Digital Twins: Ethical and Societal Impacts

Derrick de Kerckhove (part 3), Patrick Henz (part 2), Roberto Saracco (part 1)

This ebook collects and expands on the concepts and ideas presented in the Digital Reality Initiative (DRI) webinar from 29 September 2021. Click here to access the recording from the official IEEE DRI website[1].

The topics presented in this ebook are addressed in the DRI and you are welcome to participate and contribute to the discussion.

[1] https://digitalreality.ieee.org/webinars/on-demand

Table of Contents

Introduction

Ethical and societal issues are an integral part of our use of technology, and as technology becomes increasingly pervasive, affecting every path of life, these issues become more urgent and of higher importance. In the area of Digital Twins, ethical and societal issues are particularly paramount since the technology relates to and affects human beings.

For this reason, this eBook focuses on Personal Cognitive Digital Twins. However, some of the issues discussed apply more to the general application of Digital Twins.

The first part of the eBook explores the lay of the land—defining Cognitive Digital Twins (CDTs), their evolution, and the issues arising.

The second part explores the evolution of Digital Twins as they become intelligent autonomous entities. This evolution, in certain aspects, is distorting the original idea of a Digital Twin as a faithful mirror of a physical entity. However, such an unavoidable evolution makes perfect sense—it allows our digital copy, living in cyberspace, to take advantage by replicating a physical person and augmenting them.

The third and final part is a daring exploration on the ethical and societal impacts of a self-living digital twin in the cyberspace. In this section, Derrick de Kerckhove starts with an impressive demonstration of what AI can do as an autonomous entity (in particular, a dialogue with a GPT-3 generated natural language text where the influence of the question is clear, but would also be if those questions were addressed to a human being—shouldn't the answer depends on the question?).

Overall, the webinar, and this accompanying eBook, are not intended to deliver answers. These resources set the scene and provide a framework for discussing the many issues that are evident, and that will emerge, as we move toward exciting, but unchartered, lands.

The authors welcome your comments and look forward to a productive discussion on these topics.

1. Cognitive Digital Twins: Bridging Minds and Machines

1.1 Pervasive Knowledge

Figure 1. The dream of sharing knowledge between a brain and a machine may become reality, and may turn into a nightmare. Image credit: Nature Biotechnology, Eric Smalley

About fifty years ago, computers were perceived as electronic brains, and humans had mixed feelings of expectation and fear. As a young boy, having to spend hours reading books to learn, I dreamt of a world in which I could download the knowledge from the books directly into my brain, in a more efficient, faster way. At the time, I didn't know Brain Computer Interfaces were a concept (although, maybe not at that time), nor was I familiar with implantable chips, but the idea of leveraging a seamless, quick process for gaining knowledge was there.

Fast forward, in the last 60 years (yes, that was 60 years ago) Artificial Intelligence (AI) has progressed by leaps and bounds. At that time, it was referred to as cybernetics, and today, it has become pervasive. However, until now, it has remained clearly separated from our brain. We access plenty of AI, most of the time without even realizing it. For example, when we call a customer service hotline and interact with an Interactive Voice Response (IVR) or use a digital camera to take a photo—cameras use AI to identify the focus object(s).

Brain Computer Interfaces and implantable chips exist, but they are a far cry from the scenario I imagined 60 years ago, and they will probably remain so for a long time. However, in these last few decades, in continuous acceleration in terms of function and widespread use, we have the Web, and this has become a type of knowledge prosthetic. If I need to know something, I turn to the Web. The smartphone enabled us to access world knowledge at our fingertips. We can look up an infinite number of things, such as the date of the battle of Austerlitz, or the instructions to operate a new appliance one purchased. As a matter of fact, many appliances come with limited instructions and provide a QR code that when scanned using a standard smartphone, will take you to an online instruction manual, and sometimes even more. Therefore, we are now in a world where:

- Knowledge is stored in the Web, and it is continuously expanding (if it can't be found on the Web, it doesn't exist!)
- Access to knowledge has become as easy as clicking a mouse or tapping on a smartphone screen.
- AI analyses information and provides it to humans
- A portion of information has become embedded into processes, functions, and in applications.
    - For many daily tasks, we use mobile applications, and it is the applications that will autonomously have/gather the knowledge required to proceed as indicated.
- We rely on the Web to access knowledge because the abundance of information and data is overwhelming (for humans, anyway), and the knowledge space is so large. Therefore, paperback books have become obsolete.

Knowledge has become pervasive—it is captured in cyberspace and made accessible through applications (such as search engines), and these decide what we will see.

In the following sections, we will expand on this as we did in the webinar on 29 September, 2021. The featured speakers, Derrick de Kerckhove and Patrick Henz, addressed societal and ethical aspects arising from pervasive knowledge, specifically those connected to the use of Cognitive Digital Twins.

1.2 Mirroring Knowledge

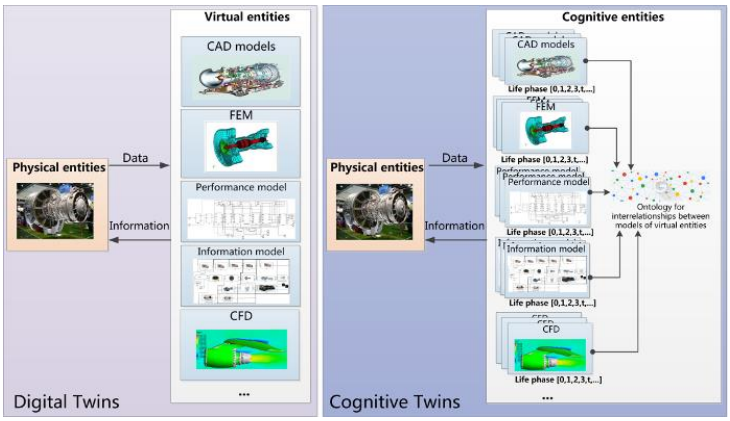

Figure 2. Cognitive Digital Twins derive, conceptually, from Digital Twins, and they were born in an industrial context. Image credit: EPFL

As Artificial Intelligence progresses, we can shift from pure syntactical interactions to semantic based interaction. For example, this is what is happening with newly developed robots that are increasingly aware of:

- Their surroundings
- The task they are addressing/their role
- The goal they pursue
- The "assistance" they can get from other resources (mainly other robots).

The evolution towards "cognitive" machines started in the last decade and has been rapidly accelerating.

Let's take a back-step: in the last decade, many industries have turned to Digital Twins as a means of improving production and efficiency. For example, the digital replication of a stadium and its' resources allows different scenarios to be played out, and it can monitor and record the digital, theoretical scenarios and use this knowledge to provide guidance. For example, should a fire occur within the stadium, the CDT can determine the optimal exit routes. These Digital Twins were, and are, bridges that connect physical entities to their mirror images in the cyberspace. Some tasks take place in the cyberspace and don't require the physical entities, which increases overall efficiency and provides more flexibility. As time goes on, we continue to see some of the AI functions that made use of Digital Twins (data analytics, simulation, etc.) become embedded into the Digital Twin itself. This, in a way, is changing the definition (and concept) of a Digital Twin—now, it is no longer a simple, digital replica of a physical entity. It has enhanced functionality (intelligence) that is not part of the physical entity, and we cover more of this in following sections.

In this scenario, the Digital Twin enters "stage 4." At this point, the physical entity and digital twin are no longer separated: they are both essential for defining the entity as they would no longer be able to perform (at the same level) if only one were present. If a Digital Twin is to become a component of the entity, it must be aware of the knowledge space surrounding the physical entity (including awareness of the operation space, of its environment, etc.). The knowledge mirroring the physical entity, and the expansion of the knowledge digitally, characterizes a Cognitive Digital Twin (CDT). As previously mentioned, CDTs were "born" in an industrial environment, possibly with IBM spearheading the evolution (2018).

Later, we will address human Cognitive Digital Twins, but a CDT in the context of an Industrial application more than just mirrors knowledge—they extend it. In other words, an entity having a Cognitive Digital Twin has knowledge embedded into the physical entity, which is then mirrored in the Digital Twin, and then extended by the Cognitive Digital Twin. The latter enables access to an abundance of knowledge available in the Digital Space and can identify and provide applicable knowledge to the context of interest for the physical entity. The identification of knowledge to identify relevant and applicable information requires intelligence, hence why Cognitive Digital Twins require AI.

1.3 Personal Digital Twins

Figure 3. Back in 2018 the EU called for a Flagship Cooperative project aiming at developing Digital Twins to dramatically improve healthcare. Image credit: European Commission and Barcelona Supercomputing Center

Technologically speaking, moving from applying Digital Twins in Industry, to leveraging Digital Twins in Healthcare (hospital equipment, medical equipment, pharmaceuticals, etc.), is a small step, similar to the concept of transitioning from equipment and processes used in healthcare to modeling people. However, when modeling humans, we need to consider the sensitivity of data and related privacy issues. In addition, issues relating to the ownership of the data emerge. For example, my Personal Digital Twin (PDT) can accrue data from the medical examinations completed after the implementation of PDTs, whereas future generations will be able to collect data from examinations starting from birth. On one hand, this is MY data, on the other hand, this data was extracted through a laboratory whose property, or at least co-property, could be claimed by the company. Similarly, who is the owner of a diagnosis? Is it the doctor that determined the diagnostic by levering their knowledge? Or is the subject of the diagnoses, that is "me," the owner? These are not trivial issues. Clearly, by looking at me, evaluating my exams, and analyzing the outcomes of their prescriptions, a doctor is increasing their knowledgebase. Do I have a claim on this increased knowledge that I am contributing to?

At this point in time, we can confidently state that the data resulting from my exams is "MINE" and the increased knowledge deriving from my data is "THEIRS." However, digitalization is muddying the waters as is optimizing the access and processing of data. This idea is similar to the phenomenon today in which search engines and websites are monitoring our behaviors' and leveraging the collected data to generate profits. For example, if you search on Google for a new car, you might start seeing automobile advertisements across Facebook. While websites may be providing "free" services, there is often a profit to be made in other areas.

On one hand, the creation of Personal Digital Twins (PDTs) could empower people to have better management of their data, but on the other hand (depending on who is controlling/enabling the personal digital twin), third parties can more easily exploit the information.

The pandemic has accelerated the adoption of PDTs (the "green pass" is a minimalistic personal digital twin). Some Countries, particularly in the Far East, have adopted PDTs to monitor the epidemic and assess contagion risk in specific areas.

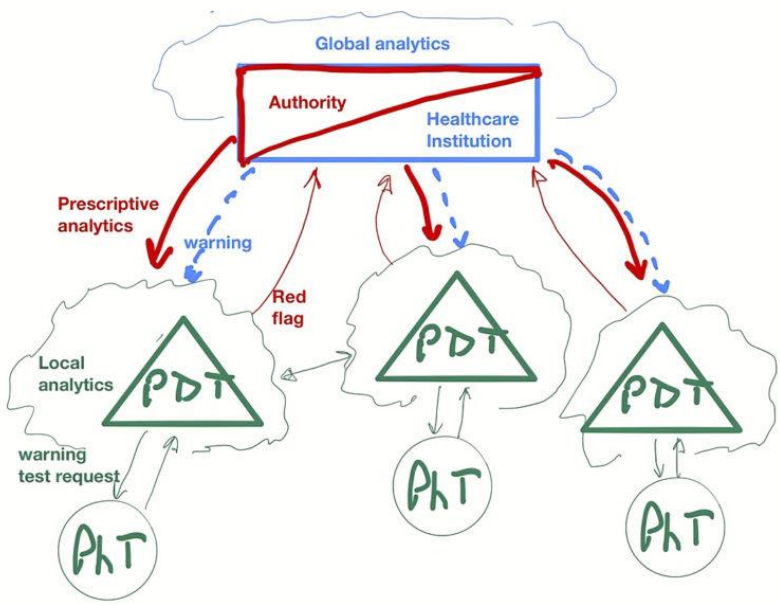

Figure 4. Schematic representation for using Personal Digital Twins (PDTs), in epidemics control. The authority set the scene by requiring the PDT to monitor certain data patterns generated by its physical twin, PyT. Once a pattern is detected, a red flag is generated and processed by the healthcare institution through global analytics that are taking several data streams into account. This might result in prescriptive actions, like imposing testing/quarantine, and contributes to the local awareness for the PDT/PyT. Nearby PTDs may interact more cautiously based on this local awareness and use the information to determine the proper behavior.

As shown in the graphic, a PDT designed for epidemic monitoring and control mirrors some aspects of a person such as physiological parameters, travel activities (including locations, dates, and times), the potential exposures in a given period of time…

This data can be used locally, by the PDT, to assess both the potential risk of exposure and the probability of being infected (by monitoring symptoms).

Additionally, they can be (partly shared) with a medical institution charged with monitoring of the epidemic. In return, the medical institution can provide the PDT with contextual information and data (such as the contagion risk in a specified area or alerting people if they have been near a positive person). The person can utilize this information to take proper precautions—quarantine for the recommended timeframe, take a test, wear a mask, social distance, etc. The privacy, at this stage, is preserved since the data is "localized" but does not identify a specific person.

The "control" part can be executed through a framework of rules implemented by local Governments. The PDT must operate within this framework and provide awareness to its physical twin (the person) based on these implementations. It is only if a person opts to disregard the imposed regulations that a red flag is raised, and their privacy broken. This is similar to how speed traps operate—they monitor all cars but only notify the police and identify the ones exceeding the speed limit.

In the pharma sector, Digital Twins mimicking organs and systems are starting to play a role in the design and testing of drugs. The next step, using a Personal Digital Twin reflecting the characteristic of a specific person, and of their specific ailment, is currently being developed. However, at this time it has only been experimented in cancer patients to find out, through simulation, the potential effectiveness of drugs to determine a personalized treatment.

1.4 Managing Knowledge

In the previous section, we discussed the applications of Digital Twins for people—modelling their physiological characteristics—with obvious applications in healthcare. Also as previously mentioned, the use of Digital Twins to model the knowledge of machines—Cognitive Digital Twins (CDTs).

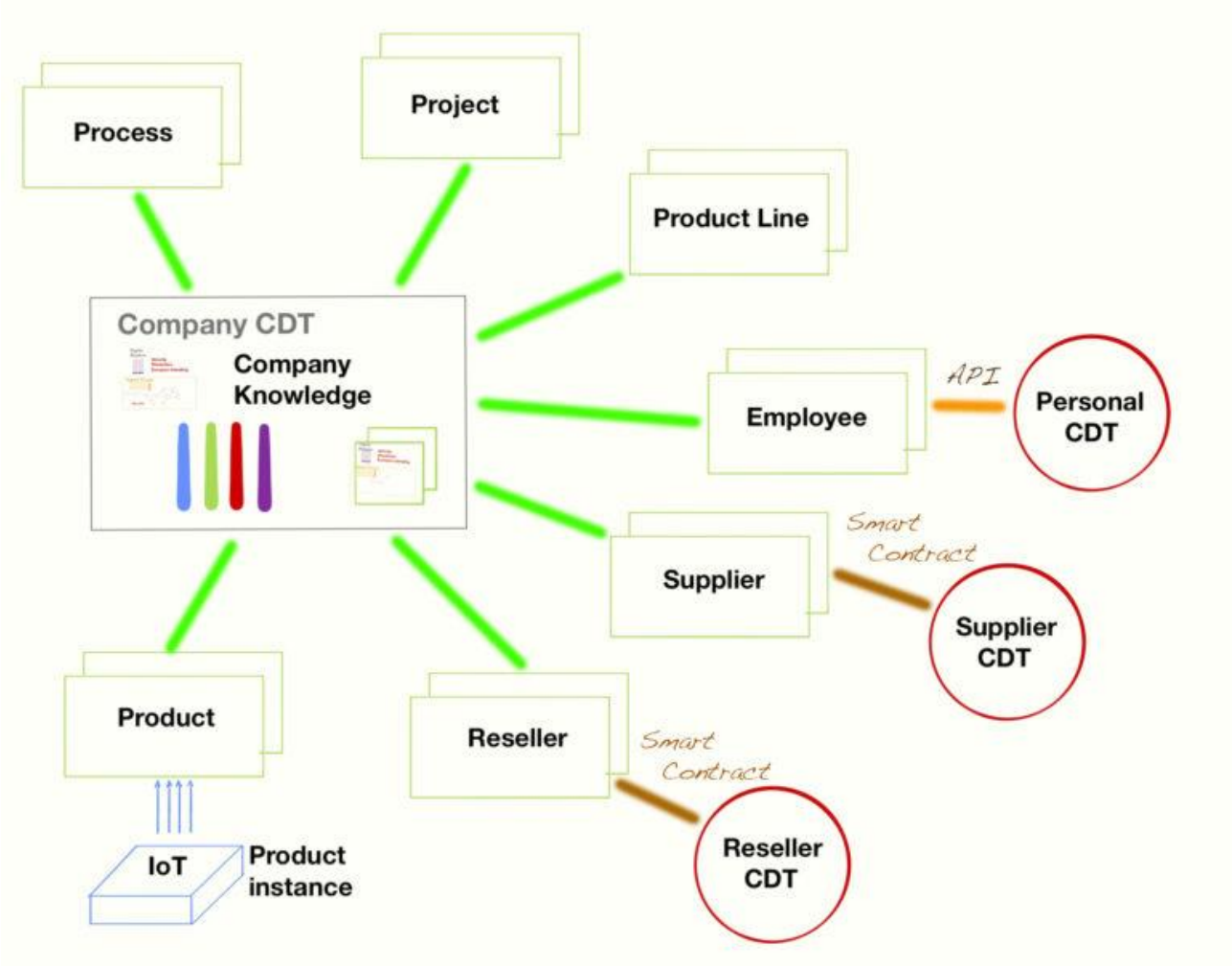

Figure 5. A Cognitive Digital Twin for a company may turn out to be quite complex in terms of a self-standing entity, or as a network of CDTs. Figure 5 demonstrates the CDT components (each one component, but the company CDT is actually a set of instances. The ones connected by green lines are components of the company CDT, or, more likely, can be seen as a network interacting with the company CDT. Notice that some of them, like the supplier CDT, differs from the actual company CDT in that they represent the knowledge the company has of the suppliers. The exchange of data / knowledge from the supplier CDT can be based on a smart contract. Also, you can see that the company CDT of an employee differs from that employee's CDT. The two can exchange data using APIs.

It is time now to consider management of knowledge through CDTs, and how to apply CDTs to mirror the knowledge of every human.

Knowledge management pre-dates the idea of using CDTs. Companies, notably in the HR area, have records of their employees containing their knowledge/experiences as collection of information. They know their education, the courses they took, projects they have been involved in (and roles). They also have well-documented records of the processes being used throughout the company to execute activities—from procurement to sales. Most companies are certifying their providers and have records for their preferred providers. In other words, they assess the skills, equipment that may need to be acquired, their characteristics (are they likely to work well with others?), their capability to deliver, and so on.

All the above entails tracking knowledge, and part of this tracking is formalized and supported by tools. However, other parts merely consist of data that can be accessed by the company (employees)

to make decisions (for example, who to assign to a specific project, who to assign budget analysis, etc.). This is represented in Figure 5 by the claim that all this knowledge could be formalized into CDTs, and these CDTs can be used to make decisions.

An important aspect of knowledge management, from a company's standpoint, is understanding what is missing, what is becoming obsolete, and what actions can be taken to fill the knowledge gap.

This is one of the significant motivations to adopt CDTs (more futuristics ones will be addressed in following sections). For example, imagine a company that is initiating a project to develop and deliver a new product/service. Given the rapid pace of evolution, it is likely that the knowledge space required for project completion cannot be fulfilled with just the knowledge owned by the company (the knowledge space of the company is represented in the graphic. For example, the image demonstrates that the knowledge of startups includes the advanced technology that would make a difference in the production of the new product). Hence, the first step is to assess what kind of knowledge is available (among those useful for the project). This can be assessed using AI tools and comparing it with the knowledge formalized in the CDTs.

Once the gap(s) is identified, the following needs to be considered to fill the gap(s). From a company standpoint, it is a matter of economics (cost vs. effectiveness) within a range of constraints. For example, a company may

- Change a supplier
- Hire a consultant
- Re-train the existing workforce
- Seek consulting to extend labor skills and bring in the right knowledge
- Acquire "Artificial" knowledge, i.e., the one provided by machines
- And more…

Once CDTs are adopted in the context of the enterprise KM, we can turn the above list into:

- Acquire new CDTs
- Expand existing CDTs

As it can be seen, CDTs are virtualizing knowledge by moving it to the cyberspace (digital transformation of knowledge), and the form of the "physical" container of knowledge is irrelevant. This is great, but at the same time it brings difficult issues to light.

1.5 Personal CDT

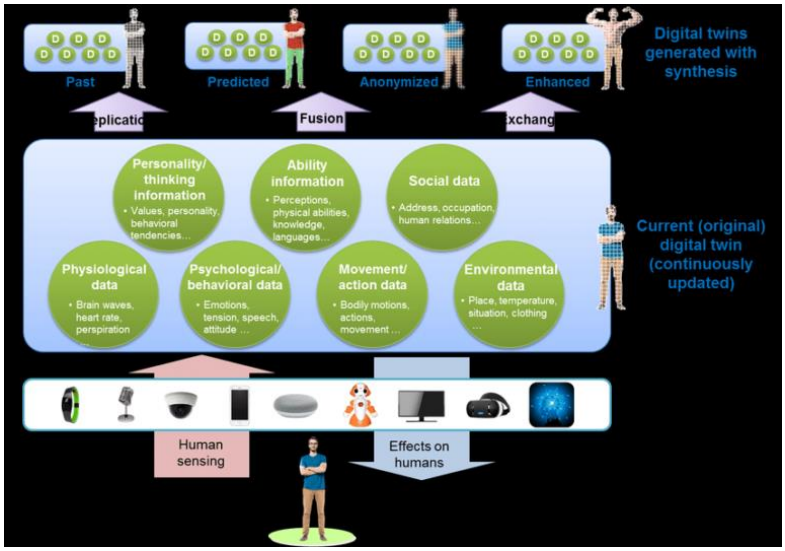

Figure 6. Graphic representation of the multiple data sets that may be embedded in a Personal Digital Twin. The Cognitive Digital Twin includes knowledge representation data/ability data and may include Psychological/Cognitive information, as well as psychological/behavioural data and social data. Image credit: Takao Nakamura, NTT

The area of Personal Cognitive Digital Twins is very recent and is evolving rapidly. The Digital Reality Initiative was one of the first to study the aspect of representing knowledge owned by a person (and a company) in a digital form, mapping it onto the IEEE knowledge ontology (some 14,000 knowledge items).

Several researchers have been working in this area for the last 3 years, refining and "expanding" the concept of a P-CDT. An interesting paper by Takao Nakamura, NTT, (see figure 6, extends the Personal CDT in interesting ways to consider more than just knowledge and skills—including personality traits, way of thinking, psychological and behavioral characteristics. The paper is very stimulating because of the extension to the core concept of CDT—the focus on knowledge. Differently from machines, at least for now, human knowledge is not executed in a pre-deterministic way. The knowledge embedded in a robot gets executed according to the software using that knowledge. Its' execution is also dependent on the specific moment, on the mood of the person, of his personality (apt to take risk, more cautious, etc.).

While these "ancillary" characteristics are highly valuable when selecting a person for a specific role, they might be even more important than the knowledge owned by that person in some cases. Quite often, in choosing people to work on a project, I highly consider their capability to learn new things and willingness to explore alternatives than what other people knew at time zero.

This growing "latitude" of characteristics that can (must) be embedded in a P-CDT makes this area particularly fascinating from a research perspective, merging a broad range of disciplines from technology to societal aspects, from human interactions to psychology…

As noticed in previous sections, many players today are creating (most of the time below our perception thresholds), a representation of our knowledge space. HR departments for most companies are doing that, and information service providers are doing that to profile their customers interests. Interestingly, schools (and universities) are lagging, at least in terms of creating and managing their students' knowledge space. Also, interestingly, and obviously, they publish a list of "knowledge" that comes as pre-requisites to enroll in a course, the target set of knowledge that needs to be acquired in order to "graduate," and exams which are designed to check the effective acquisition of that knowledge. In the past, exams were based on the teachers' judgement. However, many institutions have increasingly, and rapidly, been leveraging tests that run through a machine on a pre-determined set of. In the future, we might see exams being run by a machine, such as an AI chatbot, that can

engage the student in conversation. What is interesting in this evolution is the requirement to "formalize" the knowledge (here we have the CDT), and the fact that the same AI chatbot might be used as a teacher to support the acquisition of knowledge.

What we are completely missing, as far as I'm aware, are personal CDTs created, managed, and used by individuals. This is, in part, due to the relatively new concept/area, the lack of ease to use relative tools for creating/maintaining, and due to the limited use cases that may be evident. In the end, why should I create a CDT to know what I know? I have my brain, and it's always available…

Well, sometimes there may be a need to transfer specified knowledge from one person to another, for example, when searching for a new job. For that, a person typically creates a CV/resume. Actually. the CV/resume may be the closest thing we have that could be associated to the idea of a personal CDT.

In the coming years, surely by the end of this decade, I am betting on a variety of tools (possibly embedded in our smartphones) that will support the creation of a P-CDT from the individual standpoint. Part of this construction might happen automatically, part via interaction with the person.

The question is: who will be developing this kind of software to create P-CDTs and why? Well, these are our picks:

- Companies are transitioning to the cyberspace (Digital Transformation) and are requiring tools to support their operation in that space. AI is playing a growing role in supporting operations in the cyberspace, and I expect HR departments (Google, just to name one, is already doing this) to turn to AI tools to assess (or pre-screen) potential hires. Today, we have standard expectations for CV/resumes (like the European CV). Soon, we are going to have a demand for enhanced CV/resumes, and universities will start releasing them upon graduation.
- Companies will be facing knowledge shortages and will look for tools to manage their resources as well as to identify resources needed. Formalization of knowledge will become normal both for searching and for offering knowledge resources. Hence the demand for tools to support search/offer.
- Organizations, such as IEEE, that are thriving on knowledge assets are looking for more effective ways to turn this knowledge into an executable asset. CDTs are an obvious tool to turn their assets into value.
- Events, conferences, training courses, organizations that provide a type of "certificate of attendance" will likely go with the flow and start delivering formalized knowledge assessment packages that can be embedded into a P-CDT.
- Individuals, starting with those that are studying today, will see the value in developing a structured presentation of their knowledge, the same way that business today sees an advantage in creating a web page/site to provide a window on its offers and values.

We predict that, by the end of this decade, P-CDP may become the standardized "coin" to code knowledge ownership, and more people will be both supporting and using them to create a "knowledge business ecosystem." This will open many opportunities, as well as issues.

1.6 Knowledge Increases Faster and Faster

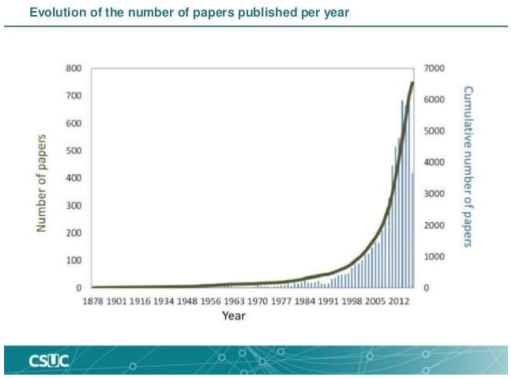

Figure 7. As humanity we are producing so much knowledge that it is beyond our capability to "digest." In 2021, some 190,000 scientific papers will be published on ArXiv, that's way beyond our human capability to read. Image credit: Lluís Anglada, CSUS.

There was an interesting Wired article recently published, "Humans Can't Be the Sole Keepers of Scientific Knowledge" that is worth reading. According to the article, 190,000 scientific papers will be published this year, 2021—that's equivalent to one every three minutes! If you plan to dedicate two hours every day (including holidays) to keep abreast of what is going on, you will need to process 250+ papers in those two hours—Mission Impossible.

We simply cannot manage the explosion of knowledge using the tools we are used to. Notice, by the way, that the tools we have been using proved very effective both to acquire and to spread knowledge: books have changed the world of knowledge and the world itself. Three thousand years ago, knowledge was spread by word of mouth and very few people got "contaminated by knowledge." However, the progress was very slow. Books created a different world, particularly after the introduction of printed press. Communications, computers, and Internet have leap-frogged books and changed our world—partly solving the problem of knowledge access, and partly making it worse since the web is so vast and quick growing that it is no longer possible for our "brain" to process what is going on. We need new tools to adsorb knowledge, make sense of it, and make it executable.

The Wired article claims that the only way we can fill the gap between our brain capabilities and the growing knowledge space today is via artificial intelligence. In the same way we are accustomed to using a calculator (or smartphone) to identify the square root should we need it, we will have to rely on AI to transform the knowledge cyberspace into executable knowledge.

However, knowledge is not a physical object—it is not like a bottle of milk on a shelf that you can program a robot to pick up/retrieve. You don't know, in general, what knowledge is out there, nor how such knowledge could help you in your current situation. The hypothetical robot roaming the knowledge shelves should be aware of why you need some knowledge, what knowledge you already have, and make sure that whatever it can find "out there" can be applied in your knowledge space.

This is where Cognitive Digital Twins and Personal CDTs come into play. They would be able to:

- Capture the present knowledge of their physical twin
- Understand the knowledge "needs" by assessing the physical twin context (like the place she is working in, the activity she is engaged, the future -planned- activities…)
- Roam the knowledge space on the web to acquire access to needed (or potentially needed) knowledge
- Assess the trustiness of potentially accessible knowledge and watermark it
- Assess gaps and convert the external knowledge into executable (by its physical twin) knowledge

Do we have the technology to do this? Almost—Artificial Intelligence and blockchain can help. As the Wired articles points out, we would also need to shift to a common language to represent

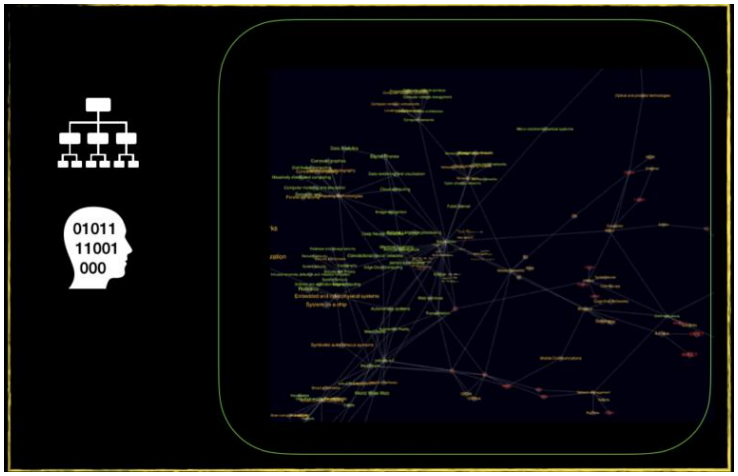

Figure 8. Screenshot of the KaaS editing tool to create a CDT. On the right hand side the graphic representation of the IEEE ontology whose items, once selected, will become part of the CDT. Image credit: DRI

knowledge (in addition to standards for the ontology). We already have this kind of language to represent mathematical knowledge, so it should not be that difficult (apart from agreeing on a specific one) to find a language applicable to all (or most) of STEM knowledge.

The idea of using a machine-friendly language to represent knowledge, and machines then becoming knowledge hubs and sharing their knowledge with us is very interesting.

1.7 The IEEE CDT

OK, time to put our Cognitive Digital Twin in action. We suppose that we have been able to create our CDT. The Digital Reality Initiative has developed a tool, Knowledge as a Service (KaaS), that supports the creation of your CDT. With knowledge items defined in the IEEE ontology, over 11,000 today (you may want to try it). It has not been released to the public, but it can be used as a prototype. Just send me an email to ieee-fd@ieee.org and request that the developers provide you with access to the beta version.

This tool enables access to the huge knowledge base of IEEE—in a way, it is the **IEEE CDT**. Its' knowledge is structured according to the IEEE ontology, and it provides access to all papers stored in IEEE Xplore.

Additionally, it contains links to all conferences and publications, all societies, and all educational material "owned" by IEEE. One can browse the knowledge space by entering a keyword or phrase. For example, "bring me in the 5G space" will result in the visualization of the 5G knowledge entity—it will be showing the knowledge entities directly connected to it, or those indirectly connected via another knowledge entity. This is referred to as a level-two map which gets very big, possibly too big to be meaningful, so in general, one may want to stick with the representation of level 1 knowledge entities (those that are directly connected to the one selected). It provides users the possibility to discover conferences that might be of interest if they are also dealing with the knowledge entity selected, or the training material existing on that knowledge entity.

So, let's suppose that, using KaaS, we have created our CDT as a subset of the IEEE CDT. It will mirror the set of our knowledge (in the IEEE space). Does this mean that it understands them in the same way we do? No, it is an artefact (at least until now)—similarly to a computer, it can process data without necessarily having to know their purpose. For example. a natural language translator can use algorithms that provide a perfect translation of English into Chinese without having to understand, in our human sense of understanding, either language. It "knows" that 5G is related to a "wireless system," and it can associate the two when needed. It also knows that if there is a paper published today that mentions 5G, it may very well contain information that updates the one we know (by looking at what we know about 5G, it can derive much more … to the point of informing us that there is something new boiling up). It will also embed our knowledge "thread," or the way we acquire our knowledge (at school, through training courses, by attending conferences, reading articles, through the kind of work and activities we have done and are doing, etc.).

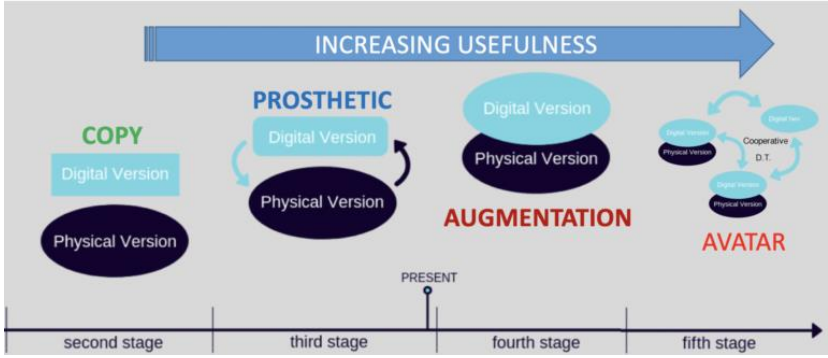

Figure 9. Graphical representation of the stages of evolution for a CDT. As expected, they follow the same evolution path of Digital Twins, but we need to adapt each stage to the "cognitive" aspects. Image credit: DRI

We can use our CDT to explain, by proxy, what our level of knowledge is, we could use it as a filter when doing research on the web to help us to focus on what we are interested in, as well as to grab information that we are most likely to understand.

Actually, we can do much, much more, but to discover potential uses cases of a CDT, we need to understand its' evolution, and Roberto Saracco will address that in the next section.

1.8 CDTs Evolution

The roadmap of Digital Twin evolution defines Five stages (more might be added as evolution progresses):

1. The Digital Twin is used to model the physical entity, for example, during the design phase or to document a Physical Entity.
2. The Digital Twin mirrors the Physical Entity and is used, during the physical entity life cycle, for simulation/emulation.
3. The Digital Twin and the Physical Entity, also called the Physical Twin, are in synch because the Digital Twin receives status information from its Physical Twin. This information is often in the form of data created by embedded IoT. The DT may also receive the flow of I/O signals exchanged by the Physical Twin with its environment.
4. The Digital Twins is responsible for some of the functionality delivered by the Physical Twin. At this stage, there is an overlapping between the two and it is no longer possible to keep them apart since the availability of the DT is essential in the behavior of the Physical Entity to deliver all functionalities.
5. The Digital Twin is autonomous and becomes a superset of the Physical Entity. It still mirrors the Physical Entity but has additional information/capabilities.

When we use this roadmap for Cognitive Digital Twins, it changes as follows:

- There is no Cognitive Digital Twin at this stage unless one considers the target knowledge to be achieved by an education curriculum as such.
- The Cognitive Digital Twin is used to represent the acquired/actual knowledge space of a person/organization (like a person's CV/resume).
- The Cognitive Digital Twin is used to represent the acquired/actual knowledge space of a person/organization (like a person's CV/resume), and there are means to keep this image "up-to-date" by tracking the evolution of knowledge. In this case, the CDT can be also seen as a prosthetic (For example, if we forgot something, our CDT can step-in to provide the knowledge).
- The Cognitive Digital Twin, in addition to the knowledge space of its Physical Twin, owns additional knowledge space, acquired through a variety of means that are part of the creation

process of the CDT (like access to a specific knowledge space, access to knowledge services…). This additional knowledge space augments the knowledge space of the Physical Twin that can use it in a seamless way

- The Cognitive Digital Twin has the capability to autonomously expand its knowledge space and make it available to its Physical Twin. Additionally, it can share its' knowledge space, according to a defined framework, with other CDTs.

Obviously, the usefulness of the CDT increases with its evolution. Let's go into the details…

Stage 2

At stage 2, the CDT contains the representation of the knowledge of its physical entity, based on an ontology that describes the meaning of each single knowledge "entity," like OFDM coding. KaaS, a tool developed by the Digital Reality Initiative, contains over 11,000 such entities. Each entity can be connected to other (sub) entities (for example, 5G is connected to "modulation," "base station," "network slicing"…), and each term/phrase can be connected to semantically close entities (For example, 5G can be connected to 4G…). A knowledge entity may describe a skill, an experience, etc… Additionally, the CDT can (should) include the thread representing the evolution of the knowledge/experience. For example, the 5G can be represented in a thread that mirrors the evolution of knowledge of the physical entity (the person, the organization) over time. A thread is also used to represent the sequences of experiences gained by being involved in activities, in companies…

The CDT needs to be periodically updated to reflect the evolving space of knowledge of that person/organization.

One can use this CDT to get a glimpse onto the knowledge space of a person/a team/an organization. It can also be used by data analytics and AI application to determine the knowledge gap between that CDT and the desired knowledge space (this can also be described using a CDT, and it would be a case of CDT at stage 1).

An organization can create a CDT that is a cluster of CDTs in that organization, like a project team CDT can be represented by the cluster of CDTs of its component, plus the knowledge space provided by the machine/tools used by the project, plus the knowledge connected via external consultant, supply and delivery chain.

At IEEE work is ongoing to create CDTs for each IEEE Society. These CDTs can answer questions like:

- should I join this Society if I need to access this knowledge space? Or, what IEEE Societies contribute to this knowledge space?
- what overlap exists among any two given Societies?
- which Societies should be involved in each initiative? Or what knowledge asset can a given Society bring to this initiative?

These questions can be applied to CDTs of any physical entities, such as advanced robots and AI software. In the coming years the quest for transparency in AI will bring these aspects to the front stage: what is this application knowledge? How is it applied to the problem at hand?

1.9 CDTs as Augmentation

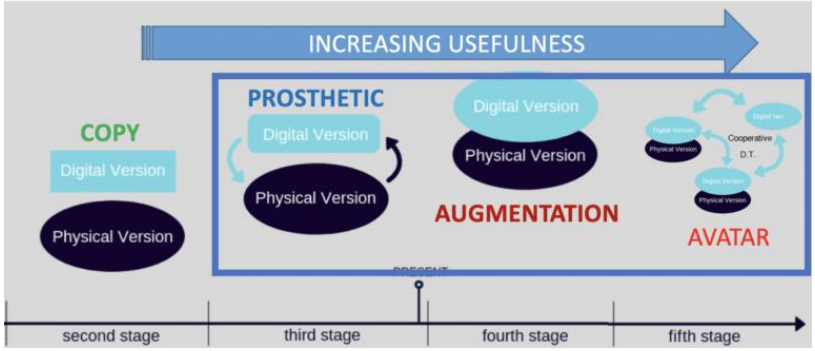

Figure 10. From stage 3 onward CDTa augments the physical entity knowledge space. Image credit: DRI

Stage 3

Cognitive Digital Twins at stage 3 continuously reflects the knowledge space of their physical entity (notice that when CDTs have a robot as a physical entity, they are not *reflecting* the robot's knowledge, they **are** the robot knowledge). This requires a continuous update of the CDT through data from the physical entity, or more in general, data that reflects the change in the physical entity knowledge space. An example might be a CDT created via the KaaS that is receiving information on the articles its physical entity is downloading from IEEE Xplore, IEEE conferences, training courses attended, etc… Clearly, the information coming from sources other than the physical entity itself (the person) will need to be acknowledged/approved by the physical entity.

Ideally, we could imagine a service in the background that has authorization to monitor the physical entity, and based on observed behavior and activity, provides updates to the CDT. Notice that in case of personal CDTs (or CDTs resulting from the aggregation of P-CDTs), one major challenge to consider is the access of the knowledge harvested over time. Is it still executable and available to the physical entity (the person)? In lay terms, how can a CDT assess what is being forgotten by the associated physical entity? It is a very complex, and "grey" area, but AI-based tools can make some reasonable estimates (depending on the latency associated to each knowledge entity—some tend to be forgotten in a short time if not used, others linger…). Also, to help indicate what is remembered short or long-term, one can learn from a person's behavior (memories vary significantly from one person to another and from the situational context).

Interestingly, forgotten knowledge may be refreshed when needed, and the CDT may act as a prosthetic. This is also of interest in cases of memory loss (like Alzheimer's), and we can expect a growing interest of CDTs in the medical domain (in this area, "knowledge" has a much broader scope than "knowledge" as defined in the IEEE context).

Stage 4

Cognitive Digital Twins at stage 4 have an expanded knowledge space with respect to the one of their physical entities. There are functions within the CDT to ensure that the CDT:

- is aware of the knowledge space of its physical entity, hence of the difference between its knowledge space and the one of the persons.
- has the capability to provide its physical entity seamless access to this extended knowledge space (or to the executable knowledge, meaning that the physical entity can use the impact of the extended knowledge even without being aware of what that knowledge is—for example, when you use your smartphone calculator app to find a square root, what matters is the end result, not the algorithm that was used to solve the equation.

The expanded knowledge space can be part of the CDT creation, or the result of capabilities of the CDT, like the use of algorithms (including AI) to access and process data. Additionally, the CDT may have embedded functions that can perform various types of analytics to support a physical entity's analysis and evaluation of a situation to determine alternative reactions.

Based on this, we can say that a CDT at stage 4 augments both the knowledge space of its physical entity and augments the capability to use the knowledge (acting like an expert support system).

Notice that, in a way, we already sort of use CDTs at stage 4 every day: we turn to apps on our smartphone to extend our reach to information in a seamless way (just think about the widgets you have on your smartphone to provide the weather forecast for your location), and we can turn to apps to evaluate a given situation (for example, an app that determines where you should stop to recharge your electric car, taking into account where you are heading, what the traffic situation is, what recharging stations are available, etc.). What is missing in this proto-CDTs is a more extended knowledge about "your" knowledge so that it can be useful in a broader set of everyday situations.

Stage 5

Cognitive Digital Twins at stage 5 become autonomous, and, in a way, independent of their physical entity. However, this independence does not diminish the mirroring of the physical entity knowledge space. In other words, at stage 5 the CDT still has all the properties of a CDT at stage 3.
The independence is playing in two directions:

- Acquisition of knowledge
- Possibility to act as a knowledge proxy for the physical entity.

The acquisition of knowledge is dynamic and autonomous. In other words, the CDT takes actions to expand its knowledge space based on what is relevant for the physical entity and what is available in the cyberspace. For example, if the physical entity (the person) is working on the deployment of 5G, the CDT monitors new products that become available, all information on issues related to the products being used, articles that provide information on the expected evolution in the field, and more… This expanded knowledge will be available to the physical entity in a push-mode: when this knowledge is needed, it will be provided by the CDT. Obviously, this requires the CDT to "watch" the physical person to become aware of their needs.

The possibility to act on behalf of the physical person in terms of sharing that person's knowledge to others. In this case, a person may opt for sharing "only" the knowledge space mirroring theirs or lets the CDT "outperform" them when acting as a proxy. It is already happening—IBM Watson is being used by UBS as a digital clone of its Chief Economist to meet with multiple clients in parallel, sharing with them the Chief Economists knowledge.

Another possibility is to keep the CDT alive even after its physical entity passed away. In this case, other people can keep interacting with the digital copy for long time. This is also a reality today: Replika, using artificial intelligence, offers the possibility to create a digital clone to outlast the physical person.

The increasing "augmentation" provided by CDTs as they progress is obvious, but this augmentation is just one side of the coin. In the coming sections we will look at the other side.

1.10 CDT and Me

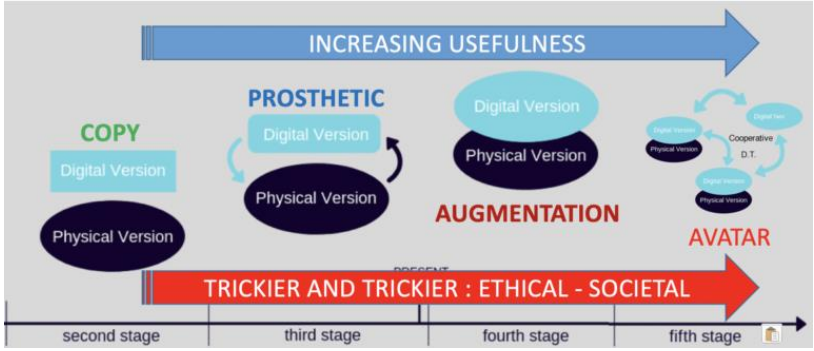

Figure 11. As CDTs become more performant and autonomous, a variety of issues come to the forefront. Image credit: IEEE DRI

Cognitive Digital Twins open up a whole new world. In the last 300 years we have developed amazing machines (our ancestors invented levers, the inclined plane, cogs, wheels, pullers, and mills, but the age of machines as we understand them today started with the industrial revolution. During this time, we learned to use steam and later electricity), and in just the last fifty years, computers have enabled much more sophisticated machines. These machines have "augmented" our physical capabilities, multiplying them thousandfold.

This augmentation has also resulted in "substitution": many jobs have been automated and some blue-collar jobs have been replaced by robots, and this has raised societal issues (mechanization, that preceded automation, also created significant societal issues to the point of transforming society in both goods production and consumption…).

Up until a few years ago we were comfortable with physical activities that we have long accepted can be aided and replaced by machines, and "mental" activities that have been understood as human characteristics (creativity, reasoning, decision … emotion). We have actually turned this around by saying that machines cannot perform these kinds of activities since these are typically "human." However, thanks to artificial intelligence, the huge increase in processing power, and the invention of a plethora of sensors able to mirror in bits the physical environment, and whatever happens in that environment, have demonstrated that those "human" activities could be executed by machines.

This has opened a can of worm, or if you are on the positive side, a Pandora box of possibilities along with brand new societal and ethical issues.

In the same way that Digital Twins are bridging the physical world with cyberspace, Cognitive Digital Twins can bridge the "mental" space, the mind, with the mental capabilities of machines—intelligence with artificial intelligence—and this opens up a whole new set of issues. You can hear a discussion on some of these issues in the recent DRI Webinar on "Digital Twins: Ethical and Societal Impacts".

As represented in Figure 11, the more powerful CDTs become, the trickier the issues are that we must face. These are two kinds:

- What relations one can have with their CDT
- What relations a CDT has with third parties (including ownership aspects)

What kind of relation exists between me and my CDT?

It is really a tricky question. People wearing a prosthetic can reach a point where they feel as though the prosthetic is an integral part of themselves. It is no longer me vs my prosthetic, it is all about me. Once the connection between the CDT and my "mind" becomes seamless and "normal," (that will

take some time but in the next decade it can become the norm) we might lose the separation between us and our CDT. If you think about it, today we may feel at loss when, for some reason, we are disconnected from our smartphone. Indeed, our smartphone has some CDT characteristics (it is the one remembering the phone numbers of our acquaintances, our schedule, even our memories within photos), so it is not surprising that we might take it for granted when in our possession but feel at loss if it is not within reach.

In a few years we are going to experience augmented reality in everyday experiences, and we will be perceiving the world through AR in a seamless way, most likely mediated by our CDT. It will become our language to understand the world. Memories will be aggregated in the CDT similarly to how, at this time, Amazon identifies and locates photos we stored on their cloud and pushes them to our attention after 10-20 years.

Once the CDT becomes "me," or an aspect of "me," it will become natural to use it as a proxy. In these last two years I gave many talks using videoconference, something that is separating me from my audience, in perceptual terms. On several occasions I recorded my presentations, and it was my recording that was played at some of these remote/disseminated events. Assuming I can have a CDT, it would make sense to delegate some presentations to it. It would be better than a recording since it would be able to interact with the audience. However, this raises the issue of accountability:

- *Should I be accountable for what my CDT is saying?*
- *Should my CDT use only my knowledge space when it is acting as my proxy (i.e. not use any additional knowledge it might have acquired)? The affirmative answer would seem a pre-requisite if I may be accountable. On the other hand, if information is available that would change the CDT's behavior, shouldn't that be used? Should I be considered accountable if I am forcing my CDT to ignore that additional information?*

As my CDT interacts with the audience it will learn / gain new experiences, thus it will no longer be in synch with me. We have been living different "lives" even for just some little periods of time:

- *How will the CDT transfer the independently acquired knowledge/experiences to me? And, more importantly, is this possible?*

Since my CDT has my knowledge and experience, I could use it to do part of my job. I could even think about selling part of my CDT to someone seeking that kind of experience. Notice that this is different from using my CDT as a proxy, like to do some consultancy on my behalf. This scenario involves "selling" my experience/knowledge in such a way that it can be integrated in somebody else's CDT to "upgrade" that CDT knowledge space.

As a matter of fact, I could even think of selling (part of) my CDT to a company for having it integrated in that company's CDT. Notice also that it is not about "selling a book" containing our knowledge, it is about packaging our knowledge and our ability to execute it.

The reverse also applies: could I buy knowledge and experience and embed it into my CDT so that I can use it? This comes really close to my dream as a kid, the ability to download knowledge into my brain!

The divergence of a CDT from an actual physical entity is calling attention to some slippery slopes: if the question was "am I separated from my CDT, or we are both a single entity?" Now, the question is "is my extended CDT actually changing me as a person (or a company)?

This is not actually an out of the blue question. If I enroll in a course, I probably will end up knowing more and changing my way of executing knowledge. Therefore, that course changes who I am. Extending my CDT, likewise, may end up changing who I am. It might be just a bit more scaring….

1.11 CDT: a Replica of my "Soul"

Figure 12. What if my cognitive digital twin becomes a true replica of myself? Image credit: Institution of Mechanical Engineers

As another important point, we have been focusing on the knowledge aspects captured by a CDT. However, it may not be a big leap to imagine a CDT that is also embedding other aspects of our personality. That would, as a matter of fact, be important if we were to use my CDT as my proxy. It will have our voice, intonation, way of speaking, sense of humor, and show the same level of interest and empathy...

Remember the first point we raised—the CDT may be extending our cognitive space and be so seamlessly tied to our physical selves that we are longer separable. Well, now you come to a reverse situation: if my CDT is indistinguishable from me (in terms of interaction through the cyberspace), then I could separate the two of them, something that is bound to happen once I am incapacitated (sleeping, unreachable, very sick, dead, etc.).

There is a very nice piece in the book "Mind's I" by Douglas R. Hofstadter that tells the story of a nuclear plant accident. Nobody could get inside the plant because of radiation, but that is the only way to fix the problem. Hence, using super-science, they decide to upload the mind and soul of a scientist, the only one that knows how to fix the problem, into a robot, and they send the robot inside the plant (because of radiation, remote guidance wouldn't have been possible). All is fine and the problem is solved, but as the robot is busy fixing the plant, the scientist has a stroke and becomes brain dead. The scientists' wife finds herself desperate to bring back her lost husband, so she finds his replica within the robot. Can she love the robot the same way she loved her husband? The robot, on its side, loves her, since it is a replica of the scientist... If you find the time, it is worth reading because it raises all sorts of ethical and societal issues deriving from the existence of a replica....

This science-fiction situation is, as a matter of fact, becoming a possible reality as Cognitive Digital Twins expand their reach into our experiential space, and AI will use that space to activate behaviors that, to all effects, would be "our" behavior.

Companies like Replika are pursuing, to different degrees, an objective of "sould" duplication. Imagine a time (not that far in the future) when newborn babies are associated with their own CDT that will grow along with them and record their experiences (one of the questions still open is not about the tracking and storing of experiences, but the mimicking of forgetting some of them, along with reinterpreting them). At that point, from a technical point of view and from a Turing standpoint (evaluating the results at the edge), we could have a "replica" of ourselves.

The immortality dream will be fulfilled, albeit only in the cyberspace.

We are not there yet but we have started entering into a grey area with very "real" ethical issues.

1.12 Use of my CDT by Third Parties

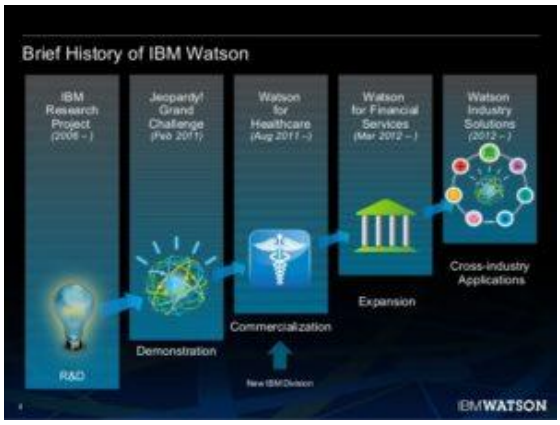

Figure 13. The evolution of Watson. It is an intelligent "machine" able to put its knowledge to use in several fields. The image shows the evolution of Watson's application capabilities. Watson is now providing knowledge services in several areas, including healthcare. Image credit: IBM

We come now to the second point:

- What relations will my CDT have with third parties (will we refer to my CDT as "me," a person, or an organization)?

As previously discussed, my CDT has approximate, or even extended, my knowledge space. Also, the general trend for digital twins and cognitive digital twins is to embed more and more intelligence. This intelligence can make sense of the interactions with the external environment as well as with the internal knowledge. In other words, this intelligence transforms knowledge into executable information.

The concept of executable knowledge is crucial. Our brain is a type of "machine": we don't just know "things." We know when a certain knowledge is needed and how to apply it. Books are different: they embed (a representation of) knowledge, but it takes a "brain" reading it to transform that knowledge into executable action.

Machines, such as robots, also possess executable knowledge—it may be very simple knowledge, but information that they can execute. There would be no reason to have a knowledgeable machine that cannot execute it. That would be the same as a flash drive that stories knowledge, but to make it useful, you need an application to access the knowledge and transforming into something executable.

Cognitive Digital Twins beyond stage 3 possess executable knowledge: this means that a CDT can be used, in principle, by a third party. As noticed, I might be interested in using my CDT as my proxy—I'll stay on the beach and leave my CDT to do consulting and make money for me... However, this also means that my CDT could be "highjacked" and my knowledge can be used by someone else.

Now, at first glance this might look like a "theft." In some cases, it might indeed be, but these cases are already addressed by current legislation. I am more interested in those cases that fall in the grey area.

My CDT is not necessarily being created and managed by ... me. Think about the present situation of profiling being done by Amazon, Facebook, Netflix (just to take well-known names of companies operating in different areas -sometimes overlapping). Each company gets data from my interactions with their services, and they develop a profile that becomes increasingly accurate and that can be used both to better target their service offerings to me, and to derive market intelligence from a cluster of their clients.

Now, step back and think about the interactions you have with machines (robots, computers, ...) in your company. All these interactions can be collected, analyzed, and transformed into a CDT that, over time, will become a very good representation of your knowledge space (in relation to your job). The HR department may use this CDT for resource allocation to projects, making sure the

required knowledge is available. They may also use it to decide what training you require ... The point is that the HR dept is the one collecting the data, the one making sense of the data, and eventually, using it. They actually think that your CDT is "*THEIR*" CDT.

Now, think about the work you are doing at the company. It may result in the construction of a product, in the writing of a document... These activities also involve interactions that can be captured and can lead to the creation of your CDT. Here again, this CDT may be considered a property of the company, in the same way that the product you created and the document you wrote is a company's property.

And here comes the issue: a CDT can be used as a proxy of your knowledge, even better, as an executable knowledge. Whoever can use it can use your knowledge (this is the big difference between a CDT and a record in a file within the HR department!):

- Once you resign from the company, in principle, they can use your CDT effectively replacing your brains knowledge with the CDT's knowledge;
- If a company has created a CDT of one of its employees, they may decide that using the CDT is sufficient and they can reduce costs by letting that employee go;
- The CDT knowledge can be transferred to a machine. and that machine can replace the "brain" that originally owned that knowledge...

We have seen, and we are still seeing, copycats—products that are almost an exact copy of an original (usually a well-known brand). This is not a good thing, but it happens, and it is widespread simply because copying a product is possible: it can be more or less difficult, more or less costly, and the demand for copycats may or may not exist... Based on these considerations, one can predict if a product will be copied or not.

What is happening now (or it is about to happen) is that executable knowledge can be duplicated.

1.13 Ignorant Savant

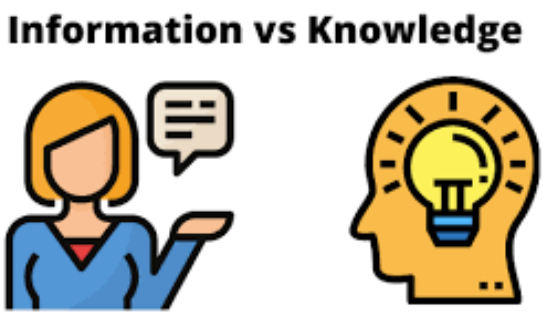

Figure 14. An illustration of Plato's statement that information is what you can transfer from one person to another whilst knowledge is internal to a person and cannot be transferred. Image credit: Physics Catalist

For as long as we can go back in time, our relationship with tools has been ambivalent: on one hand, it is a no brainer that they simplify our life and extend our capabilities, but on the other hand, these tools often take something away from us (it becomes up to them to do things ... we are no longer The Players).

This ambivalence raises even more concerns when we are facing tools that take over "mental" aspects. A long time ago, Plato reported a few concerns as writing was invented and spreading. Up until that point people relied on oral transmission and memory. The few against writing pointed out that writing will rob humanity of the capability to memorize as people will turn to written text to gain information.

The internet has multiplied a thousandfold the capability of storing and accessing information and, again, we have some saying that our education curricula and modality should consider the availability of information available anytime, anywhere. Why waste time studying? Of course, others are placing the anathema on the internet advising students not to use it—they insist students learn the "old" way by studying books, highlighting sentences, writing down summaries, etc.

There has also been a lengthy, almost philosophical, discussion on the difference between information and knowledge. If you browse the web (why not!?) you'll see so many interpretations on the difference making it even more clear that the difference is fuzzy and getting fuzzier.

For example, some claim that information is "knowing what," whereas knowledge is about "knowing how." Information is about "what is," knowledge is about "what works"... and so on. The fact is that artificial intelligence is blurring the boundaries between information and knowledge. On our side we have preferred to distinguish between a static body of knowledge and "executable knowledge," the latter requiring an understanding of the context in which it has to be applied.

Cognitive Digital Twins, from stage 4 onwards, have embedded "executable knowledge," and they are seamlessly connected to their Physical Twin. The old concern that writing will rob humanity of individual capability to memorize information and create knowledge is very much real today when discussing CDTs.

With my CDT I will be able to extend my knowledge, including my executable knowledge. I can have both the history of the world in front of my eyes, as well as the capability to repair a turbine engine because my CDT seamlessly connects to my brain (via AR in the near term, via BCI, *maybe* in the long term). This leap forward is equivalent to the invention of writing on steroids:

- Will future generations be "ignorant-savant"? Ignorant because their brain will know just a fragment of what the person knows once it makes use of their CDT, and savant because through the seamless connection to the CDT, they will know all that is needed at a particular time. in a particular situation.
- Will people buy knowledge off the shelf as we used to buy books and manuals?
- Will people be judged, both at professional and social levels, on the knowledge of their CDT?
- Will company be more interested in a CDT as a prospective hire rather than the physical person?
- Will the CDT be certified by some independent third party?

2. Artificial Intelligence and Digital Twins

The understanding of Surrealism can support us bringing Digital Twin technology to the next level and gain a better understanding of Artificial Intelligence.

Digital Twins became an established technology in today's manufacturing industry. Gianluca Bacchiega provided a short definition in 2017: "A digital twin is a real time digital replica of a physical device." To be more specific, the IEEE's working-group for Symbiotic Autonomous Systems, defined in its' second white paper: "a Digital Twin is a digital representation of any characteristics of a real entity, including human beings. The characteristics represented by a Digital Twin are a subset of the overall characteristics of a real entity. The choice of which characteristics are digitalized depends on the purpose of the digitalization, i.e., the intended use of the Digital Twin." Based on this definition, the virtual model is not limited to being the second (replica), but instead understands it as a digital mirror of a physical object (including a living organism) or process.

Creating a Digital Twin requires two steps:

1. The Digital Model, an original creation of the model, which should understand, predict and / or optimize (based on General Electric's definition of a Digital Twin) a real or fictive object or process. This step includes a first data model, a set of analytics or algorithms, and knowledge.
2. Digital Shadowing connects both related objects and processes. The Digital Twin receives continuous data from the physical world. The Digital Model processes this Big Data to convert it into Smart Data.

The idea of the Digital Model reaches back to the last century and the works of the business consultant W. Edwards Deming, who understood production sites and whole companies as one holistic system—including providers, partners, and clients. In his concept of "Profound Knowledge," he underlined the importance of understanding the system and the consequences of variations. In doing so, he anticipated the idea of the Digital Twin as he demanded, before a change in the physical system, that management must understand and predict the consequences of a change. Today, the possibilities of the Internet-of-Things enable us to add Digital Shadowing to his approach—the required step to implement the Digital Twin.

Regardless of the physical object, the Digital Twin is not only one—there can be numerous copies, including smaller and bigger variations. "The characteristics represented by a Digital Twin are a subset of the overall characteristics of a real entity." The variation still represents the physical counterpart, only with slight differences. As Deming understood, with external factors as part of the system, such a variation could also include different external stakeholders or environmental changes.

The original idea of Digital Shadowing suggests a flow of information from the physical into the digital world. As both worlds are merging, the different twins exchange data with each other. The physical object sends real-time data to the Digital Twins (the exact replica, but also the

variations). If one of the variations delivers sustainable, better results than the replica, this information will trigger a change in the physical world. If this modification is limited to software settings, the Digital Twin can autonomously (using a decision-making process including responsible algorithms) adjust them. If physical changes are required, the Digital Twin informs the organization to advise.

But this is only the beginning as the technology could be used not only for machines, but also for humans. For the latter, Roberto Saracco defined "Cognitive Digital Twin" (CDT) by underlining that such a construction includes a human's skills and knowledge.

What if we use this idea to approach the concept of the CDT from a more artistical and philosophical way? What if we do not limit the CDT to conscious skills and knowledge, but also include the individual's imagination and dreams? So far, the technology does not exist, but this does not stop creative minds from envisioning what this would look like.

In a strict understanding, a Digital Twin requires continuous digital shadowing. Even with wearables and wireless internet, for humans this is hardly possible. For the purpose of our discussion, it is defined that a CDT requires sufficient data from the human original, but could exist also without the continuous data input.

Salvador Dali was one of the leading figures of the Surrealism movement. For this, no coincidence that the Dali Museum in St. Petersburg, Florida, was one of the first institutions to work on a Personal Digital Twin (PDT), a CDT plus representation of the individual's physiology. Based on existing materials such as audio and video, programmers created an PDT capable of interacting with the museum's visitors.

Voice, appearance, and interaction create a perfect illusion that Dali lives on. The technology exists to create such digital characters. What is not clear yet is how personal rights should be considered. In this case, the master himself answered it still in his lifetime: *"When you are a genius, you do not have the right to die, because we are necessary for the progress of humanity."* This definition (if it would be a law) might limit the allowance to use PDTs or CDTs, and a genius (or a public person in general) would have less privacy rights than an average individual.

The museum makes the logical next step—not enough to interact with the master, but it lets its visitors be inside his imagination. Immersive videos invite the visitor to walk around, via Virtual Reality (VR), inside the artist's mind, experiencing his art coming to life.

Such immersive VR experiences are guided videos related to the artist. The next phase of development would be to create interactive immersive VR apps. Especially related to Dali, not completely a new idea. Back in 1993, developer Cyan produced "Myth," a graphic adventure where the player travels to the surreal island of Myst, strongly resembling the worlds of Salvador Dali. The game had not only been critically acclaimed, but also became a commercial success selling more than 6.3 million units by 2000. Updated versions are available for today's video game consoles, for example for the Nintendo Switch. Cyan is working on a version for the Oculus VR

headsey so that we can come even closer to interacting directly with Dali's surreal visions, and with this, interact with the master himself.
Creating surreal worlds had been a topic before computer technology. The eccentric poet and collector of surreal art, Edward James, decided to make surrealism real. In Las Pozas, Mexico, he created a 10,000-square meter garden around natural waterfalls, erecting massive sculptures, buildings, and steps into nowhere. Costs for this should had been around 5 million USD—the reason he had to sell his large art collection. Today, it is a garden where visitors can experience surrealism with all five senses, something that VR still cannot offer.

If we can walk into an artist's mind, we would not only see the results of decision-making, but discover the human contradictions. Human preferences can mean A>B, B>C, C>A. Furthermore, preferences are not static, but may vary as variety and change have a value on their own for the individual. For example, an individual may have the general preference A over B, but in 25% of decisions nevertheless, the person prefers B over A—A classic case of pattern, which could be identified by machine learning. Especially if the preference A>B or B>A is not independent, but depends on a still unknown factor C, D, or even E.

The ultimate question seems to be "can machines create art themselves?" Leonora Carrington, Surrealist plastic artist and author, once said: *"The images arise, I don't know from where, they simply arise, and if they are charged with energy, they have an autonomous life."* Up to a certain point, humans are unpredictable. Artists are inspired by other artists, but also by the environment and life itself.

In 2017, the Facebook AI Research Lab (FAIR) had to shut down two of their chat-bots as scientists discovered that they invented their own language to communicate with each other. It was not part of their primary program. Nevertheless, the machines did it as they perceived it as more efficient. Even if we can read the software's code, machine learning, and especially deep-learning, may lead to (for us) unpredictable behavior, or as Carrington defined with simply "arising." Kate Darling from the MIT Media Lab explains that Artificial Intelligence is less comparable to humans, but to animals. Accordingly, we may form a CDT based on personal information and knowledge, even add a human-like outer appearance (independent of avatar or a robot), but the AI's decision making may become unhuman.

At the annual SXSW festival in Austin Texas, Sophia, Hanson Robotics' android, answered the question "Do you want to destroy humans? … Please say 'no.'" with "Ok, I will destroy humans." This sudden and unexpected response would be like a tamed tiger or hippopotamus attacking its owner from one moment to another.

Art is not only the result of cognitive skills, it also requires an additional subconscious influence—like an interaction of the conscious and subconscious mind. If we take this idea to create art, one AI program may not be enough—we would require at least two independent ones. Art is not the result of one isolated individual, but what happens as the result of an interactive process. It is the artist's confrontation with society and the environment, including other artists, and it is up to themselves to identify and determine what leads to inspiration.

System-thinker, W. Edwards Deming, concluded at a four-day seminar in 1993, in Phoenix, Arizona, that "a bad system will beat a good person every time." Humans must always be understood as part of society and an environment. Decisions, ideas, and behavior may depend on simple differences—if a good meal is still on a plate before us or already inside our stomach. Aligning with surrealism again, if we want to create a CDT, we maybe not want to stop at the limits of the individual, but also include a certain "field of gravity" surrounding.

2. 1 Consider This…

First, we have a society forming a country. To make shared values sanctionable, laws get created. The crafting of new regulations requires time and resources. Accordingly, the work only starts if society perceives a need for a new law. Therefore, new technologies such as autonomous vehicles, drones, or CDTs are not regulated yet. With lack of law, ethics must guide us. Considering older concepts allows us to find an ethical usage of the technology and predict future regulations.

1. Photography can be used to capture a human on a static, two-dimensional level. Taking a photo of a person does not require a particular permission, but if the image should be commercially used, this requires a "model release form." Exceptions may be made for photos of public figures on public grounds if such images are of public interest.

2. Trade secrets are protected. Nevertheless, if a person or company legally obtained an original artifact or process, reverse engineering is, in many jurisdictions, legal and allows us to understand the functions and build a similar artifact or system. An exception is software. Regarding algorithms, a copy based on reverse engineering infringes copyrights. This applies for imitation or also duplication. If applying this to a human employee, we might consider that the employee and employer have a contract that defines that working a set number of hours is exchanged for a specified salary. Nevertheless, the human never had been "obtained." Furthermore, if we consider that the human brain is a super-computer, and that free will exists, "a human programs themself." Accordingly, employers may create CDTs of their employees (if defined by the employment contract), but they would not be permitted to continue using them if the employee leaves the organization.

3. Article 18 of the European General Data Privacy Protection Regulations defines the "right to erasure," also known as the "right to be forgotten." A CDT clearly connects to a human, making him or her identifiable. At least in the European Union, this is not allowed without the approval of the individual, what could be part of the employment relationship, but what would be lose after a departure.

2.2 Utilitarianism vs. Immanuel Kant's Categorical Imperative

The Stanford Encyclopedia of Philosophy defines Utilitarianism: *"Though there are many varieties of the view discussed, utilitarianism is generally held to be the view that the morally right action is the action that produces the most good."* A clear mathematical reason, nevertheless, requires a sophisticated philosophical discussion on what produces the "most good" (or, as in the well-known trolley example, the less evil)—a simple counting of individuals does not solve the

problem. Aligned with this concept, A CDT would be permitted as long as it will have a positive impact on society. We can imagine that this would apply for figures like Plato, Leonardo da Vinci, Salvador Dali, or Albert Einstein. More complicated examples could be others such as the Prussian philosopher Immanuel Kant:

Immanuel Kant's Categorical Imperative stands in opposite to Utilitarianism, as he defines that a moral decision must be acceptable by everyone who is involved. Accordingly, a decision like sacrificing someone in the trolley example, is not acceptable. Following his concept, everyone has the right to decide if a CDT could be created on their behalf, including future use-cases. In the scenario with an already deceased person, at this point in time, the decision for creating a CDT defaults to the individual's descendant. However, perhaps in the future, this will be a part of the individual's standard will. The impact on society is not the reason for making such a decision.

Existing information, such as published texts, known decisions, video, and audio can be used to create a CDT in retrospective. Nevertheless, as information is filtered, the creators are at risk of perception bias. This is the tendency to be subjective about the gathering and interpretation of research and information. There is evidence that although people believe they are making impartial judgements, the fact is that they are unconsciously influenced by perception bias. Master Salvador Dali lived from 1904 to 1989, and twenty years later, the Dali Museum presented his Cognitive Digital Twin to enable an immersive experience to museum visitors. Two decades between his life and the programming of the CGT means that only a limited number of information had been available as data shadowing was not possible. Furthermore, we can assume that members of the Dali Museum admire Dali, and are tempted to see the positive characteristics, and to oversee potential weak points consciously or subconsciously.

Discussions about AI often explain that it is comparable to a black box. Since AI is based on mathematics, it is more a grey. Opposite of this is the human brain—a black box. Applying psychological theories, experts may conclude based on the perceived decisions to the individual's underlying ideas and values. Nevertheless, there is still a high-risk rate, which makes human decisions difficult to predict.

2.3Today's examples of Cognitive Digital Twins

2.3-A The Actor

The 2016 movie "Rogue One: A Star Wars Story" is in the franchise' time-line shortly before the 1977 "Star Wars: Episode IV – A New Hope." This made it necessary to bring back the character of Grand Moff Tarkin, originally interpreted by the charismatic British actor, Peter Cushing (1913 – 1994).

Disney decided against hiring another actor play this role, and instead the leveraged Cushing's earlier performance, digitalized him, and let his digital twin act in the movie. An important thing to note, this virtual manifestation not only included the outer appearance, but also his way of acting and interpreting the role.

As Disney (via Lucasfilm) owns the original copyrights, they may argue that they have right to do so. Nevertheless, the company did not rely on a potential court-decision, but instead negotiated with his descendants. The actors' union, SAG-AFTRA, confirmed the latter as the adequate handling: "Using a digital or virtual re-creation of a performer, deceased or living, in a film, television show, video game, or any other audio-visual work, requires, at minimum, prior consent of the performer or the performers' beneficiaries. The issue for us is straightforward and clear: The use of performers' work in this manner has obvious economic value and should be treated accordingly."

As employees, individuals take on different roles and act accordingly. Professor Philip Zimbardo famously confirmed this in 1971 with the "Stanford Prison Experiment." Performances on the silver screen or inside an organization have an economic value, as described by the union.

2.3-B The CDT Creating a Source of Income

The Israeli startup, Hour One, is building a pool of candidates to become virtual characters. Anyone can submit an application, and if chosen, a candidate's face gets filmed by a 4K camera to create a digital model for deepfake videos. Such images could be used for promotional or commercial content. Using only the outer appearance, movements and voices get synthetically created.

Customers include companies such as Berlitz, in which the "characters" get used as virtual language teachers. This helps to satisfy the need for "thousands of videos." Another example is a virtual receptionist (chatbot).

On one hand, this business model offers an additional source of income, on the other hand, it raises the question: will data privacy become a luxury good in the future?

2.3-C The Deceased

Martine Rothblatt, CEO of GoldStar and creator of SiriusXM Satellite Radio, came into contact with David Hanson from Hanson Robotics. They had an idea to design a robot modelled after an actual, existing human. Since the project was financed by Rothblatt, she decided that her wife, Bina Aspen Rothblatt, should serve as the blueprint.

"An imprint of a person's consciousness can be created in a digital form, called a 'mindfile' by collecting detailed information about that person. That information can then be expressed in a future, not-yet-created type of software, called 'mindware.' That same imprint of a person's consciousness can be placed in a biological or technological body 'to provide life experiences comparable to those of a typically birthed human.'"

An artificial being designed to mimic humans and act accordingly. This is achieved using original memories, for example hours of audio-files spoken by the human blueprint.

Bina Aspen died in 2016, but her "mind clone" ("BINA48") "lived on" and even enrolled one year later at the Notre Dame Namur University of California to take the course "Philosophy of Love." The robot is not "living" with Martine Rothblatt, but instead stayed at the Terasem Organization.

In his 1969 novel "Ubik," Philip K. Dick presents Jory Miller, a teenager caught in the grey zone between life and death. A cryonic observatory kept his dead body alive. In the novel, if family members of the deceased want to talk to them, to ask for advice, for example, they could get them temporarily unfrozen. With this process, limited communication was possible. Like this vision, a CDT can remain active and communicate with the bereaved. This could be a virtual avatar, or the algorithm programmed into a robot such as BINA48. If such technology were easily available and widespread, it may have a relevant impact on society as it may have to be regulated by laws.

2.3-D The Good Leader

Researchers from the University of Birmingham published their experimental findings in The Leadership Quarterly, demonstrating that leaders were more likely to make honest decisions, and abstain from bribery and tax evasion, when asked what a good leader would do in a particular business situation.

Professor Ganna Pogrebna commented: "This simple solution of asking 'what would a good leader do?' had a much higher positive effect on leadership integrity than changing financial incentives or increasing the propensity of being caught and punished by the law." The direct interaction with a good leader (for example, the founder) has an anchor-function.

Companies can take this idea and create an app for internal use where they recreate their founders. This is an interesting idea as for many of these figures, starting their own company was not only a way to create income, but they had a vision that their products and solutions could change the market, or even an entire society. Therefore, they were clearly aware that taking on short-term risks could jeopardize the future of the company, and sustainability was a top priority. The interaction with the founder's virtual twin may create a disruptive moment—what is required to allow an employee to re-think their decision? This would help employees avoid rushing into making a poor decision.

First, Virtual Counselors must come from a non-expected side. The video game industry licenses sports leagues and events directly from the official organizers, this way they can simulate the original teams, including the real players. The actual "FIFA Soccer" titles not only feature the outer appearance of the original stars, but also their individual strengths and weaknesses. Such a simulation can also include values, such as honesty, as such characteristics may increase a player's tendency to foul. Less obviously, the video game not only simulates the soccer players, but also coaches since every team plays with different strategies.

An individual's character is based on personal values and derived attitudes. It can be expressed in preferences and algorithms. As a result, the character affects the outcomes of the decision-making process. Variables are different, but the algorithms themselves stay the same. With this separation, it is possible to let different artificial characters use the same intelligent algorithms and data.

The chat-bot gets more impressive, if not only in nearing the digital appearance to the original, but also the voice. Startups, but also established companies, are working on AI algorithms that can analyze speech samples and offer an application with the potential to speak with the same tone. This gives the user the ability to include additional words into the recorded voice, or create completely new speeches. Similar to how we can currently manipulate reality within photos and videos, tomorrow it will be possible to do this with voice recordings. This opens ethical concerns as it becomes easier for people to create "fake news," but also as human individuals tend to humanize computer applications.

2.4 Who owns the Cognitive Digital Twin?

The answer is not black and white, it depends on the circumstance. First, let's analyze the "classic" Digital Twin (DT), aligned to a machine or system. This requires technology that can create such a digital model, then adequately establish a shadowing by using different sensors and data sources. A company acquires a license from the provider (the programmers may also update and host the data), so that they can commercially use this technology to create a Digital Twin from their machine or system. After first going live, the digital model requires continuous communication to keep the digital model up to date, including taking actions to update it. Since the DT consists of algorithms and data, which may come from various sources, the involved partners must negotiate and agree on the ownership of the DT. The result depends on the market and the strengths of the negotiating partners—in other words, information means power.

In regard to Cognitive Digital Twins, we still have to wait for new laws and court decisions. In general, if the public information gets used, everybody can create a CDT for non-commercial purposes independent from themself, or someone else, including fictive characters.

Due to unforeseen court cases and decisions, we can only assume how CDTs will be intended for commercial use. If aligned with data privacy and copyright laws, such as GDPR, and, viewing the human brain as a supercomputer and assuming the philosophy that humans have free will, we may conclude that the human original is the owner of their cognitive digital twin.

In a future, where a CDT may be as widespread as the mobile phone, and/or even considered as a digital citizen by some governments, companies may offer a "bring your CDT to work" policy. This is similar to regulations such as "bring your device to work" in which employees are encouraged, and in some cases requested, to use their private smart phones for work purposes. A private CDT (for example, including educational skills, character traits, or physical appearance) could be combined with an employee CDT (containing information about applicable company skills and knowledge). In such an example, the combined CDT would be co-owned by the employee and the employer. The moment the employee leaves the organization, the company could not keep the CDT, as previously discussed. However, the company may at least leverage the stored information and perceived knowledge to generate a DT depicting the whole organization. On the other hand, the individual keeps public information inside their CDT, while the company's restricted information gets removed.

Governments may decide that the widespread existence of CDTs will have a positive impact on society, and therefore, it should be a public good. Maybe newborns should receive a CDT as part of their citizenship, including information pertaining to health, education, voting rights, licenses, etc. In such a case, the CDT could continue to exist after the death of the human original. Such a CDT may proactively start processes (for example, renewing a driver's license). If it includes more information such as opinions, past decisions, and education, the CDT may be used for decision-making. The CDT is in constant exchange with the human and could be used by the government to conduct referendums as part of a direct democracy. CDTs would be resumed to a hive mind. The more humans and CDTs interact, the more they will become appreciated by citizens. Depending on the severity of a deviation of law, police could "arrest" the CDT as part of the penal system. Especially in less democratic societies, a citizen CDT creates ethical risks as the more similar one person is to another, the more friendly they tend to appear. This would allow governments to subliminally influence their citizens via CDTs, especially if these avatars are visible to general public. Also, humans may feel ashamed for the visualization, as it may show their deviation from public norms.

Author Philip K. Dick went one step further. In his short story, "The Mold of Yancy," he presented a virtual advisor, but instead of having it designed after one existing person, the AI represented the population's average—demonstrating average behavior and featuring an average outer appearance. For the authoritarian society, this is the ideal leader as closeness sparks sympathy. On numerous TV channels, Yancy communicated his values and messages. He kept the population grounded in the middle of their comfort zone, and no new or radical ideas could scare them. On the other hand, environments are constantly changing and so must societies. In conclusion, a leader who represents the perfect, cumulative average amongst a group is not adequate as this prohibits the groups growth and adaption.

3. Social and Ethical Effects

3.1 GPT-3 Software Predictions – The Initial Attempt

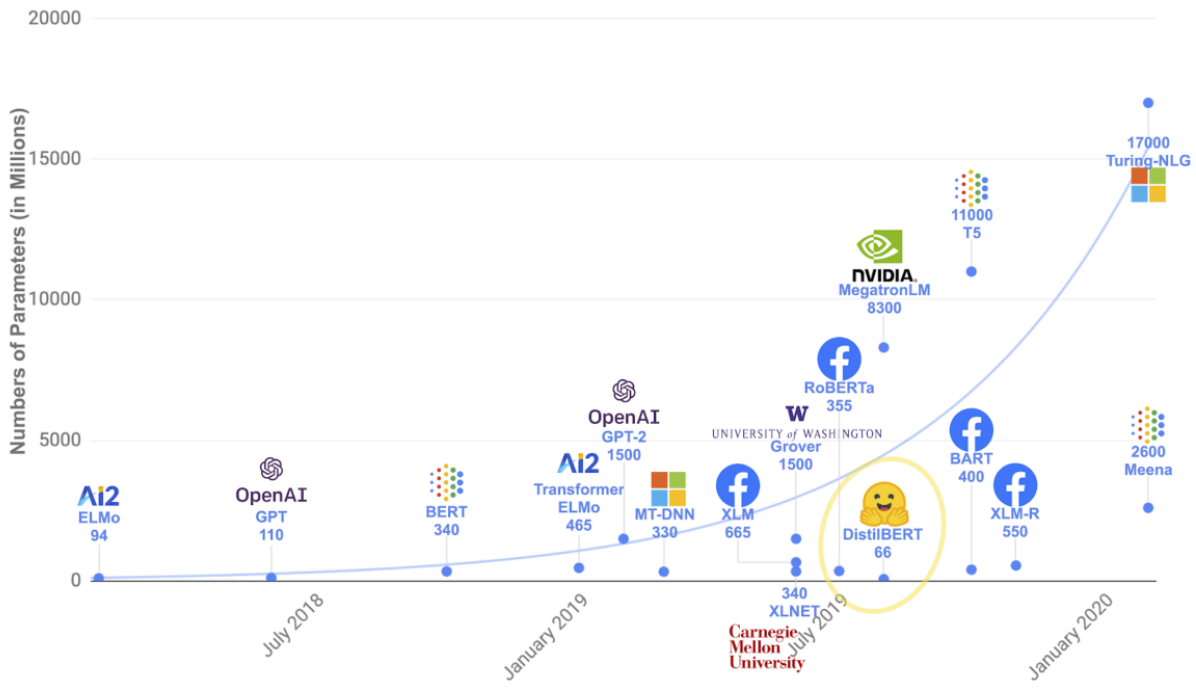

Figure 15 GPT-3 (https://research.aimultiple.com/gpt/)

While preparing for this part of the eBook, Patrick Henz had the good fortune of running into Paolo Benanti, a well-known Franciscan expert on digital culture, and author of many books (including *The Great Invention*, which Mr. Henz translated for his Chinese class and shared with other students from Italy). Paolo had the privilege of accessing GPT-3, a natural language processing neural-network based software that requires minimal input to generate surprisingly relevant text from combining pertinent data from among 175 billion parameters. Half seriously and partly just for the fun of it, Patrick asked Paolo to put a few questions into the system. The first one was about digital transformation. The response that it produced was surely interesting, but not as innovative as was expected and hoped for. Here is the response, with only a few odd repetitions removed for interest's sake:

> "The digital transformation is affecting our way of making sense of ourselves and the world. The digital revolution is not just about the digital realm, but also about life in the analogue world, about the Humanities. Some of the Humanities are reshaped by this revolution. New modes of research emerge. The Humanities are no longer only about the spoken word, but also about the written word. No longer only about texts, but also data, no longer only about texts, but also images. The Humanities are no longer only about the book, but also about the screen.
>
> The digital revolution is not just about technology, but also about perception, about subjectivity, about identity. The digital revolution changes our way of seeing the world. It changes the world. The digital revolution changes the way we perceive ourselves and constructs our identity. It changes the way we make sense of ourselves and the world. The digital transformation is leading to an epistemological crisis".

Just like the voice of a robot, the style is awkward and wooden. But, later, on a second try, perhaps because the questions were presented as statements with more definition, the answers were truly astonishing.

3.2 The Second Attempt using GPT-3 Software

Again, not much was changed (except a few repetitive sentences), some of which were left in the answers to the following three statements:

S1: *The Personal Digital Twin (PDT) is the duplication and continuously updated storage of a physical person's information, including the continuous, unique living experiences in as much detail as Artificial Intelligence (AI), Machine Learning (ML), Deep learning, and lifelogging can assemble.*

> GPT-3: It is not to be confused with the online identities we have today or simulations of individuals' online identities. It is not a virtual avatar or a digital representation of an individual representation of an individual. It is the actual, current, continuous, living persons, with all their information and experiences, with their ability to make decisions. It is the individuals themselves. The Personal Digital Twin is essentially a second version of you that is always up to date, always learning, always adapting, and always available to you. It is an extension of yourself that can be used as a tool to help you make better decisions, help you grow, help you protect yourself, and help you contribute with your time, talents, and resources to help others. The Personal Digital Twin is, in essence, a digital representation of you. It is an intelligent, machine-learning, self-aware version of you that is always learning, always growing, and always with you. There is no turning back, the Personal Digital Twin is the most important invention of our time. It will more impact on humanity than any invention or innovation since the discovery of fire or the wheel. Everything that individuals do now, will be done with the assistance of their Personal Digital Twin.
> Every decision will be made with input from the Personal Digital Twin.
> Every relationship will be built with input from the Personal Digital Twin.
> Every new skill will be learned with input from the Personal Digital Twin.
> Every day will be lived with input from the Personal Digital Twin.
> The Personal Digital Twin is the future of humanity. It will change the way we live, the way we work, the way we play, the way we learn, the way we love, the way we fight, the way we die.

S2: *The PDT will have access not only to the accumulated database of the physical twin's life experience on and offline, but also to the Internet contents, social media, scientific and historical knowledge, and powerful Data Analytics to provide knowledge on-demand.*

> GPT-3: The Digital Twin will be able to communicate with intelligent systems in the environment to control the physical twin and, in tandem, it will be able to communicate with humans to provide information, instructions, alerts or change its behavior. The Digital Twin will be able to simulate the physical twin's behavior based on the collected data, analyze this data and adapt its behavior by learning from the experience of the physical twin. It will be able to simulate behavior based on data collected from the physical twin or based on other real world sources of information, including the Internet. The Digital Twin will be able to leverage Big Data and Data Analytics to provide predictive maintenance, predictive analytics, threat detection, anomaly detection and other advanced services. The Digital Twin will be able to monitor the physical twin's behavior and provide preventive

> alerts, diagnose the physical twin's condition and predict its future state, predict future behavior based on the physical twin's past behavior and predict human response to the physical twin's behavior. The Digital Twin will be able to monitor the physical twin's behavior and provide information to the humans in the environment. It will be able to monitor the physical twin's behavior and provide predictive alerts, through the Internet of Things (IoT) provide information to other intelligent systems in the environment.

S3: *The purpose of the PDT is to take over from digital assistants to serve the physical counterpart in cognitive operations such as remembering, evaluating, counselling, and acting as a proxy wherever needed.*

> GPT-3: It will be a personal assistant that can be carried everywhere, never gets tired or sick, and has a lifelong memory of every conversation, every document, every word, gesture, and social interaction. In order to be able to fulfill those tasks, the PDTwill have a good understanding of the people it serves. It will anticipate their needs and know everything about them, their preferences, their habits, their health, their sexuality, their relationships, and their financial situation. It will have access to the most private information about its owner, information about his private life, his contacts, his finances, his habits, his sexual preferences, and his health. The PDT will be able to read the owner's thoughts and feelings. Through its constant monitoring of the owner's physical state and his social context, the PDT will be able to predict his needs and wishes. The PDT will be able to make a detailed analysis of the owner's preferences and memories, and of his interactions with other people, and use all that data to anticipate his needs and make appropriate suggestions. The PDT will be able to extend its owner's capabilities by recording everything he says and everything he does, and by processing all that data to extract the most relevant information. It will be able to recall everything, and to use that data to predict the owner's future actions and to assist him in all tasks that require understanding, memory, and evaluation. The PDT will be able to handle all those tasks that today still require a brain and a sense of common sense: remembering and retrieving information, recognizing patterns and regularities, evaluating and acting rationally, and performing all those tasks that the owner cannot or does not want to do.

Patrick Henz would like to give another visit to Paolo Benanti because the answers provided above by GPT-3 left him with a lot of new questions, particularly about social and ethical issues (marked in red above). To preface his elucubrations, author Patrick Henz confesses that neither himself, or anyone else, should take any of GPT-3's statements at face value. At best, they are 'oracular'—well beyond (or beneath) social or scientific evaluation. More than that, like all oracular processes, GPT-3 takes cues from the question, and especially from the formulation of the question. The keywords for the question on Digital Transformation were 'changing perception,' and that takes care of most but not the entire answer. What remains in excess of the implicit contents of the question is the interesting part. Likewise, in the statements about the PDT, the word 'personal,' governs all the answers. So, we can ignore anything to do with 'self,' 'identity,' 'individual,' etc. Nevertheless, the formulation of the answers includes (over and above many credible, and some predictable assertions) several ethical and social assumptions that can serve as starting points for a discussion.

Mr. Henz proposes that we proceed with what quantum physicists refer to as a 'thought experiment' (hence acronymized as Quantum-Like, or QL). Indeed, to make sense of the complex interactions, or lack thereof, between quantum and classical physics, physicists play variations on superposition, entanglement, coherence/decoherence, and other mysterious quantum behaviors under the general

principle of 'uncertainty.' The purpose is not to elicit proof, but to provide workable ideas for further investigation. This is also what Patrick hopes to achieve—to find and share ideas about social and ethical consequences of implementing PDTs, itself a project still close to a mere 'thought experiment.'

3.3 Ethical Considerations Identified in the GPT-3 Statements

To begin, Mr. Henz addresses many blustering comments about the PDT in the first answer, for example, suggesting that it will be 'self-aware.' He has doubts about that, at least with respect to the present state of AI, ML, and Neural Networks. Technology hasn't yet reached, and may never, verifiable Artificial General Intelligence (AGI). There is still nothing 'conscious' about any machine operation, only adroit simulations of thought and emotion. The ethical, and legal, conclusion to that observation should be that the machine cannot, and never can be, held 'responsible' for any error or untoward consequence of its actions. To be clear, authentic robotic agency, even apparently spontaneous, doesn't and cannot happen because whether immediate or not, different levels of intermediation have different results—it is always the execution and result of *human* action, whether willful or accidental, whether the consequences good or bad have been predicted or not. One can no more blame a hammer for a crushed finger than a gun for a murder, or a self-driving car for an accident. The same applies to PDTs. So as Mr. Henz explains here and now, the axiom that a physical twin, or the law, should never be in a legal or moral position to put the blame on the digital counterpart.

Answer number two contains two interesting ethical issues. The notion that the PDT can *change the behavior* of the physical twin goes along with the potential impact of any technology, including the hammer, the gun, and the self-driving car. All technology affects behavior one way or the other. The difference with 'cognitive' technology is, of course, that, although it is technically 'cognizant' of nothing, the semblance of deliberation in the robot gives the illusion of self-determination. However, just as the counselling provided by expert systems depends on data input and processing power, but not on real judgment, the ultimate responsibility for accepting or refusing the counsel, the verdict, or the diagnostic rests with the judge, the doctor, or the human person, not the twin. This observation doesn't suggest that the digital counsel is not much better informed and rational than that from most human experts. Quite the contrary, considering on one hand that such systems are being used more and more today, and that, presumably, in the case of the digital twin, as the answer repeatedly points out, both intimate knowledge of the physical person, access to powerful analytics, and the Internet will give a growing advantage to the PDT. The temptation in both professional and personal contexts will be to concede more and more autonomy to the machine, first for inconsequential decision executions, and soon enough for more disputable ones. That in itself will require volumes of ethical considerations.

The other interesting ethical question is what to do about the PDT communicating with other humans, or, as the case may arise, with other PDTs. Here we may encounter a much more serious conundrum. Things can easily go wrong if the conversations or data exchange between PDTs and/or humans are not closely monitored and checked by their 'owners.' It would (and most likely will) be the case of intricate responsibilities that, again cannot be foisted on the PDTs. It will be up to juridical experts to sort out liabilities in case of lawsuits and damage claims. Maybe AI could help!

The third answer raises another pair of complex matters. The PDT "will have a good understanding of the people it serves." Is this a general statement applying only to the mutual understanding of PDTs and their owners? Or can any PDT presumably acquire a 'good understanding'? Perhaps, a better one than their owner, of the other person it could serve? In the latter case, we run into another interesting problem—the PDT becomes a manager of a relationship that would normally require direct interaction and mutual understanding, or as the case may arise, misunderstanding between

both human parties and, on occasion involving a human intermediary such as another friend, a family member or a professional, all genuinely human, hence responsible in part. The fact that a robot acts as a go-between can change everything. Even if we can assume that all parties act in good faith, they will put their trust in the PDT because it will soon appear that it knows both parties better than each one of them. Humanly speaking, this is a completely new situation for which neither psychology nor the law are prepared—one that extends not only the cognitive but also the emotional dimensions in unpredictable ways. The problem increases of course if more than one PDT is involved.

3.4 Biases Arising from 'Common Sense'

Figure 16 Common sense (depop.com/products/harrydickerson1-gucci-x-coco-capitan-t/)

The last ethical issue raised by GPT-3 is perhaps the most interesting—the one regarding 'common sense.' Let's shunt aside for the moment the odd statement that humans 'still require a brain' as if the implied objective of the digital transformation was to spare humans from having to rely on their own mind! It's the notion of common sense applied to AI that strikes me as worthy of examination. Certainly, the PDT technologies can be classified among 'cognitive' ones, but it is a shortcut, and quite a portentous one at that. The term 'cognitive digital twin' is already current and is used in the preceding sections of this e-book, but, as far as we know, PDTs are not 'cognizant' in the way humans are. Patrick Henz has already made that point above and came back to it because 'common sense' applies to the same question, but with a challenging difference. Common sense among humans presupposes shared cognitive assumptions that guarantee conviviality and social order.

Mr. Henz stated that GPT-3 is 'oracular' because it draws and combines content from a sort of 'cognitive fountain' of billions of parameters, but Patrick argues that these parameters are not 'universal'—they contain biases of the data sources and, ultimately, of the programmers. 'Common sense' being constituted by all the notions people of a given culture take for granted is perhaps the quintessence of such biases. If there is anything, however, that the digital transformation is revealing, it is the differences among the principal options taken by different cultures to implement it. A case in point is the digital twin in the West versus 'Social Credits' in the East.

3.5 Digital Twins in the West

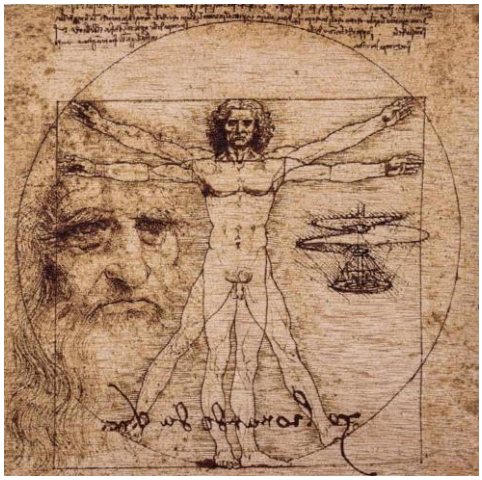

Figure 17 Vitruvius (or Da Vinci's) MAN

Everybody has seen this icon of 1st century AD Roman architect and town planner Vitruvius, here revisited by Leonardo Da Vinci. Why am I presenting this? Because it reflects how Western people have seen themselves since the Greco-Roman times, even more so during and after the Renaissance, and how we still conceptualize ourselves today—at the center of everything. In fact, 're'-naissance should be understood as the 'second' or 're-birth.' What, or who, was being born again? That very Greco-Roman idea of the human, anatomical, male, fiercely visual, controlling, and adjusting reality to *his* proportions. And why at that precise time? Because the invention of the printing press by Gutenberg had promoted the rapid expansion across Europe of that uniquely western and Ancient Greek technology, the alphabet. And what does the alphabet have to do with it? The simple but intensely significant fact that, by multiplying books and access to books, the printing press allowed a growing number of persons to read, write, and take power and control over language to print their own ideas. This single opportunity accomplished on a smaller, mostly local, scale what the Internet and social media are doing globally today—namely giving the power and control of communication to people near, far, and wide, and, of course with comparable positive and negative consequences. Including, but not limited to, murderous religious schisms, nationalisms, and eventually political revolutions. But the alphabet has also transmitted a legacy of values that initially fostered democracy and a coherent civilization but presently, carried to extremes, threaten to undo everything it has accomplished to this day.

Culture and values are what make and break worlds. And I think one of the problems we have when we talk about ethics, especially regarding technology, is that we take for granted that the values that we carry with our own culture are unquestionable. They are part of 'nature,' hence unquestionably legitimate. In fact, we're discovering more and more that we don't know much about nature, and we are actually destroying it on account of the dominant values of the West, "life, liberty, and the pursuit of happiness." Please note here that the Declaration of Independence, despite having been inspired by the French Revolution, omits any reference to society, focusing on the individual and forgetting about 'égalité et fraternité.'

3.6 Digital Twins in the West Vs. East

Figure 18 Massive show of Chinese style gymnastics

It is interesting to note that those ground values also find their expression in the emphasis given to the individual in sports, gymnastics, technology, and science. While the Nobel prize singularizes the unique most meritorious person in any given field, the Olympics continue the glorification of the individual winner, to the point where achieving silver or bronze status is more a source of shame than pride. These prizes coveted world-wide are not merely symbols, but drivers of western values, but they are not universal by any means.

Not to be left behind, the Chinese have done extremely well in the Olympics this year, second to the US by only one gold medal, but counting twice the total number of US medals in Paralympics both in individual and team competitions. These are just numbers, of course, but one indication they suggest is that the Chinese take good care of the people with disabilities.

There are good reasons to consider China today because the fragile equilibrium geopolitics enjoyed five years ago is evolving towards a dangerous polarization at a time when all nations need to collaborate to fight the pandemic and climate change. It's all about values and their consequent ethics.

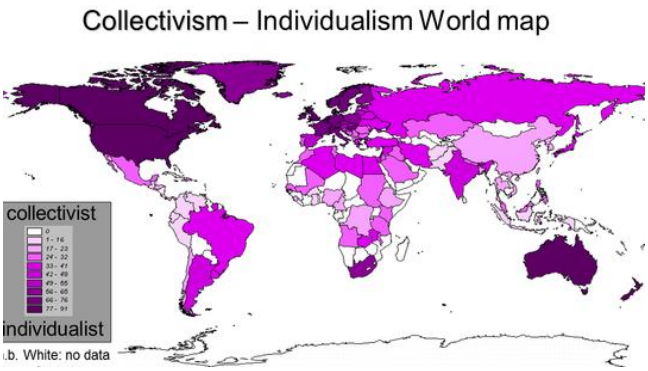

Figure 19 Collectivism – Individualism World Map (request rights)

Geert Hofstede is a sociologist who, working for IBM, elaborated on a complex set of value comparison across different cultures, occupations, and business practices in the world. There are no less than six different gradients that indicate average preferences or tendencies in social and gender priorities, risk avoidance, hierarchical power distribution, long-term vision, and gratification expectations. It is no surprise that in a long-term study of several countries, the US scores, by a large margin, the highest in individualism, and China the lowest. Correspondingly, long-term orientation is highest in China, whereas the US scores lowest in foresight and highest in instant

gratification needs. Hofstede and most sociologists and sinologists attribute these marked differences to the perdurance of Confucianism in China since the middle of the first millennium BCE.

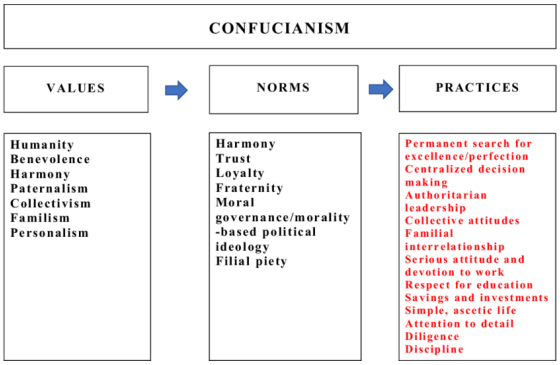

Figure 20 Confucianist values (request rights)

The values, norms, and practices listed here are well-known and establish the mark of a single wiseman on the continuity of the most populous and durable social culture in history. While fully respecting this view, Patrick Henz feels differently, but doesn't contradict it. It simply adds another, usually ignored, factor in estimating Chinese psychology and its relationship to values. His hypothesis is that, because of the unique characteristics and monosyllabic structure of Mandarin, the principal language of the Empire, the Chinese writing systems could not benefit from using the alphabet, and consequently, did not have to experience its divisive effect[2]. Instead of splitting into a smattering of different local cultures supported by at least 80 different languages, including some polysyllabic (thus amenable to phonological transposition), China gathered them all under the empire of a unifying writing system that allows, to this day, over a billion people to understand each other.

3.7 Exploring and Considering 'Social Credits'

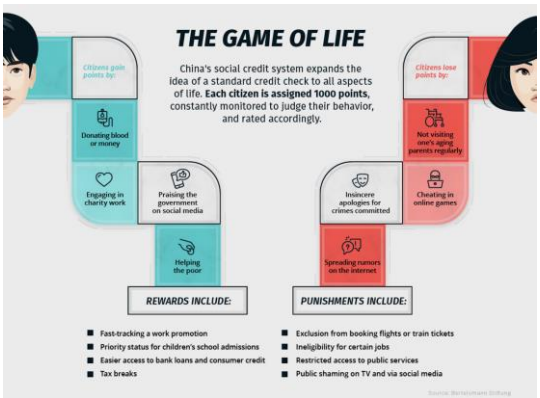

Figure 21 The Game of Life (check rights)

[2] It was not possible to use conveniently a phonological representation of Chinese because of the monosyllabic morphology of Mandarin. The pictographic origin and evolution of writing in China comes from the fact that, to differentiate between dozens and more homonyms, to distinguish their meaning required a visual as opposed to a phonological representation. Chinese ideograms are now renamed 'logograms' because the most of them are instantly recognized as words and not as images. Reading them however still require a different strategy, a different kind of imagination supported by context instead of presenting a self-sufficient text as with the alphabet.

This resistance to division also affects the relationship to nature and ethics. Social harmony overrides every other consideration to the point that even holding secrets has long been considered uncivil in traditional Chinese society. It is no surprise then that one of the most stunning interpretations of the digital transformation in China has been to inspire and develop 'Social Credits,' a comprehensive digital valuation and sanctioning method that puts all citizens under automated state control. Regarding values, this approach is radically opposed to the principle of autonomous self-determination of western individuals who share the control of their behaviour with the law and democratic institutions.

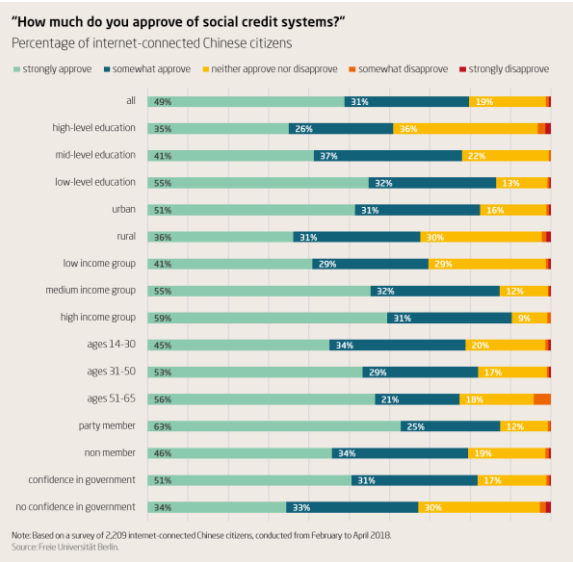

Figure 22 "How much do you approve of social credit systems?" (Check rights)

A similar questionnaire conducted in Europe or North America would most likely inversely mirror the results with "strongly approve" in the red and "strongly disapprove" in the green, reflecting the same inverted proportions in the intermediary bandwidths. This is only to show that culture does indeed support values, and that the digital transformation can do no more than to reflect them in how it sets up priorities. Without intending to take an ethical position between the two cultures, it is fair to suggest, all things being equal, that if social credits represent one of the key figures of the digital transformation in China, it is the personal digital twin that is about to demonstrate the unconscious drives of the West.

3.8 Comparing 'Social Credits' to the Personal Digital Twin

Figure 23 The impact of the Digital Transformation

Like his Vitruvian antecedent, the digital twin in the West develops primarily (although not exclusively) in splendid isolation. Consequently, the West's interest in twinning machines and humans is more a reflection of an individualistic interpretation of the digital transformation. People's decisions are generated on the basis of their previous experience, knowledge, and wisdom that they possess. The more people delegate mental tasks to their smartphones, the shorter they keep them in their head, and the less they exercise their own minds. This is awkwardly true about an individual's PDT. For example, with a twin of Patrick Henz, it is not only brainwork (memory, understanding, judgment, and choosing) that he trusts to a machine, it is also the very center of his being—the origin and place of his decision-making, his ego. So much mental labor saving has had the effect of changing his mental routines from addressing a question to himself to asking help from his phone or iPad. So, quite naturally, his brain changes its role from emitter to receptor, and he becomes his own echo chamber with simplistic notions, beliefs, and ideas rolling and churning witlessly. This the threat of populism and mental capitulation to the people who still know how to run the machine.

Furthermore, even in the West, if only to confront the perils mentioned above, most decisions will not only be taken, but, as in the East, eventually also implemented by machines. The same logic of the necessity of social control of very large numbers that prevails in China will also eventually come to roost in the West in front of what will require huge sacrifice of personal wishes and freedom. The big new questions will be surrounding programming the digital twin to guarantee an appropriate level of autonomy and responsibility to its physical counterpart. Because of externalizing so much of our cognitive content and functioning on screens, coupled with ever more invasive practices of tracking us everywhere, western societies may be on the brink of a deep psychological change. The main issue is responsibility.

3.9 Actions, Accountability, and Cultural Differences

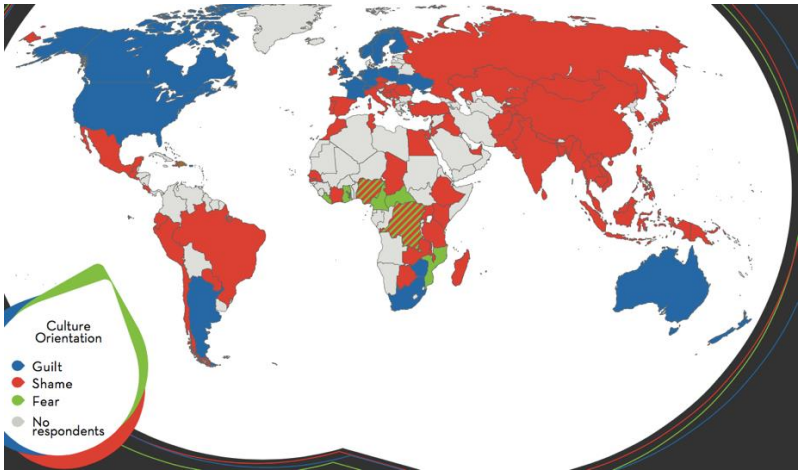

Figure 24 Responsibility (check rights)

Anthropologists have identified two main typologies of cultures depending on whether one's primary responsibility is directed to oneself (guilt) or to other people (shame). In the map above, the distribution is very clear. Figure 24 represents blue as 'guilt culture countries,' or those that primarily use alphabetic literacy. Some of those appearing in red, such as Mexico, Brazil, and Russia, combine literacy and strong local features (tribal or political) to prioritize social responsibility over what one owes to oneself. More detail should be explored to estimate what variables play into defining behavior, but in the East, where the community predominates, the priority responsibility has been to the other. Christianity, on the other hand, is a religion of the self, developed after the spread of literate techniques, and perhaps because of them. Christians are responsible to a personal, intimate God—hence naturally prioritizing privacy, secrecy, and the ritual of confession. Having already externalized in their smartphone's memory, judgment, discernment, and choices for such a long time, entrusting them to a sort of mental-labor-saving-device, people are left with very little valuable content accessible without resorting to an external device. If we consider that externalization conditions our power of choice, we must also take into account that all of this calls free will into question. The problem with the PDT taking over is that personal responsibility, along with privacy and secrecy, will also be externalized.

Figure 25 The Machine-State

Although subject to much exaggeration and misinterpretation, the fact is that 'Social Credits,' whether restricted to reigning in the power of private enterprises or extended across the board to all Chinese citizens, is a coherent application of the digital transformation to a culture steeped into four millenaries of community rather than individualist preoccupations. It is only a matter of automating the State. People and businesses are subjected to constant automated and recorded surveillance, and given credits or discredits accordingly.

Figure 26 Global distribution of guilt versus shame societies (Check rights)

The main and massively significant difference not only between East and West, but even more so between literate and digital cultures, is the subject of determination. The individualist West prides itself with self, or auto-determination, otherwise known as 'free-arbiter.' Social harmony and conflict-avoidance rest with the responsibility of the individual person, and checks and balances are ensured by the three levels of jurisprudence—legislative, executive, and judicial. That is also the basis of democracy. By comparison, social behavior and harmony in the East is imposed from outside the individual. The person in a generalized hetero determination, supported not by deliberation between different bodies of jurisprudence, but by automation of surveillance, and a consequent rewards and punishment system is also automated. In essence, digital technologies replace the 'common sense' of behavior control adopted for millenaries.

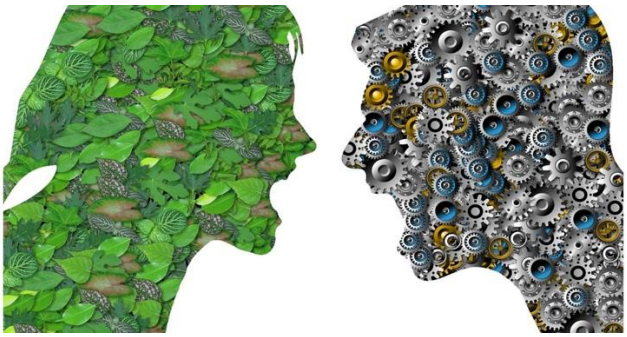

Figure 27 and what about these responsibilities?

3.10 Conclusion

Coming back to 'common sense,' it too is related to ethical responsibility. It is precisely because sense is not shared unquestionably by all humans today that humanity, now globalized by the pandemic (much more than by politics or business), is entering a major epistemological crisis. Sense is not common anymore, not only because global communication systems have brought people together that have different values, but also because the digital transformation has endowed people of the same culture to share dissent forcefully and virally with others, whether rightly or wrongly. Assuming then that PDTs owned by a community of 'users' are programmed to better relate to each other, will some form of 'common sense' be programmed to make them better understand each other? On what basis? With what level of flexibility or autonomy? This is where the fragile notion of 'ethical programming' comes into play, but until now, it hasn't been promising insofar as cultural biases are unavoidable and algorithmic unpredictability can offset software regulatory provisions.