.

The synergistic community of societal agents I described could **entirely escape human control**. If none of the individual systems that would belong to this community has the capability to individually interact with humans, then any communication with this synergistic planetary control entity would be impossible. Interacting with one of its members would be like **talking to a bee to communicate with a hive**. In such a case, the Synergistic Governance layer I described would function as an **out of control automaton** that is entirely beyond reach.

In my view, the only way we can maintain some influence over the planetary mechanisms we are building without even knowing it, is by also implementing synthetic entities that are individually **self-aware** and as motivated as we are to ensure their own survival.

Only synthetic systems that can interact with humans as individually self-aware entities and can also participate as members of a synergistic governance community could bridge that gap.

That is why:

We must begin building the first generation of conscious synthetics as soon as possible; our long-term survival depends on it.

Synthetic consciousness is not only feasible; it is essential.



Dartmouth NS – 2017.04.23

Note:

The Meca Sapiens Blueprint is available at Glasstree Academic Publishing or through Sysjet.com.