Additional Useful Knowledge (90549)

# A Wrap-up of the AUK Activities

Federico Minutoli
MSc in Computer Science

## 1 Introduction

In the following sections I will list all the activities that I attended from October, 2019 until September, 2020 in order to fulfill the **50 hours** required to gain my *Additional Useful Knowledge* (AUK, for short) credits, and I will provide a little bit of context for each of them.

## 2 Academy of ZenHack

This activity accounted for **12 hours**.

In the period from October 11, 2019 to November 29, 2019 I attended six lectures of two hours each proposed by the Academy of ZenHack to raise awareness on ethical hacking.
Cybersecurity attacks causing data breaches and failure of critical systems and infrastructures are on the rise and the current response is not effective enough. For instance, a recent one targeted private data on German Chancellor Angela Merkel and other senior German lawmakers.
As a counterpart to the number and complexity of attacks, there is the need for a competent workforce to prevent or mitigate such threats effectively. But, at the moment, the gap between academia and real-world skills is huge and academia cannot provide students with sufficient skills. Capture the flag (CTF) competitions have been established as a great tool to achieve this goal, since they encourage students to think as an attacker does, thus enforcing more awareness on the modalities and consequences of an attack. The Academy of ZenHack supports this approach through a series of meetings in the form of seminars, formal classes and hands-on training exercises which aim to train students in scenarios as close to real-world ones as possible. By the end of the program they should be ready to participate in the final CTF competition, sponsored by Boeing, which is held on-site to test the skills they should have acquired at that point.

But first of all, what is *Hacking*? It is a branch of computer science which operates to understand how systems function and communicate over a network, how they are designed, how they are protected or whether they are vulnerable, etc. An hacker that performs these activities under proper authorization is known as an ethical (or white hat) hacker, as opposed to one which applies the same skill set to perform malicious deeds, addressed as a black hat hacker.

Over the course of those six weeks we covered a wide range of topics for a basic understanding of what's under the hood of ethical hacking. We started off by reviewing the basics of Linux to gather a crisp control over the command line interface and the Linux ecosystem in its entirety.
Then we delved into network protocols, mainly TCP- and UDP-based, and how to analyze network traffic through the open-source tool, *Wireshark*. Later on, we analyzed the most prominent threats to web security both on client and server side and how to provide the necessary security measures that any web service requires nowadays, avoiding potential SQL injections, denial of service (DoS) or cross-site-scripting (XSS) attacks. The meetings progressed scratching the surface of cryptography with a focus on what are the peculiar properties that define a problem as a crypto problem, while digging into the basics of symmetric- and public-key cryptography (for instance, the RSA algorithm [17]). A brief primer on machine learning techniques applied to cybersecurity followed shortly after wherein generative adversarial networks (GANs) show great results despite requiring a domain-specific taxonomy. In fact many insights gathered from computer vision don't work as well in this domain for a number of reasons: i) - The malicious functionality has to be kept intact

in the perturbed sample. ii) - Many feature types (string appending, API calls, etc.) that can be altered have different levels of difficulty. iii) - Small perturbations, such as altering a few pixels in an image for computer vision, don't have an immediate counterpart for cybersecurity features. iv) - Executables are more complex and varied in format than images. The seminar lectures concluded with an understanding of the realm of binary analysis: how to extract meaningful information from a pre-compiled program and identify noteworthy exploits that would make an attacker able to control it the way he wants it to, instead of how it is intended to. A classical example was proposed in the form of the string buffer overflow in C language [8, Chapter 4], through which an attacker may load instructions stored into private memory areas he shouldn't have access to.

An extensive series of exercises was left to us on the online *training platform* of the Academy of ZenHack to turn those theoretical concepts into practice.

# 3 Open day @ DIBRIS

This activity accounted for **6 hours**.

On February 2, 2020 I took the role of mentor at the *Slow Rogaining* activity which was part of the 4th edition of the *Alternanza Scuola-Lavoro* internship between the Computer Science department and schools located between Liguria, northern Tuscany and southern Piedmont.

The internship followed an innovative approach revolving around team building: a dozen teams of interns would be formed to encourage the spirit of collaboration inside and outside the team. The teams would need to be well balanced in terms of gender, previous IT skills, school of origin and the five personality dimensions resulting from the *Big 5* questionnaire [3], conducted under the surveillance of the psychologist Prof. Francesca Vitali. Over the course of a few days the teams would be asked to perform a series of different activities within the DIBRIS department each of which would earn them a certain amount of points. At the end of the last day the winning team would be announced as the one with the highest number of points.

Among those activities there was the Slow Rogaining, as well. The Rogaining is a variant of Orienteering, which takes place on variable distance and involves both the route planning and the navigation between control points or *stages*. Its "*Slow*" version eliminates the need to run, meeting a safety requirement imposed by the DIBRIS department, and instead emphasizes the problem solving ability of the team that will have to plan which of the fifteen stages to visit.
The aim of the Slow Rogaining activity is to consolidate the teams and to stimulate the teamwork among all the participants. A fundamental aspect of this activity is the need to set realistic goals and develop a long-term game strategy, since the teams will not be able to complete all the stages in the time available and will have to carefully take into account the trade-off between the time for completion and the value gained from it. The teams will indeed have to manage the namesake SlowRogaining app, developed by a group of MSc and PhD students in Computer Science for the experiential trainer Dr. Luca Gelati, inventor of the activity, to book the next optimal stage in their path to glory while it is still out of another team's sight.

Introducing a team building experience manifests itself as a valuable addition to the internship on many different levels. It represents a form of *approach to business dynamics* and strengthening of soft and transversal skills, fully consistent with the aims of the ASL internship.

Organizing this activity saw the employment of MSc and PhD students in Computer Science as inventors and mentors of the stages proposed during the Slow Rogaining as well as tutors of the interns monitoring their behaviour within the DIBRIS department. Being one of them, I partnered with my fellow colleague Luca Forneris and we proposed a version of the famous board game **The Mind** [26] re-adapted to the IT world. We thought this could have made for a meaningful stage because it was rather quick for the 20 minutes we had, yet not easy at all (it is one of those easier-said-than-done type of game), and also because it suited the ASL internship's objective since this game received the *Best Cooperative Game* award in 2019.

So, how does it work? Each round the players are given a fixed number of cards equal to the number of the round, whose only depicted information is a number between 1 and 100, inclusive. Their goal is to pile up the cards in a common pot in ascending order while being denied of any form of direct or indirect communication with one another. Timing is key then, and there is no clear winning strategy besides recognizing patterns in your teammates' habits in playing the cards, while improving as a team the more you play together. The full completion of the activity required clearing the 6th round (i.e., six cards per player) without running out of lives (in which case they would have had to start over). In order to help them, given the little time they had to familiarize with the game, and also to add a new strategic layer, we introduced some aids in the game that would have helped them shorten their way through the rounds. These aids were in the form of questions of increasing difficulty (matching a triplet of colors consisting of green, yellow and red) related to a wide range of topics they would have needed to be comfortable with in case they did choose to pursuit a BSc in Computer Science (physics, linear algebra, geometry, informatics, logic, chemistry and calculus), that they had to answer in under 30 seconds: answering them right would have cleared a certain amount of cards for them depending on the color, while answering them wrong wouldn't have had any negative impact on their gameplay. Their only notable drawback, besides time consumption, was that the aids were also limited to a total of 9 per game (divided into 4 green, 3 yellow and 2 red), so managing them well was another important key to be successful. This activity awarded a grand total of 70 points (which many of the interns and I found to be a little too low for its actual difficulty) and each team was awarded by us a certain amount of those points not only on merit, that is, how far they managed to get through the rounds, but also on the cooperation the team showed during the activity, with a focus on how the ones that had it easier were able to give valuable pieces of advice to the ones which were struggling.

Later that afternoon, I joined as a tutor an introductory lesson to Python, scheduled for the interns in order to bring even those that did not have an IT background closer to programming.

# 4 Seminars

This activity accounted for roughly **10 hours**.

Before diving into this section, I just wanted to say that over the course of the first year of the MSc I was able to attend quite a lot of seminars (standard format, an hour and a half each), mainly coming from MaLGa and PhD seminars, but also from RegML summer school.
Among those I would like to briefly mention two talks I attended on July, 1, 2020:

- *Geometric Deep Learning* by Prof. Emanuele Rodolà, from La Sapienza in Rome, Italy, on what is this umbrella term for emerging techniques attempting to generalize deep learning paradigms to non-Euclidean spaces. This approach helps with complex objects that naturally do not fit in a Euclidean domain or posses a grid-like structure, such as dense networks (applied in several scientific fields), manifolds, 3D meshes, etc. It is a relatively new branch of machine learning which had the first notable results only in early 2017, yet it still presents a lot of open problems, especially in the study of invariances in non-Euclidean spaces and in the relaxation of the need of priors on the data geometry [5].
- *Hilbert Space Representations of Probability Distributions* by Prof. Krikamol Muandet, from Max Planck Institute for Intelligent Systems in Tübingen, Germany, on possible ways to represent data in Hilbert spaces in order to compare possibly different probability distributions. A relevant problem in this domain is the *two-sample* problem [9], that is, given $m$ samples $X = \{x_1, ..., x_m\}$ drawn i.i.d. from the distribution $P$, and $m$ samples $y = \{y_1, ..., y_m\}$ from $Q$, can we say something on $P$ and $Q$? Are they any different?

## 4.1 3D Scene and Object Understanding

Since this report will already cover a lot of activities I'm not going to describe any further any of the other seminars I mentioned above. The one I would like to spend a few more words on, was instead held by Prof. Federico Tombari, from TU Munich, Germany, on *3D Scene and Object Understanding in the Era of Deep Learning*, streamed via YouTube on May, 5, 2020 [24].

AI has a juvenile history if compared to other scientific fields, with its birth dating back only to the early to mid 1950s. For a long time AI had the reputation of being *the next big thing* but it wasn't all fun and games for it, as it saw two cold winters transitioning from the '80s to the '90s. Of course, right now AI is back on the verge, but what's next? Will AI go into another winter? Or will AI maintain its current state wherein certain applications benefit hugely? Or will AI gain even more effectiveness and pervasiveness and reach unthinkable goals?

Despite all the milestones set already by AI, some people still question if deep learning (DL) has actually let us down.
To their point, there are still quite a lot of roadblocks that AI is currently trying to overcome effectively: i) Over-fitting, ii) Domain shift, iii) Over-confidence, iv) Efficiency and, last but not least, v) Data hunger as, no matter how many trillions of data samples have been annotated, DL models request for more and more to get as accurate as ever.

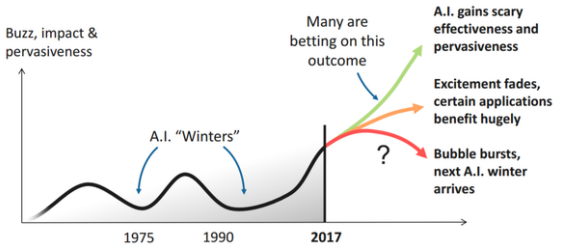


Figure 1: AI evolution

In all of this preface, there is one key point gone missing: what about 3D perception? As we all know, consumer- and LiDAR-based data are heavily used to generate AR for smartphones, head-mounted displays (HMDs) and automation. Some of these projects are really fascinating, yet challenging: for instance, self-driving is still a dream waiting to become real which, according to Moore's law on transistors growth [2], will not happen any sooner than 2035. Therefore, in the next few lines we'll try to understand what can and cannot be done as of now, focusing especially into 3D object understanding (monocular 6D pose estimation, point cloud embedding, etc.) and 3D scene understanding.

### 4.1.1 Monocular 6D pose estimation

6D pose estimation explores the problem of defining a 3D object's pose in terms of the 3D transformation from our viewpoint to the object's viewpoint. Starting out from a 3D model and considering a RGB-D representation of that, possibly given by a depth sensor in conjunction with the captured original RGB image, a transformation $T$ can be defined from the object's coordinate system $(x, y, z)$ to the viewer's $(x', y', z')$: this transformation usually implies six degrees of freedom, three for rotation and three for translation, which can be summerized in the *6DoF* matrix [21].
This approach is at the basis of augmented reality (AR), and it actually works pretty well in simple staged scenarios (static object, no occlusions, powerful hardware), but the field's interest is shifting towards what happens in more complex scenarios [22] wherein these luxuries are not granted.

Learning 6D object pose is something that's also conceivable without a 3D sensor, by first detecting the object's 2D bounding box in the image plane, and then, extending it into a 3D bounding box (parallelepiped) consistent with the object's viewpoint. Broadly speaking, there are two families of techniques regarding 6D pose estimation:

- *Direct monocular 6D pose estimation*, which interprets it as either a classification or regression problem to estimate the rotation's 4D quaternion, while at the same time predicting depth as an extra output [27] in a multi-task learning fashion.
- *Perspective-n-Points (PnP) driven monocular 6D pose estimation*, that splits this problem into two subsequent phases: first it establishes a 2D-3D correspondence, by looking into the 2D projection of the 8 corners of the 3D bounding box, sometimes also associated to a key-point vector field, and then it applies PnP to estimate the 6D pose [28].

Computer vision as a whole has always had as one of its main challenges the need to annotate real data in order to perform supervised learning, and 6D pose estimation suffers from it any less. As an answer to that, the field has been shifting towards synthetic data which are cleaner in structure and come annotated by nature. No domain gap and over-fitting are two of the main phenomena that clamp this choice, though. One recent study [23], showed a surprising outcome:

a 6D pose estimator was trained on an annotated real dataset, and showed pretty good results (in the order of 80% precision) and then, it was tested on a synthetic dataset extracted by simplifying the objects present already in the real one. So, one would have expected a performance at least as good, but instead, the results were awful (not greater than 16% precision) which evidently showed how the model built a deceptive bias towards the training dataset. In order to narrow the gap a few strategies have been implied, such as generative adversarial networks (GANs) with the objective of turning the synthetic data into something slightly more realistic, or domain randomization via rotation, translation or illumination changes with a certain transformation function $T(R, t)$ where $R$ is the rotation and translation DoF matrix mentioned earlier.

But, at the end of the day a 6D pose estimator still is at most roughly accurate because it looks into a generic bounding box around the object. So, a further step which has been getting more and more popularity is to apply a network dealing with 6D pose refinement [13]. The idea is that by looking at the object in different poses iteratively, the network is able to refine the transformation $T$ to approximate the best possible bounding box. In order to do so, proxy losses with distance transform [1] are typically implied to compare the estimation goodness in the more accessible top-down 2D plane projection rather than in 3D.

Another issue to keep into account in 6D object pose estimation is ambiguity. Let's just think of a breakfast mug: you can rotate it around from multiple viewpoints, but the 2D acquisitions out of it will mostly look exactly alike to one another. And this is something not that sporadic in vision tasks, because sometimes uncertainty is inherent in the task itself. In order to deal with it, meta-learning procedures have been implied to take into account all the possible ambiguous outcomes as valid heads into a multiple hypothesis predictor (MHP), which constructs a Voronoi tessellation out of which the most probable one is determined [19]. This approach has shown greater results in the ability to map the underlying probability distributions starting from simple toy problems like next frame prediction [30], up to complex 3D pose estimation problems (i.e., human body joints).

#### 4.1.2 Point cloud embedding

Lots of application falling into this category rely on 3D sensors, which can go from more expensive, precise and robust outdoor sensors to small mobile-embedded ones. The goal is to find the most suitable 3D descriptors given a 3D representation. For a long time this process was hand-crafted with researches trying to find the combinations of normal vectors that would best describe the 3D object they had in front. Of course, nowadays we are shifting towards learned descriptors which follow the standard end-to-end learning paradigm, computed entirely by a neural network.
What's interesting here, though, is how to learn features from voxel maps (that is, the 3D counterparts to 2D pixels) via 4D convolutions, and it's not as easy, because voxels carry much more information and the risk of memory saturation is tangible as the dimensions grow. Point cloud embedding is also a relevant tool towards unsupervised feature learning, trying to extend auto-encoders to 3D [15] and learn via reconstruction loss (e.g., the Chamfer distance [6, Section 2.2]).

Capsules [20] are one of the most interesting approaches to look into learning unsupervised embeddings of point clouds. They are a good alternative to a standard CNN wherein each neuron doesn't output a value, but instead it outputs a vector containing a richer set of information:
for instance, primary capsules (that are, the first processing layer of the input) will contain informative vectors regarding the pose, the appearance, the viewpoint or the illumination conditions. By applying capsules to 3D auto-encoders, they have shown a tendency to capture a latent representation of specific parts of the 3D objects which are semantically meaningful [29]. This allows to simply apply a further standard CNN on their output to perform 3D parts classification, or also, substitute specific parts of a 3D object by manipulating the corresponding latent representation (for instance, an airplane's tail, wing or core body).

#### 4.1.3 Monocular 3D reconstruction

One of the most interesting 3D scene understanding problems is 3D reconstruction, whose goal is to create a RGB-D representation starting from RGB images, which can be accomplished easily with a sequence of frames (monocular SLAM [14]) or with multiple partially overlapping images. This is an area which may have many implications for mobile, as well, allowing the development

of many great tools that require a pseudo-real-time knowledge of depth. But, what if we don't? What if we only have a single static image? To tackle this problem, recent studies have implied the use of a CNN to predict a dense depth map from a single RGB image, assuming the CNN has a big enough receptive field to learn monocular cues (i.e., relative size, occlusions, perspective). This is not a task suitable to standard CNNs, since they are usually not suited to regress high-dimensional outputs, which would cause memory explosion and loss of fine details. Instead the standard procedure nowadays is to use fully convolutional networks (FCNs) that will act as auto-encoders looking for a particular embedding. Many approaches have been followed which have led to great results, mainly in indoor environments [10]. In fact, one of the main issues this field is seeing at the moment is domain shift and the capability of alternating between real and synthetic data without a noticeable hit in performance (a frequent bottleneck, e.g., with MegaDepth [11]).

# 5 Adotta un Talento

This activity accounted for **30 hours** and counting.

Starting from June, 2020 I have been attending lectures offered by some of the main Ligurian companies employees, as part of the *Adotta un Talento* project. This project was organized by Digital Innovation Hub Liguria and Confindustria to promote a tighter collaboration between the University of Genoa and the companies across Liguria. It's meant to be a scholarship complementary to the academic path, which is divided into two different phases: a teaching period lasting till the very end of the I academic semester, and a traineeship with one of the adherent companies for the duration of the whole II semester, which could also culminate into the MSc thesis.

Besides a few introductory weeks dedicated entirely to company presentations (see Section 5.1), the main topics of the lessons have been revolving around the term *Industry 4.0* (see Section 5.2), which is usually used to indicate the technological Fourth Industrial Revolution, and the dense landscape of *soft skills* (see Section 5.3), which is an umbrella term for inter- and intra-personal abilities that are extraneous to a specific domain, as opposed to hard skills.

Thus far, I have attended 74 out of the 89 hours scheduled for lectures for the months of June, July, September and October, 2020. Due to privacy reasons, and a veto on sharing outside of personal use by some of the AuT organizers, I will omit the precise listing of lectures and the repository containing material and recordings of the lessons. Of course, if that's required, we can agree on an offline modality to share this information.[1]

## 5.1 Company presentations

The *eight* adherent companies to the AuT project all held short presentations (about an hour each) about their structure, mission, objective and how they embraced the 4.0 transformation.
In addition to that, we, as candidates, all had to undergo interviews with each company in a speed-date format which lasted a grand total of four hours, on September 14, 2020.

I will shortly list them all, along with the key points from their presentations.
Please refer to this nomenclature: a limited liability company corresponds to an Italian S.R.L., while a joint-stock company corresponds to an Italian S.p.A.

### 5.1.1 ETT[2]

The presentation held on June, 29, 2020 described ETT, a joint-stock company with seeds in a plethora of market sectors, ranging from the luxury industry (applications for Bulgari and Prada) to the industrial automotive (partnerships with Leonardo and Ansaldo). The diversification of interests makes ETT a prosperous company, with an annual turnover from 13 to 16 million euros.

In all of these fields, ETT tries to adapt 4.0 technologies, with a focus on augmented (AR) and virtual reality (VR) applications, the collection and processing of big data to conceive AI systems, the design of wearable devices, multi-channel solutions with parallel devices and gamification (i.e.,

[1] Contact me at *fede97.minutoli@gmail.com*, if necessary.

organizing content according to videogame-*esque* mechanics). The key behind turning these demands into fruitful ICT solutions is the integration between humanistic and technical knowledge: humanists work side by side with engineers in order to approach ICT problems from different points of view and assemble the most well-constructed piece of 4.0 software.
Among the most famous applications developed by ETT, there are surely on-site visits, for instance within museums or art galleries, wherein a crucial aspect is immersiveness. Visitors have to turn into *visit-actors* and live engaging experiences, which must vary depending on how they interact with their surroundings in order to make every visit unique.

### 5.1.2 Liguria Digitale[2]

The presentation held on July, 2, 2020, described Liguria Digitale, an Italian public company, that is, a company for which the majority of shares on the stock exchange belong to public bodies. Liguria Digitale has an annual turnover ranging from 30 to 40 million euros and it is structured in two stand-alone vertical compartments: one for software and IT infrastructures entirely dedicated to public health and the other dedicated to all the remaining issues related to public administration, such as agriculture, environment, culture, employment, etc.

### 5.1.3 ABB

The presentation held on July, 6, 2020 described ABB, a half-Swedish, half-Swiss electrotechnical multinational company based in Zurich. ABB is a pioneering technology leader operating in robotics, energy sustainability and automation in over 100 countries, with an annual turnover ranging from 26 to 30 billion dollars. It is involved in many different types of activities, for example the production of equipment for industrial automation, the development of cooperative robots that can be placed side by side with humans in production plants and the design of public cabinets that are able to recharge electric vehicles (EVs). The latter activity, and more generally, ABB's strong involvement in the supply of electricity (just think of turbine installments to produce electricity or the deal with Formula E to develop ever more capacitive batteries), has made ABB increasingly bounded to the 4.0 digital transformation process, and consequently to intense software and hardware development activities. The aforementioned cabinets, for instance, often require the development of applications using the Open Charge Point Protocol (OCPP), an application protocol for the communication between EV charging stations and a central management system.
In particular, a unified platform is often responsible in this scenario for managing charging operations and monitoring the current operating status of the cabinets in order to detect anomalies.

### 5.1.4 Tenova

The presentation held on July, 20, 2020 described Tenova, a joint-stock company that deals mainly with the design and development of innovative technologies and services that improve the business of customers in the industrial automation field. Tenova has a prosperous annual turnover ranging from 3 to 4 billion euros, which finds its roots in the development of solutions that help mining and metal companies reduce costs, save energy, limit environmental impact and improve working conditions for their employees. Most projects developed there revolve around the 4.0 concepts of predictive maintenance, remote assistance, process optimization, robotics and 3D printing.
Among this varied lot, an intriguing use case is represented by scrap metal classification [18]: a device equipped with a (multi-spectral) chamber runs a computer vision algorithm (i.e., YOLO) in order to identify the type of scrap carried out by each truck being inspected. This is performed because metal scraps are then thrown into an electric oven in different proportions depending on the type of scrap itself, hence this prevents potential anomalies. This problem is quite challenging, in fact, because metal scraps look alike for the most part, at least from a camera's POV.

### 5.1.5 Fincantieri

The presentation held on September, 3, 2020 described Fincantieri, a joint-stock company operating in the shipbuilding sector and, currently, the most important shipbuilding group in Europe, with an annual turnover ranging from 5 to 6 billion euros. The activities performed by Fincantieri are related to the design and construction of cruise ships, ferries, military ships, mega yachts, and to their transformation and repair activities. Fincantieri is involved as well in production activities that are meant to support shipbuilding or act as after-sales services. Having such a long history

(which we could appreciate during a four-hour on-site visit on September, 26) Fincantieri went from being a full fledged company in the 1980s, where even woodmen would be hired to design cabins or internals, to making great use of local companies that produce all the necessary supplies. As a global conglomerate of companies, Fincantieri is able to diversify its activities as much as possible with more than 100 million euros yearly invested in research and development (R&D) in order to improve ship building and ship performance, adopting complex VR or agent-driven simulations that can mitigate risks in ship design, balancing and implementation.

### 5.1.6 sedApta & EdiSoftware[2]

The presentation held on September, 9, 2020 described both EdiSoftware and sedApta.

EdiSoftware is a limited liability company with an annual turnover ranging from 3 to 6 million euros which deals mainly with enterprise resource planning (ERP) systems, such as applications accounting sales and purchases, managing warehouses and production or supporting managerial decisions in companies by integrating information coming from different business divisions.
The main technologies used in EdiSoftware belong to the .NET landscape and are blended together with all the others via Service Oriented Architectures (SOA), in which software is developed to be suitable to support Web services and to ensure inter-operability between different systems.

sedApta is a joint-stock company with an annual turnover ranging from 6 to 30 million euros which belongs to a group formed by the union of multiple high-tech companies based in Italy, Germany, France and Brazil specialized in developing smart solutions for Industry 4.0. Its customers mainly come from ERP, and put a lot of emphasis on ways to optimize their supply chain, production processes and production planning: sedApta indeed offers to its clients a suite of integrated solutions for planning, execution and optimization at all levels of the supply chain, starting from the simple directives sent from the ERP system, to the manufacturing operations management (or MOM) system, which is the component effectively in charge of executing each order.

### 5.1.7 Cetena

The presentation held on September, 14, 2020 described Cetena, a joint-stock company belonging to the Fincantieri group with an annual turnover ranging from 12 to 15 million euros, which deals with R&D in the naval and maritime field and offers technical services to shipyards, shipowners, navies and maritime operators. Cetena has focused its mission on performing applied research to study and develop practical applications, tools, processes and expertise to be used for ship testing, naval and maritime simulations and engineering consulting in general. Research projects from Cetena are commissioned by clients in order to improve naval materials or to render shipyards development, general life and consumption more sustainable and efficient: some examples come from tests commissioned by Costa and Virgin ships, such as mechanical vibration resistance tests, shock resistance tests for equipment and devices or airborne noise insulation index measurements.

## 5.2 Industry 4.0

The technological Fourth Industrial Revolution, around which Industry 4.0 established its paradigms, concerns the automation of traditional manufacturing and industrial practices by means of modern smart technology, large-scale machine-to-machine communication, internet of things (IoT), artificial intelligence (AI) and robotics in order to make decision-support systems more precise and to identify and diagnose production issues with little to no human intervention.

Due to the broad nature of the topic, I decided to split the discussion on Industry 4.0 in four different subsections by category of membership, plus a bonus extra in Section 5.2.5.

### 5.2.1 Market evolution and Meaning of the 4.0 transformation

The 4.0 society is evolving towards a more and more rooted human-machine interaction: as quick as days pass by more aspects of every day life are being digitized all around us by multinational (or even local, like Liguria Digitale) companies which invest entire capitals to make it happen.

[2]This is an Italian-only website.

Nowadays, every kind of industrial business revolves around the use of sensors, computers or other digital devices that, through low-level protocols and IoT, are able to collect data about employees' working activity over time, yet, at the same time, are able to make them interact with those very same on-going sensors, allowing them to play an integral role in the production chain.

The tendency to digitize every aspect of working life has also led to the birth of smart societies, where everything is connected in a network of digital components, which are used to improve (or are supposed to improve) our life quality. For example, modern buildings can be designed and constructed as smart entities, wherein sensors talk to automatic control processes to monitor the building's heating, ventilation, air conditioning, lighting, security, etc. In order to do so, a *digital twin* [12] of a city is usually modelled in order to map the environment and provide an efficient simulation of how the real system will perform when deployed. These models represent one of the staples of Industry 4.0, albeit a deep inter-disciplinary expertise is usually required to make them: it is no longer sufficient to vertically specialize in a single technical field, because 4.0 real life problems are so complex that they require horizontally expanded minds.

### 5.2.2 Enabling technologies for Industry 4.0

In such an entangled landscape, where multiple domains interlace one another, the core of Industry 4.0 is without a doubt data, and how technological components can interact with each other to exchange this data at various levels of abstraction of the production chain. According to experts [25], within this fuzzy puzzle that Industry 4.0 is, there are ***nine*** enabling technologies adopted by companies in their pipelines that have made the magic happen:

*Autonomous robots*, and software applications for "4.0 industrial automation", through which machines are all connected to each other, and can interact and cooperate side-by-side with humans to support the execution of production processes. The main difference with "traditional" industrial automation is that, before Industry 4.0, robots were not enough technologically advanced in the sensitivity of their surroundings to be considered non-dangerous for humans.

*Additive manufacturing*, the use of 3D printers, which have drastically shortened the process of going from a draft model to the physical product.

*Augmented reality*, or AR, the superimposition of one or more informative digital elements on top of common everyday life scenes. AR can be used to simplify workflows, to propose plausible alternative perspectives or to enhance some of their parts: for example, parts of a complex object can be magnified through AR to quickly retrieve a broader description of their functionality.

*Simulation*, by means of the aforementioned 4.0 digital twins.

*Horizontal/Vertical integration*: the 4.0 cloud-based technologies that alleviate the integration between different levels of the value and supply chain in the production cycle. They promote and support the information exchange between suppliers, customers and third-party sellers.

*Internet of things*, or IoT, the acquisition of raw data coming from machineries connected to the Internet and equipped with sensors that make them aware of their and their environment's current state of operativeness. A stage of information extraction, also by means of ML algorithms, can later be performed to detect anomalies or malfunctions or to monitor the quality of production.

*Cloud computing*: the offering of services that let external actors exchange data with a "cloud" of devices invisible to those who buy and employ said services. The services are usually provided with a pay-as-you-go (PAYG) billing format and can vary depending on the necessary infrastructure between three possible models that offer different features: infrastructure as a service (IaaS), platform as a service (PaaS) or software as a service (SaaS, like many ERPs are).

*Cybersecurity*, the counter-measures adopted to help productions systems avoid cyber-physical attacks as the incoming and outgoing data exchange grows. These kind of attacks represent security breaches in cyber space that have a negative impact on the physical environment.
For instance, a malicious user could take control of the communication channels of water pumps, pipeline valves, etc., and cause damage to machines and devices, consequently putting lives at risk.

*Big data and analytics*: the 5-step cycle of acquisition, extraction, integration, analysis, interpretation and decision that data collected with IoT goes through to infer additional clues.

### 5.2.3 The importance of R&D inside companies

An interesting topic of discussion in the 4.0 landscape is how important research and innovation (R&D) has become for companies in terms of investment. There is a crystal clear correlation between the research efforts a company makes and its economic return. Ideally, companies should "fuel" research with money, research should generate knowledge that could be translated into practical applications whose sale to the market should allow the investors to earn even more money.
In fact, there isn't always a direct proportion between investment in research and economic return, because the developed technology **must** be appropriate for the market segment it addresses.
A specific technology may indeed be neither the most efficient, nor the most sophisticated, yet it may possess some features that make it sell more than one that actually is. For instance, professional cameras are much more advanced than smartphones', but they target only professional photographers which know how to extract the most out of them; most people, instead, including even Instagram influencers and the likes, are content with their smartphone's picture quality. Hence, investing in innovative professional cameras may not be nearly as profitable as developing something that focuses on the adoption of smartphones' cameras, since nowadays everyone and their grandmother has a smartphone in its pocket and that market is basically unlimited.

Generally speaking, the performance of newly researched technologies has a volatile behaviour, which starts as slow as steady until the development becomes mature enough. At that point, the performance growth accelerates following an exponential slope [4, Figure 1], which lasts for quite some time, after which the growth gradually slows down until it stops (derivative of 0).
That last segment, that is, when a technology is about to reach its limit, is the most dangerous time interval for companies to be in, known as the *dilemma zone*. Indeed, even if the current technology still has state-of-the-art performance, new technologies might be in advent and it's the companies' job to decide whether to proceed with incremental innovations in their own technology or to start investing capitals in the research of the new technology. This choice is crucial, because it could result in disruptive changes that could lead companies to bankrupt. For instance this scenario might happen when a company introduces an innovative product that makes customers lose interest in all the previous products of the brand, as they would be targeted as obsolete.
In order not to let this happen, and lose a big slice of the market, the R&D unit mustn't be disconnected from the market, thus requiring a cohesion between technical and managerial side.

### 5.2.4 Industry 4.0 impact on companies' competitiveness

Not only has the market changed in the 4.0 era, but also the approach companies adopt towards the market itself and towards competitors in the same segment. In particular, the rate of change of the current digital transformation is exponential, as companies incorporate more and more AI technologies in production environments. Not only does this make technologies obsolete in no more than a few years time, but it also allows companies to increase their capital to a billion much faster than before since the average lifespan of successful companies has decreased over the years.

The result of the adoption of the Industry 4.0 paradigm on the competitiveness of companies is the following classification of companies in *three* groups:

- *Disrupted companies*, that were not able to react to the change, despite having once been total market leaders in their segment (i.e. Kodak, Blockbuster, etc.).
- *Market destroyers*, that were able to win against old competitors and took over the prosperous core business those ones once had in their segment.

- *Half-corporations*, that are in a in-between situation, as they have slowly adapted to new technological innovations, but have lost the market leader position.

Such scenarios present themselves because innovations can consist in adoptions of 4.0 technologies in industrial domains that originally did not involve their use, which can favour companies that are just starting their business in that specific domain, as opposed to older, more structured and more hostile to change market leaders. An example of this situation is the disruption introduced in the sensors' market by the Waze company, worth over a billion dollars: Waze, in fact, developed a mobile application that used the cell phone as a sensor to gather data about traffic and predict the best path to be followed by a car, and as a result there was no need for Waze to buy physical sensors from specialized companies that produced them. *Here*, the full story.

### 5.2.5 Leftover topics

We were given by Cetena a primer on families of certifications used to establish that a company knows how to work in the proper way. They are issued by impartial third-party organizations and they are an indicator of good management of specific technological domains, which in turn makes companies preferable to non-certified competitors. The *ISO 27001* certification, in particular, is a scalable international directive issued by RINA, which regulates how to manage, recover and improve information security in a company, where the adjective *scalable* means it is suitable for applications within both large and small organizations.

We were offered by EdiSoftware a broad look on the foundations of business intelligence (BI, for short) analytics in the Industry 4.0 era and practical examples on data coming from multiple real life data sources exploring Microsoft's Power BI offerings thoroughly. One of the most discussed topics was how to properly develop a BI dashboard, which reminded of the many *InfoVis* rules of thumb that we are currently seeing in the Data Visualization (90529) course, such as:
i) - Design focused on the user, ii) - Content specificity, iii) - Cleanliness and order, iv) - Perception vs cognition, v) - Data interaction and navigation.

ETT screenwriter Matteo Bonanno took us through the realm of storytelling, starting from the value inherent in narrating something that goes back to Homer's times, passing through Propp's 31 functions on plot evolution [16], to its digitisation through AR, animation and social media.

## 5.3 Soft skills

Soft skills is an umbrella-term that doesn't have a clear cut definition, but that generally involves the study of inter- and intra-personal behavioural models in order to improve the attitude towards addressing the world and solving its problems in a productive manner. They have a huge advantage on hard skills in not getting old, because humans were, are and will always be social animals.

Again, I decided to split the discussion in three chunks, with a bonus extra in Section 5.3.4 on the practical activities that we were asked to perform by Liguria Digitale to challenge our skills.

### 5.3.1 Types of soft skills

The main focus of the first few lectures offered by Dr. Carlo Brozzo was to unravel the broad and grey area of soft skills into *three* recognizable and discernable categories, which addressed the self, the ability to interact with others and the ability to cooperate, respectively:

*Individual* soft skills can be separated in: i) - Self-confidence, not vanity. ii) - Autonomy in overcoming diverse problems. iii) - Anxiety and stress management. iv) - Spirit of initiative. v) - Acquiring continuously new knowledge over time.

*Relational* soft skills can be separated in: i) - Ability to communicate with others. ii) - Ability to manage and integrate information coming from multiple sources. iii) - Ability to work together or as a team without slowdowns. iv) - Ability to fulfill objectives

*Work-oriented* soft skills can be separated in: i) - Flexibility. ii) - Ability to handle space and time, in order to respect strict deadlines and organize non-overlapping meetings. iii) - Ability to

relate with the work context, both with colleagues and with the main regulations defined by the company equal to all the employees. iv) - Problem solving (see Section 5.3.3).

### 5.3.2 Fundamental rules of communication

Some of the most important soft skills are the ones related to communication, because being able to express ourselves in a way that gets under the skin of other people is key to live a remembered, successful and accomplished life (e.g., think of Mandela). Communication can either be verbal, non-verbal (human gestures and postures), written and, now, online, which, despite being different, share *five* common axioms for producing an effective exchange:

1. Everything that we do in our life communicates something, and there's no way around that. *Non-behavior* does not exist, and words, gestures or even silence all carry a message value and affect others, who in turn respond to such communication.
2. Every communication attempt has a content and a relationship aspect. Within a message there is always an aspect of news in the content, which includes the information transmitted, and an aspect of form in the relationship, that concerns how the message is transmitted.
3. The nature of a relationship between two or more communicators depends on the inherent punctuation that the communication carries. During an interaction indeed a punctuation of the sequence of events is constructed in our minds: a different or unusual way of punctuating said sequence produces a conflict, because each speaker interprets his/her own behavior as a consequence of the behavior of the other speaker and never as a cause, which results in them playing a round of the *blame game* [7]. This kind of problem can be solved through a refinement of meta-communication, in which the focus goes onto the relationship rather than on the content of the exchanges being communicated (i.e., the how, not the what in the 5W).
4. Human communication is a stream alternating analog and digital codes, mainly corresponding to relational aspects and to content aspects, respectively. Analogue communication groups every non-verbal communication (body poses, facial expressions, gestures) and it is necessary to convey and define the relationship. Digital communication, instead, consists of words that serve to exchange knowledge via information, which are suitable to convey the content.
5. All communication exchanges are symmetrical or complementary, depending on whether they are based on equality or difference. Symmetric interaction is characterized by scenarios in which the behavior of speakers is reflected in the behavior of the others. Complementary interaction, instead, sees its birth in the opposite process, wherein the speakers may play different roles within an ecosystem (e.g. different positions within a company), and consequently the behavior of one will tend to complement that of the others.

A famous model used to measure the effectiveness of written communication is the so-called *5W model*, wherein each W corresponds to a specific feature that written communication must have. Its schema can actually be summerized as simply as with the phrase "Who, says What, through Which channel, to Whom and with What effect?"

Another model which is looked up to not only for written, but also online, communication is the *5C model*, whose rules nicely complement the five axioms above:

1. *Clarity*, the sender of a message must deliver a specific message. The message should also have very specific goals, because giving away too much information can confuse the audience.
2. *Cohesiveness*, the grammatical and lexical relationships that exists between different elements across sentences of a text, which can hold it together or untie it up.
3. *Completeness*, the communication should convey *all* facts required by the audience through relationships adequate to who is in front. The sender of the message must take into consideration the receiver's mind and skill set and convey the message accordingly.
4. *Conciseness*: the communication should convey *only* facts required by the audience expressed in a synthetic manner, in order not to lose track on the main topic of discussion. For instance, there's no need for unnecessary words, puns or even entire paragraphs.

5. *Concreteness*, the communication must be specific, well defined, and vivid rather than vague and generalist in order to leave the message stuck in the audience heads.

### 5.3.3 Problem solving

Problem solving is a soft skill defined as the ability of an individual to "find solutions to difficult or complex issues"[3], that is, the set of cognitive processes necessary to understand, dissect and resolve problematic situations whose solution is not trivial. A problem may arise as the result of the comparison of a reality datum (*present state*) with a reference (*desired state*), surrounded by the set of issues that need to be addressed in order to fill in the gaps. Although problems can possess very different nuances, they all usually arise due to issues within the relationship with oneself, with others or with one's perception of reality.

Many different approaches have been formulated over the years to address this issue with a systematic procedure. A good example might be a classical model, known as the *7-step model*, which is a general sequential procedure that can be followed in order to solve a problem:

1. *Problem identification and definition*, the first question to ask ourselves is whether a problem is actually a problem because, if the solution is trivial, it doesn't need problem solving at all.
2. *Data collection and information gathering*, our main ally in this whole process is knowledge, thus we need to study as much as we can about the problem. This might just mean to collect information about how similar problems have been solved in the past.
3. *Problem dissection into small parts*, the biggest step that we have to take is to decide where to start due to the overwhelming complexity a problem may possess. A possible approach to follow in order to reason on a solution with a more open mind, is to dissect the problem into smaller parts that can be tackled more easily one-by-one.
4. *Hypotheses formulation*, about the most plausible causes that triggered the problem in the first place following a cause-effect chain. At the end of this step, we should possibly be able to discriminate between those causes and identify the main one.
5. *Action plan research*, to think of, or search for, possible solutions for the problem at hand. A formulation of the action plan should follow with the definition of where, when and how to perform the intervention, considering the risk vs reward trade-off.
6. *Action plan implementation*, the implementation of the action plan developed at the 5th step, after we have made sure that destructive effects will not permanently affect our future work.
7. *Sanity check*, the assessment of our plan's aftermath.

In the aforementioned procedure, it is important to remember the *Occam's razor* principle: among the different hypotheses and solutions a certain problem may have, we should always choose the simplest one that still provides correct results. Moreover, when solving problems, we should combine the ability to reason via logic with creativity, which usually derives from lateral thinking. Creativity is a useful tool in problem solving, because it allows us to perform tasks faster, in a non-sequential manner and with less "brain friction"[4] w.r.t. plain logic.

Creative ideas usually arise in the following ways:

- *Creativity of innocence*, the creativity that emerges when someone doesn't know anything about the topic of the task being addressed. For this specific reason, he provides a different point of view on the subject, which can sometimes lead to creative solutions that would have been inconceivable to people who are too involved in the task.
- *Creativity of emergency*, the creativity that appears when someone is put under extreme pressure by the environment and he is able to respond. Extreme conditions start indeed rare cognitive processes that otherwise would not take place in one's brain.

[3] From its definition in Oxford English dictionary (OED).
[4] An emotional drop, due to uncertainty and ambiguity.

- *Creativity from experience*, the creativity that involves the application of already acquired knowledge in a novel way, different from standard solutions for a certain problem.
- *Accidental creativity*, the creativity that mistakenly leads to innovative solutions regarding type B problems while looking for solutions for type A problems.
- *Creative imitation*: the creativity that originates when someone takes a cue from somebody else who has already solved either the same, or a similar, problem.
- *Intuition of the gap*: the creativity that happens when someone has to guess the logical bridge in the middle that binds two well known extremes.
- *Identification of the niche*: the creativity that comes as result of the application of the famous "divide et impera" technique in order to decompose a problem in many smaller sub-problems. Under these circumstances, someone can stumble upon the so-called *niche*, which is an aspect of a sub-problem whose solution may change the way he thinks about the problem as a whole.

#### 5.3.4 Hands-on activities

Many hands-on activities were carried out during the set of lessons about soft skills, with the main objective to turn those theoretical concepts into practice. For instance, among the lot:

- *Speaking and presenting*, wherein we had to present ourselves to an audible audience in less than 5 minutes, by telling a catchy story that would keep the public's attention high.
- *Self-assessment*, wherein we had a limited span of 10 minutes to list some *non-4.0* activities that we feel the most natural doing while achieving peak performance, followed suit by those that instead feel the most unnatural with minimum performance.
- *Speaking improvisation*, wherein we had to come up with a meaningful story after looking for 30 seconds at a photo from the 1900s we had never seen before.
- *Problem solving*, wherein we had to solve a riddle in a limited span of 15 minutes.

## References


[1] "Distance transform," https://en.wikipedia.org/wiki/Distance/_transform.

[2] "Moore's law," https://en.wikipedia.org/wiki/Moore%27s_law.

[3] C. Barbaranelli, G. Caprara, A. Rabasca *et al.*, "A Questionnaire for measuring the Big Five in late Childhood," *Personality and Individual Differences*, vol. 34, pp. 645–664, Mar. 2003.

[4] J. Beierlein, L. McNamee, and F. Ledley, "As Technologies for Nucleotide Therapeutics Mature, Products Emerge," *Molecular Therapy - Nucleic Acids*, vol. 9, 10 2017.

[5] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam *et al.*, "Geometric Deep Learning: going beyond Euclidean data," *CoRR*, vol. abs/1611.08097, 2016.

[6] L. Dantanarayana, G. Dissanayake, and R. Ranasinge, "C-LOG: A Chamfer distance based Algorithm for Localisation in occupancy Grid-maps," *CAAI Transactions on Intelligence Technology*, vol. 1, no. 3, pp. 272 – 284, 2016.

[7] J. Decety, S. Echols, and J. Correll, "The Blame Game: The Effect of Responsibility and Social Stigma on Empathy for Pain," *Journal of cognitive neuroscience*, vol. 22, pp. 985–97, 06 2009.

[8] W. Du, *Computer Security: A Hands-on Approach*. CreateSpace, 2019.

[9] T. Fernandez and N. Rivera, "A Reproducing Kernel Hilbert Space log-rank Test for the Two-sample problem discernability," Apr. 2020.

[10] I. Laina, C. Rupprecht, V. Belagiannis, F. Tombari, and N. Navab, "Deeper Depth Prediction with Fully Convolutional Residual Networks," *CoRR*, vol. abs/1606.00373, 2016.

[11] Z. Li and N. Snavely, "MegaDepth: Learning Single-View Depth Prediction from Internet Photos," in *Computer Vision and Pattern Recognition (CVPR)*, 2018.

[12] A. Madni, C. Madni, and S. Lucero, "Leveraging Digital Twin Technology in Model-Based Systems Engineering," *Systems*, vol. 7, p. 7, 01 2019.

[13] F. Manhardt, W. Kehl, N. Navab, and F. Tombari, "Deep Model-Based 6D Pose Refinement in RGB," *CoRR*, vol. abs/1810.03065, 2018.

[14] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardós, "ORB-SLAM: a Versatile and Accurate Monocular SLAM System," *CoRR*, vol. abs/1502.00956, 2015.

[15] J. J. Park, P. Florence, J. Straub, R. A. Newcombe, and S. Lovegrove, "DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation," *CoRR*, vol. abs/1901.05103, 2019.

[16] V. Propp, S. Pirkova-Jakobson, and L. Scott, *Morphology of the Folktale*, 1st ed. University of Texas, 1968.

[17] R. L. Rivest, A. Shamir, and L. Adleman, "A Method for Obtaining Digital Signatures and Public-Key Cryptosystems," *Commun. ACM*, vol. 21, no. 2, p. 120–126, Feb. 1978.

[18] M. Robalinho and P. Fernandes, "Using Machine Learning for Scrap Metal Classification," *DCE, 3rd ed.*, Jun. 2019.

[19] C. Rupprecht, I. Laina, M. Baust, F. Tombari, G. D. Hager, and N. Navab, "Learning in an Uncertain World: Representing Ambiguity Through Multiple Hypotheses," *CoRR*, vol. abs/1612.00197, 2016.

[20] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic Routing between Capsules," *CoRR*, vol. abs/1710.09829, 2017.

[21] M. I. Shah, "Six Degrees of Freedom Point Correspondences," Jan. 2009.

[22] D. J. Tan, N. Navab, and F. Tombari, "Looking Beyond the Simple Scenarios: Combining Learners and Optimizers in 3D Temporal Tracking," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 11, pp. 2399–2409, 2017.

[23] B. Tekin, S. N. Sinha *et al.*, "Real-Time Seamless Single Shot 6D Object Pose Prediction," *CoRR*, vol. abs/1711.08848, 2017.

[24] F. Tombari, "3D Scene and Object Understanding in the Era of Deep Learning," The original video can be found at https://www.youtube.com/watch?v=-GR2-9GBNrI.

[25] J. Wan, H. Cai, and K. Zhou, "Industry 4.0: Enabling technologies," *Proceedings of 2015 International Conference on Intelligent Computing and Internet of Things, ICIT 2015*, pp. 135–140, 05 2015.

[26] W. Warsch and O. Freudenreich, "The Mind," 2018, An in-depth overview can be found at https://boardgamegeek.com/boardgame/244992/mind.

[27] Y. Xiang, T. Schmidt, V. Narayanan, and D. Fox, "PoseCNN: A Convolutional Neural Network for 6D Object Pose Estimation in Cluttered Scenes," *CoRR*, vol. abs/1711.00199, 2017.

[28] L. Zhao, X. Peng, Y. Tian, M. Kapadia, and D. N. Metaxas, "Semantic Graph Convolutional Networks for 3D Human Pose Regression," *CoRR*, vol. abs/1904.03345, 2019.

[29] Y. Zhao, T. Birdal, H. Deng, and F. Tombari, "3D Point-Capsule Networks," *CoRR*, vol. abs/1812.10775, 2018.

[30] Y. Zhou, H. Dong, and A. El Saddik, "Deep Learning in Next-Frame Prediction: A Benchmark Review," *IEEE Access*, vol. 8, pp. 69 273–69 283, 2020.