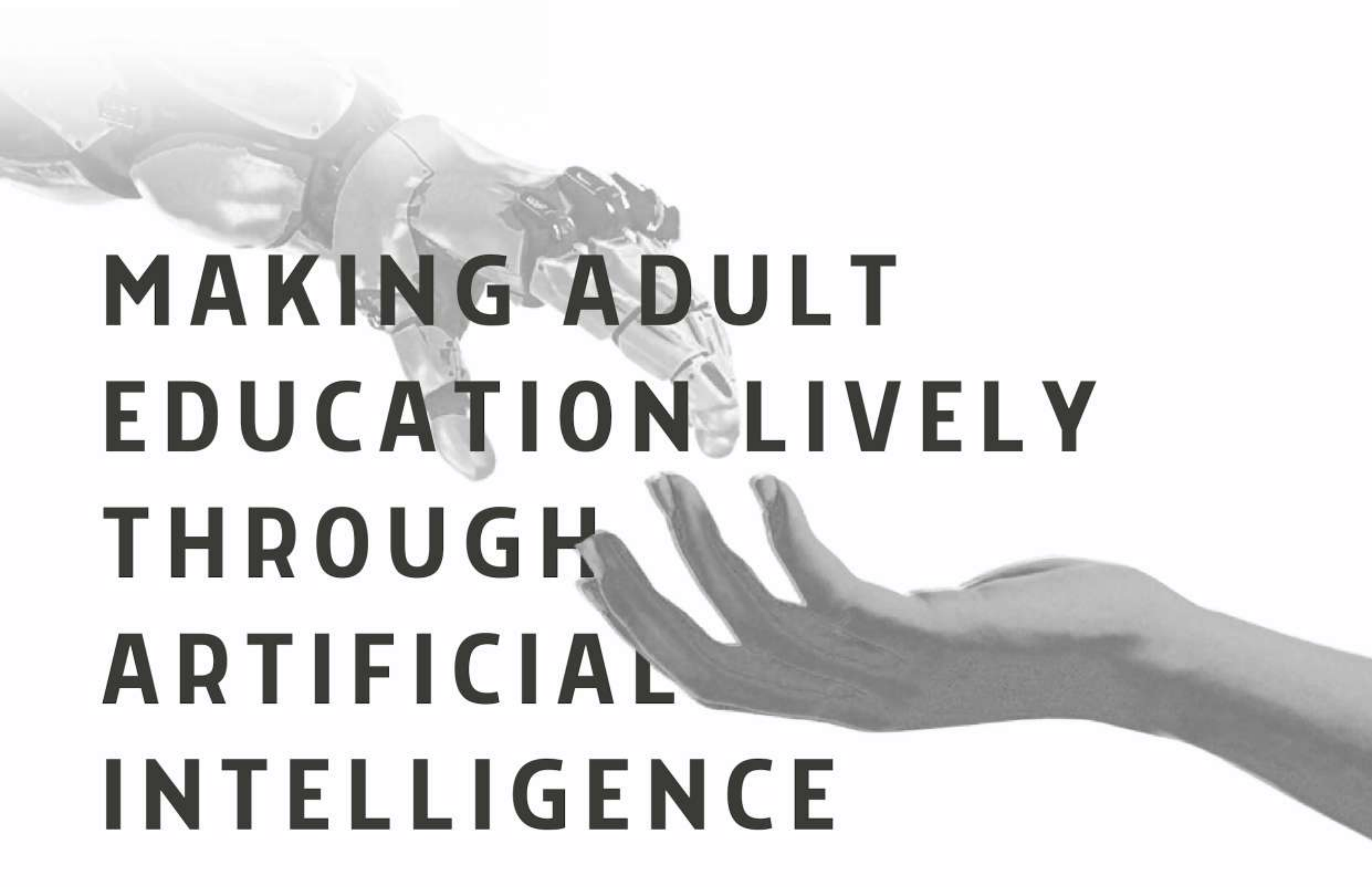
LivAI

Project no: 2022-1-ES01-KA220-ADU-000087381



MAKING ADULT EDUCATION LIVELY THROUGH ARTIFICIAL INTELLIGENCE

sepie

Funded by the Erasmus+ Programme of the European Union

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

This document was developed in the framework of the project LivAI with the collaboration of all the partners of the project.

"This document reflects the views only of the authors, and the European Union cannot be held responsible for any use which may be made of the information contained therein."

DISCLAIMER

The information, documentation and figures in this document are written by the LivAI project consortium (Grant Agreement KA220-ADU-D54476C0) and do not necessarily reflect the views of the European Union.

The European Union is not liable for any use that may be made of the information contained herein.

All the pictures used in this document are made by AI or are roaylty free.

ACKNOWLEDGMENT

This document was carried out with the kind collaboration of all the partners involved.
EFCoCert (Switzerland)
EProject Consult (Italy)
Finnova (Belgium)
Konnektable (Ireland)
Ubitech (Greece)
Universitat Jaume I de Castellón (Spain)

MORE INFO:

livai.uji.es

CONTENTS

000110011111000

01

Defining AI

.INTRODUCTION

The development of intelligent machines that can learn and carry out activities that would ordinarily need human intelligence is the focus of the quickly developing field of AI. Face recognition, virtual personal assistants, driverless vehicles, and predictive analytics are just a few of the areas that AI is already revolutionising. It is also altering how people live, work, and interact with technology.

LivAI sepie Funded by the Erasmus+ Programme of the European Union

1.1 WHAT IS ARTIFICIAL INTELLIGENCE

1.1.1 THE PROBLEM OF DEFINING ARTIFICIAL INTELLIGENCE

There is no broader consensus definition for the term ”Artificial Intelligence”. Just as there is no single consensus definition of intelligence.

In the book by Stuart Russell and Peter Norvig [1] ”Artificial Intelligence. A modern approach”, the authors present a quadrant to show the different approaches to defining AI (see Table 1). The quadrant has two dimensions, One dimension presents thinking versus behaviour, and the other dimension presents human/emotional versus rational. Each intersection of this quadrant shows a different approach to the definition of what AI is.

	THOUGHT	BEHAVIOUR
HUMAN/EMOTIONAL	Cognitive modelling	Turing Test
RATIONAL	Logicist	Rational agent

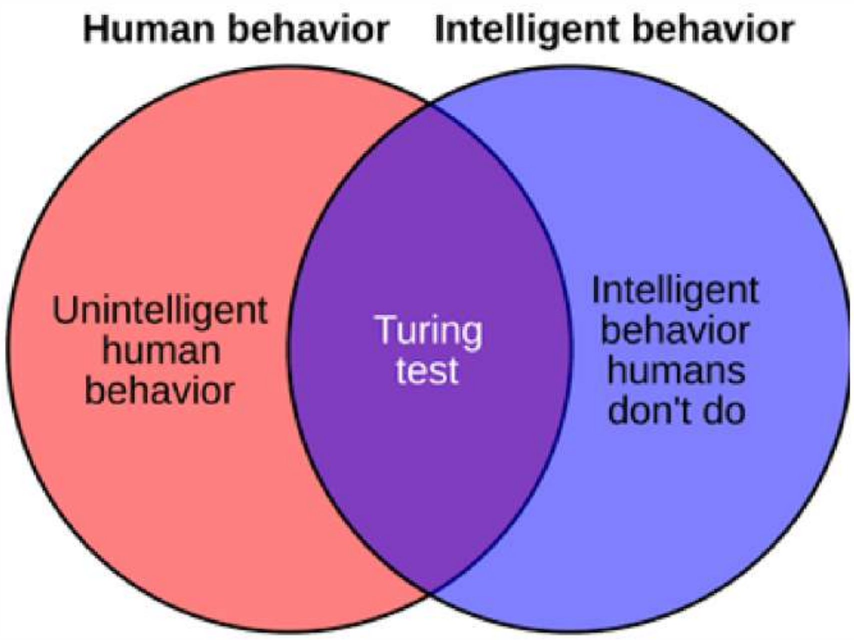
Table 1: Russell and Norvig quadrant on AI.

LivAI sepie Funded by the Erasmus+ Programme of the European Union

HOWEVER, there is a broader consensus on how to decide whether AI is present in an artificial computational system through a test. One of the first attempts to create a test to distinguish whether an artificial system can be considered intelligent or not is due to Alan Turing, who presented it in his famous article ”Computing Machinery and Intelligence” [2].

The Turing test measures a machine’s ability for showing intelligent behaviour that is comparable to or impossible to differentiate from that of a person. Alan Turing proposed the test in 1950, and it is still one of the most well-known and hotly contested issues in the field of AI today.

The Turing test seeks to ascertain whether a machine can pass for a person in a discussion using natural language. The exam involves a human assessor conversing with a machine and a different person while being unaware of which is which. The Turing test is regarded to have been passed by the machine if the assessor can not reliably tell the difference between the machine and a human. The Turing test continues to be a key milestone in the development of AI and influences how scientists and programmers approach the creation of intelligent machines.


Human behavior
Intelligent behavior
Unintelligent human behavior
Turing test
Intelligent behavior humans don't do

Human behaviour against intelligent behaviour.
The Turing test is not able to cover both of them.

However, the Turing test presents some weaknesses. The Turing test does not directly assess a computer's capacity for intelligent behaviour. It merely checks whether the machine acts like a person. Due to the fact that human behaviour and intelligent behaviour are not identical, the test may under or overestimate IQ in one of two ways: some human behaviour is unintelligent; some intelligent behaviour is inhuman.

Table 1 shows this weakness in a graphical way.

LivAI sepie
Funded by the Erasmus+ Programme of the European Union

1.1.2 A COMPUTATIONAL DEFINITION OF ARTIFICIAL INTELLIGENCE

AI is the simulation of human intelligence processes by computer systems. This includes learning, reasoning, and self-correction. AI systems use algorithms and statistical models to analyse and draw insights from large datasets, as well as to make predictions and decisions based on that data. The ultimate goal of AI is to create machines that can perform tasks that typically require human intelligence, such as visual perception, speech recognition, decision-making, and language translation, among others.

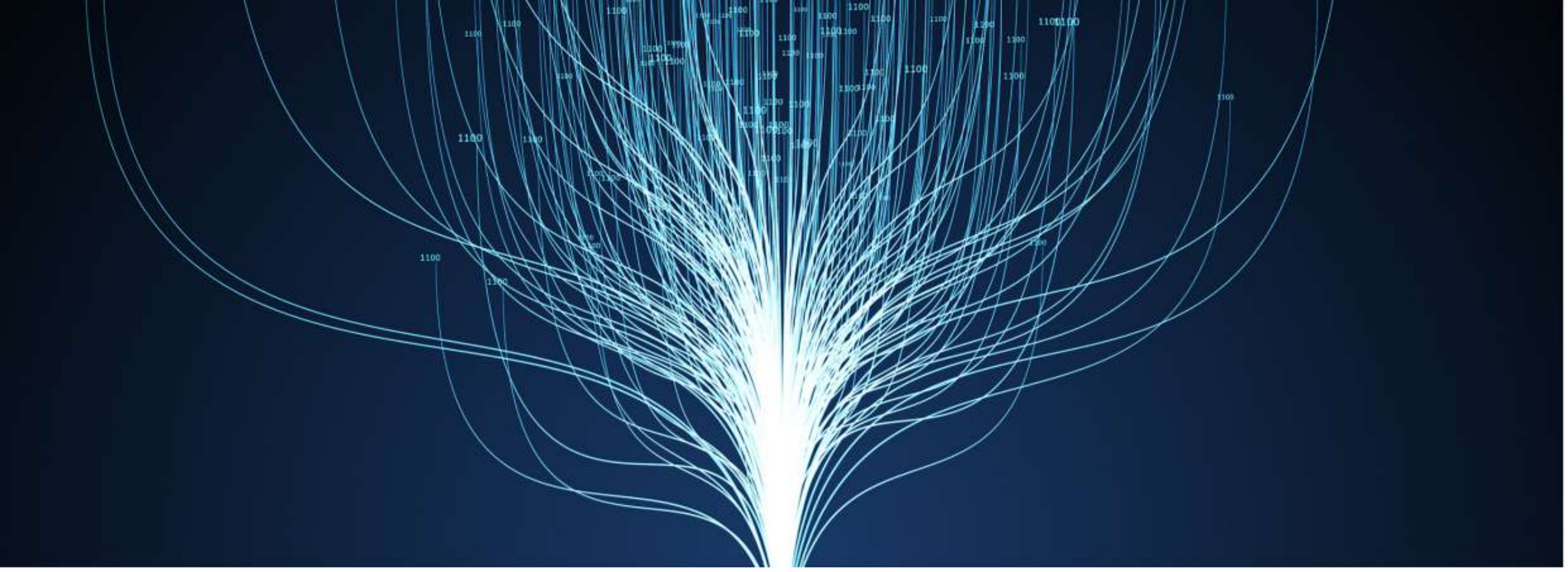
Broadly speaking, there are two types of AI: specific or weak AI and global or strong AI. Narrow artificial intelligence is created to carry out a certain activity or set of tasks, such as playing chess or spotting fraud in financial transactions. General AI, on the other hand, is capable of carrying out any intellectual work that a person can, and perhaps even much more. Although the creation of general AI is still a long way off for many academics in the field, narrow AI has already had a big impact on a lot of businesses.

LivAI sepie Funded by the Erasmus+ Programme of the European Union

1.2 A BRIEF HISTORY OF ARTIFICIAL INTELLIGENCE

Although attempts to create devices, or agents, endowed with artificial intelligence can be found in humanity's distant past, it was not until the twentieth century that advances in this field began to bear fruit. These fruits have developed rapidly over the last few decades.

The origins of AI can be traced back to the mid-20th century when the field was first established as a distinct discipline. The early pioneers of AI envisioned creating machines that could replicate human intelligence, and these ideas laid the foundation for the development of modern-day AI technologies.



1.2.1 EARLY HISTORY

The origins of AI can be traced back to the Dartmouth Conference of 1956, where a group of researchers came together to discuss the possibility of creating machines that could ”think” like humans. This conference marked the birth of AI as a distinct discipline, and it laid the foundation for the development of the field in the coming decades.

In the early years of AI research, the focus was on creating programs that could solve simple problems and perform basic tasks. One of the earliest successful AI programs was the Logic Theorist, developed by Allen Newell and J.C. Shaw in 1956. This program was capable of proving mathematical theorems, and it demonstrated that machines could be programmed to perform tasks that were traditionally considered to be the domain of human intelligence.

During the 1960s and 1970s, AI research continued to advance, and new breakthroughs were achieved in areas such as natural language processing, expert systems, and computer vision. In 1964, the first natural language processing program, called the ”STUDENT” was developed by Daniel Bobrow. This program was capable of understanding and responding to simple English sentences, and it laid the foundation for the development of more advanced natural language processing technologies in the years to come.

Another major breakthrough in the field of AI came in the form of expert systems, which were designed to replicate the decision-making abilities of human experts. One of the first successful expert systems was MYCIN, developed by Edward Shortliffe in 1974. This system was capable of diagnosing bacterial infections and recommending treatments, and it demonstrated the potential of AI to solve complex problems in medicine and other fields.

In the 1980s and 1990s, AI research shifted towards machine learning, which involves training machines to learn from data and make predictions or decisions based on that data. This approach has led to significant advances in areas such as computer vision, speech recognition, and natural language processing. One of the most important breakthroughs in machine learning was the development of deep learning algorithms, which are capable of learning complex representations of data and making accurate predictions.

Today, AI technologies are used in a wide range of applications, from self-driving cars and voice assistants to medical diagnosis and financial analysis. While AI has come a long way since its early days, there are still many challenges that need to be overcome before machines can truly replicate human intelligence. Nevertheless, the field continues to evolve at a rapid pace, and it is likely that we will see many more breakthroughs in the years to come.

1.2.2 PIONEERS IN THE FIELD OF AI

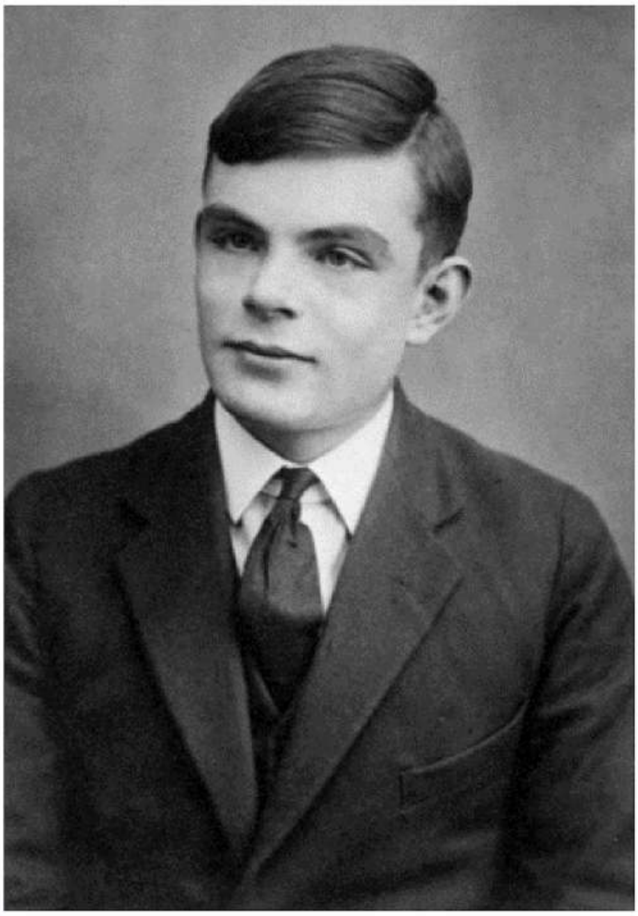


Figure 1 Alan Turing at age 16.

Alan Turing (see Figure 1) is a prominent figure in the history of computing and artificial intelligence. Born in 1912 in London, England, Turing is widely regarded as one of the most important figures in the development of modern computer science and AI.

Turing's most notable contribution to artificial intelligence was his work on the concept of a ”universal machine” or what is now known as the Turing machine.

This theoretical device laid the foundation for modern computing, providing a way to formalise the process of computation in a way that could be applied to any problem.

LivAI sepie
Funded by the
Erasmus+ Programme
of the European Union

In the early 1950s, Turing published a paper entitled ”Computing Machinery and Intelligence” in which he proposed what is now known as the Turing Test. The test involved a human judge communicating with both a machine and a human in a way that prevented them from knowing which was which. If the machine could successfully convince the judge that it was human, it would be considered to have passed the Turing Test and be considered intelligent.

Turing’s work on the Turing Test helped to spark interest in the concept of artificial intelligence and led to the development of early AI programs like ELIZA and SHRDLU. The Turing Test also remains a popular topic of discussion and research in the field of AI today.

However, Turing’s work on AI was cut short by his tragic death in 1954 at the age of 41. Despite his untimely passing, Turing’s legacy in the field of AI and computing lives on. Today, he is considered one of the founding fathers of AI and a hero of computer science.

In recognition of his contributions to the field, the Turing Award, often called the ”Nobel Prize of Computing,” was established in his honour in 1966. The award recognizes individuals who have made significant contributions to the field of computing and is considered one of the most prestigious honours in computer science.

Figure 2: John McCarthy at age 79.

John McCarthy (see Figure 2) is another prominent figure in the history of computing and artificial intelligence [3]. John McCarthy was an American computer scientist who made significant contributions to the field of AI. He was born on September 4, 1927, in Boston, Massachusetts, and passed away on October 24, 2011, in Stanford, California. One of McCarthy's major contributions to AI was the development of the Lisp programming language in the late 1950s.

Lisp was designed to be used for AI research, and it quickly became one of the most popular languages in the field.

McCarthy believed that programming languages should be designed to support the concepts and techniques of AI, and he saw Lisp as a step towards that goal. McCarthy is also known for coining the term ”artificial intelligence” in a 1956 conference at Dartmouth College. The conference, which was attended by McCarthy and other leading researchers in the field, is now considered the birthplace of AI. McCarthy and his colleagues hoped that AI would one day be able to solve complex problems that were too difficult for humans to solve on their own.

One of the areas that McCarthy focused on was the development of logic-based systems for AI. In the early 1960s, he created the first version of the programming language known as Prolog. Prolog is based on the principles of logic programming, which represents knowledge and reasoning about it. Prolog is still widely used today in areas such as natural language processing and expert systems.

LivAI sepie Funded by the Erasmus+ Programme of the European Union

McCarthy also contributed to the development of AI in other ways. He was one of the founders of the Stanford Artificial Intelligence Laboratory, which is still a leading research institution in the field. He also served as president of the American Association for Artificial Intelligence (AAAI) from 1982 to 1983 and received numerous awards for his contributions to the field, including the Turing Award in 1971.

Figure 3: Dr. Edward Feigenbaum.

Edward Feigenbaum (see Figure 3) is a renowned computer scientist who has contributed significantly to the development of AI. He was born in 1936 in Weehawken, New Jersey, and obtained his undergraduate degree in mathematics from the Carnegie Institute of Technology in 1956. He then received his Ph.D. in electrical engineering from Carnegie Mellon University in 1960.

Lisp was designed to be used for AI research, and it quickly became one of the most popular languages in the field.

LivAI sepie Funded by the Erasmus+ Programme of the European Union

Today, AI technologies are used in a wide range of applications, from self-driving cars and voice assistants to medical diagnosis and financial analysis. While AI has come a long way since its early days, there are still many challenges that need to be overcome before machines can truly replicate human intelligence. Nevertheless, the field continues to evolve at a rapid pace, and it is likely that we will see many more breakthroughs in the years to come.

In addition to his work on expert systems, Feigenbaum also contributed to the development of knowledge-based systems, which are computer programs that are capable of reasoning about complex problems by representing knowledge in a structured format.

Feigenbaum received many awards and honours throughout his career, including the National Medal of Science in 1994, which is the highest scientific honour awarded by the United States government. He was also inducted into the Computer History Museum Hall of Fellows in 1998.

LivAI sepie Funded by the Erasmus+ Programme of the European Union

Joshua Lederberg (1925-2008) (see Figure 4) was an American molecular biologist and geneticist who made significant contributions to AI in the 1950s and 1960s. Lederberg was particularly interested in how computers could be used to aid scientific discovery and was one of the pioneers of computational biology.

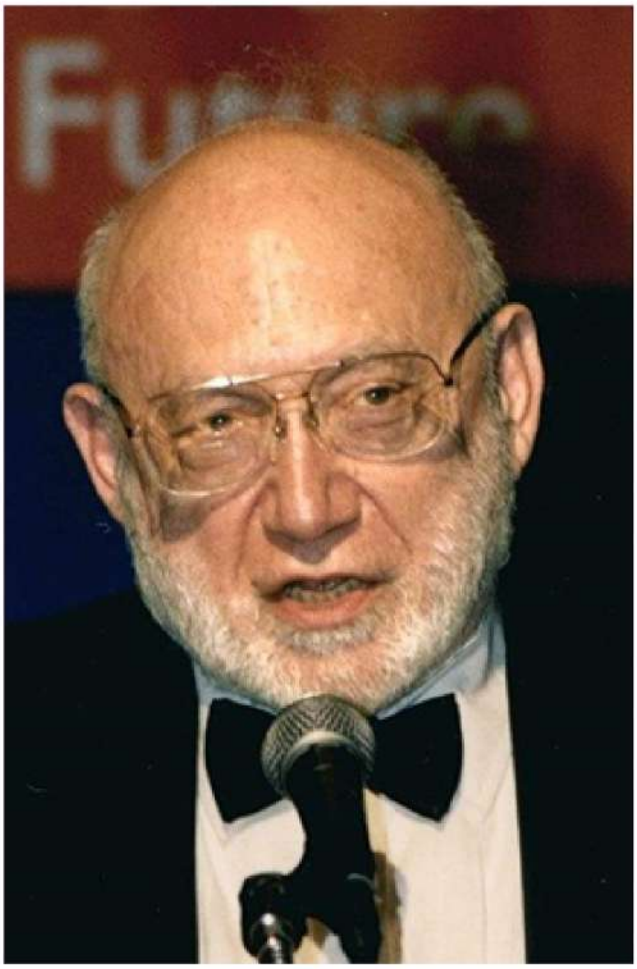


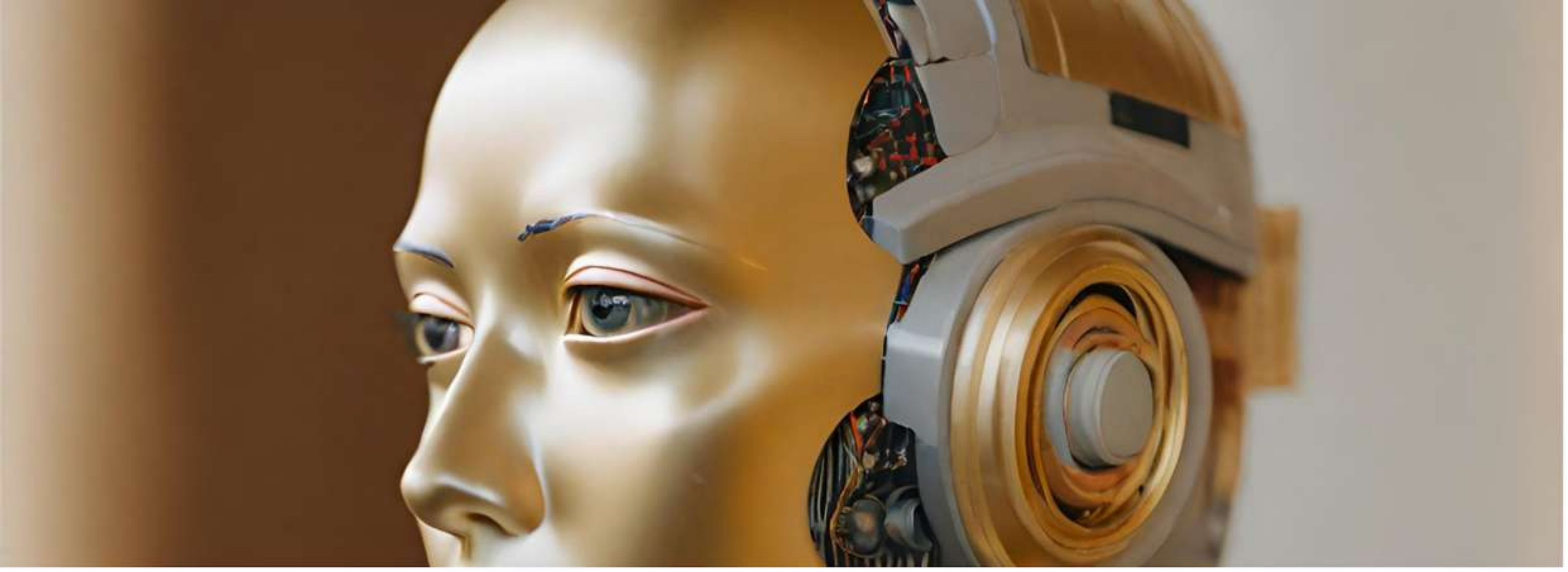
Figure 4: Joshua Lederberg.

Lederberg's work in AI focused on developing algorithms that could simulate the process of biological evolution. He was one of the first scientists to recognize the potential of genetic algorithms, which use principles of natural selection and evolution to solve complex problems.In 1959, Lederberg co-authored a seminal paper with Edward Feigenbaum, entitled "Simulation of Genetic Systems by Automatic Digital Computers," which described a computer program that could model genetic systems and predict the outcomes of genetic experiments. This work laid the foundation for the field of computational biology, which uses computers to study biological systems.

LivAI sepie Funded by the Erasmus+ Programme of the European Union

Lederberg also played a key role in the development of expert systems, which are computer programs that use knowledge and reasoning to solve problems in specialised domains. In the 1970s, Lederberg founded a company called BioComp Systems, which developed expert systems for a variety of applications, including medical diagnosis and drug discovery.

Lederberg's contributions to AI and computational biology have had a lasting impact on science and technology. His pioneering work helped establish the use of computers in biological research and paved the way for new approaches to scientific discovery.



1.2.3 GOLDEN AGE

The ”golden age” of AI is a period in the history of AI that occurred from the late 1950s to the early 1970s. During this time, researchers made significant advancements in the field of AI, particularly in the areas of problem-solving, natural language processing, and machine learning.

The term ”golden age” of AI was coined by Edward Feigenbaum, an American computer scientist, and AI pioneer, in the late 1980s. Feigenbaum used this term to refer to the period when AI research received a significant amount of funding and attention from government agencies, academic institutions, and private companies.

One of the key developments during the golden age of AI was the creation of expert systems. Expert systems are computer programs that can perform tasks that would typically require human expertise, such as diagnosing medical conditions or detecting faults in complex systems.

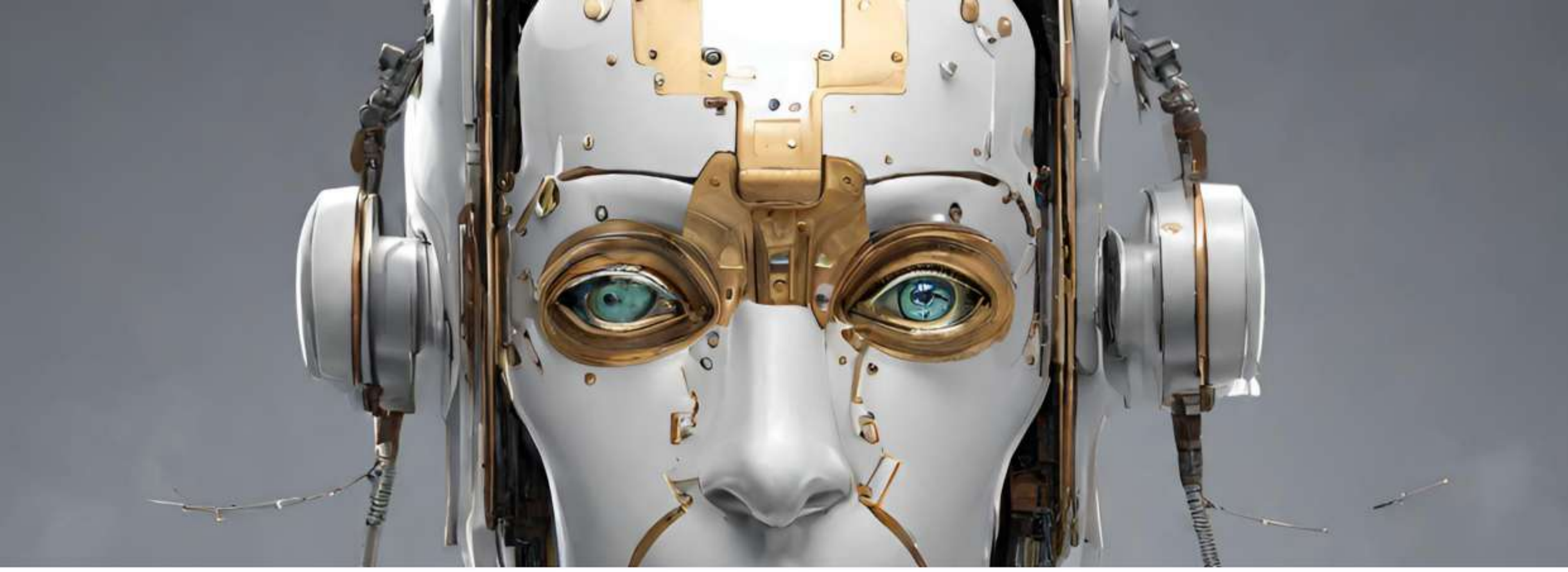
LivAI sepie Funded by the Erasmus+ Programme of the European Union

The first expert system, called Dendral, was developed in the early 1960s by Edward Feigenbaum and Joshua Lederberg. Dendral was used to analyse chemical compounds and was considered a significant breakthrough in AI research.

Another significant advancement during the golden age of AI was the development of natural language processing (NLP) techniques. During the golden age, researchers developed early NLP systems that could perform tasks such as machine translation and speech recognition.

Perhaps the most famous achievement of the golden age of AI was the creation of the first machine learning algorithms. Machine learning is a subfield of AI that focuses on creating algorithms that can learn and improve from data without being explicitly programmed. The first machine learning algorithm was developed by Arthur Samuel in 1959, and it was used to play checkers at a high level.

Despite the significant advancements made during the golden age of AI, the field experienced a decline in the 1970s due to a lack of progress in some areas and the failure of some early AI applications to live up to their hype. However, the lessons learned during this period laid the foundation for future AI research and continue to inform current developments in the field.



1.2.4 AI WINTER

The AI winter refers to a period of time during the 1970s and 1980s when funding and interest in AI research significantly decreased. This was due in part to several factors, including a lack of significant progress in the field, the inability to fulfil early promises of AI, and the perception that AI was not living up to its hype.

During this time, funding for AI research was significantly reduced, and many researchers left the field or shifted their focus to other areas. There was also a general sense that AI was not capable of fulfilling the lofty promises that had been made about it, such as creating truly intelligent machines.

However, despite the decline in funding and interest, AI research continued to advance during this period. Some important developments included the creation of expert systems, which could make decisions based on rules and knowledge, and the development of machine learning algorithms, which allowed machines to learn from data.

The AI winter eventually came to an end in the 1990s, as advances in computing power and the emergence of the internet created new opportunities for AI research. Today, AI is a rapidly growing field with applications in a wide range of industries, from healthcare and finance to transportation and entertainment.

1.2.5 RENAISSANCE

The 1990s marked the beginning of the renaissance of AI research, which was fueled by advances in computing power and machine learning techniques. During this time, researchers started to explore new approaches to AI, including neural networks, deep learning, and reinforcement learning.

Neural networks were first proposed in the 1940s, but it wasn't until the 1990s that they became a prominent tool in AI research. Neural networks are modelled after the structure of the human brain and consist of layers of interconnected nodes that can learn from input data. In the 1990s, researchers began to develop more advanced neural network architectures, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), which are still widely used today in applications such as image and speech recognition.

Deep learning is a subset of machine learning that uses neural networks with many layers, allowing for more complex and sophisticated modelling.

Deep learning has had a significant impact on AI research in recent years, particularly in the fields of computer vision and natural language processing. With deep learning, researchers have been able to develop AI systems that can accurately recognize images and speech, translate languages, and even play complex games like Go.

Reinforcement learning is a type of machine learning that involves training an agent to make decisions based on rewards or penalties received from the environment. In the 1990s, researchers began to explore reinforcement learning as a way to develop AI systems that could learn through trial and error. Today, reinforcement learning is used in a wide range of applications, including robotics, gaming, and autonomous vehicles.

The renaissance of AI research in the 1990s paved the way for many of the AI technologies we see today. Thanks to advances in computing power and machine learning techniques, researchers were able to explore new approaches to AI and develop more sophisticated AI systems. Neural networks, deep learning, and reinforcement learning have all had a significant impact on AI research and have opened up new possibilities for the future of AI.

1.2.6 RECENT DEVELOPMENTS

In recent years, there have been several significant developments in the field of AI, ranging from self-driving cars to facial recognition systems. These advancements are due to the growth of machine learning techniques and the availability of large amounts of data, as well as increased computing power.

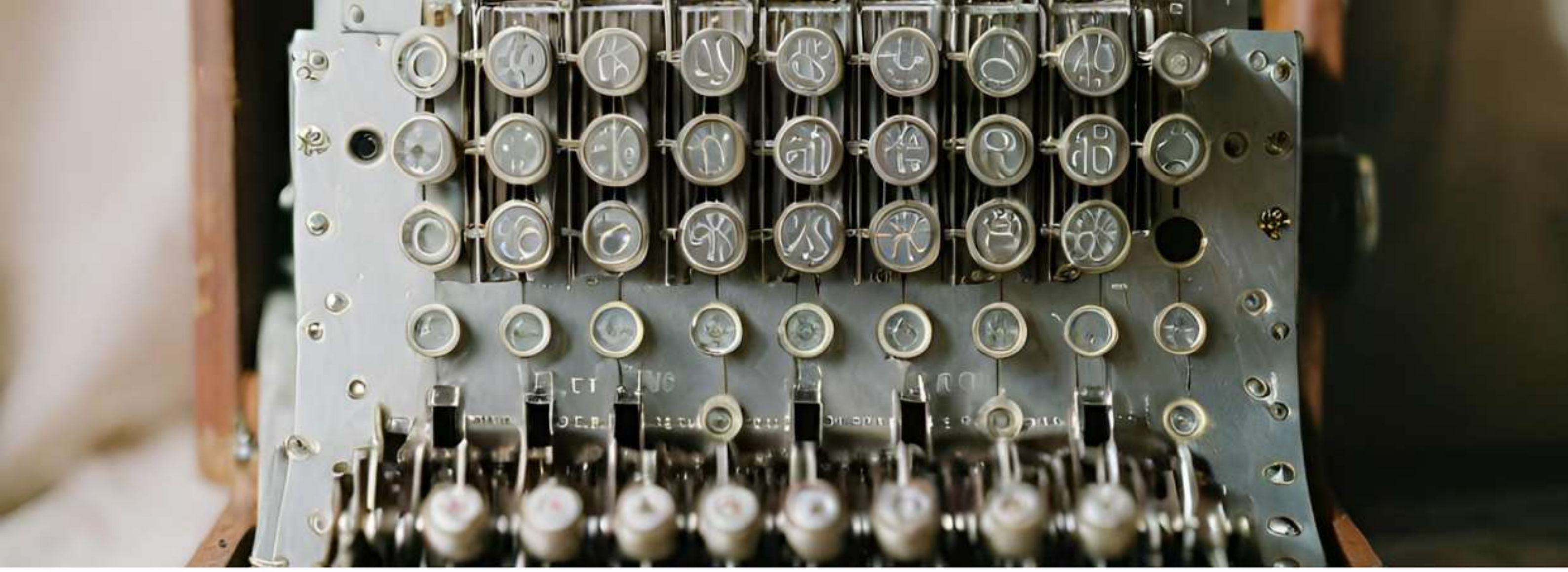
One of the most notable developments is in the area of autonomous vehicles. Self-driving cars use a combination of sensors, cameras, and machine-learning algorithms to navigate roads without human input. Companies such as Tesla, Google, and Uber have invested heavily in this technology, and there are already some self-driving cars on the road. However, there are still significant challenges to be overcome, such as ensuring safety and handling unexpected situations on the road.

Virtual assistants, such as Amazon’s Alexa and Apple’s Siri, are also becoming increasingly popular. These AI-powered assistants can understand spoken commands and perform a variety of tasks, from playing music to setting reminders.

As these systems become more advanced, they may be able to perform even more complex tasks, such as scheduling meetings and making purchases.

Facial recognition systems are another area of AI development that has garnered attention in recent years. These systems use deep learning algorithms to analyse images and identify individuals. While this technology has many potential applications, such as enhancing security measures and improving healthcare, there are also concerns about privacy and bias. For example, some studies have shown that these systems are less accurate when identifying people of certain races or genders.

As with any new technology, there are both challenges and opportunities presented by these developments in AI. Ethical concerns, such as privacy and bias, must be carefully considered to ensure that these systems are deployed in a responsible and fair manner. At the same time, these technologies have the potential to revolutionise a wide range of industries and improve our lives in countless ways.



1.2.7 SOME MAIN ACHIEVEMENTS IN THE HISTORY OF AI

THE ENIGMA MACHINE

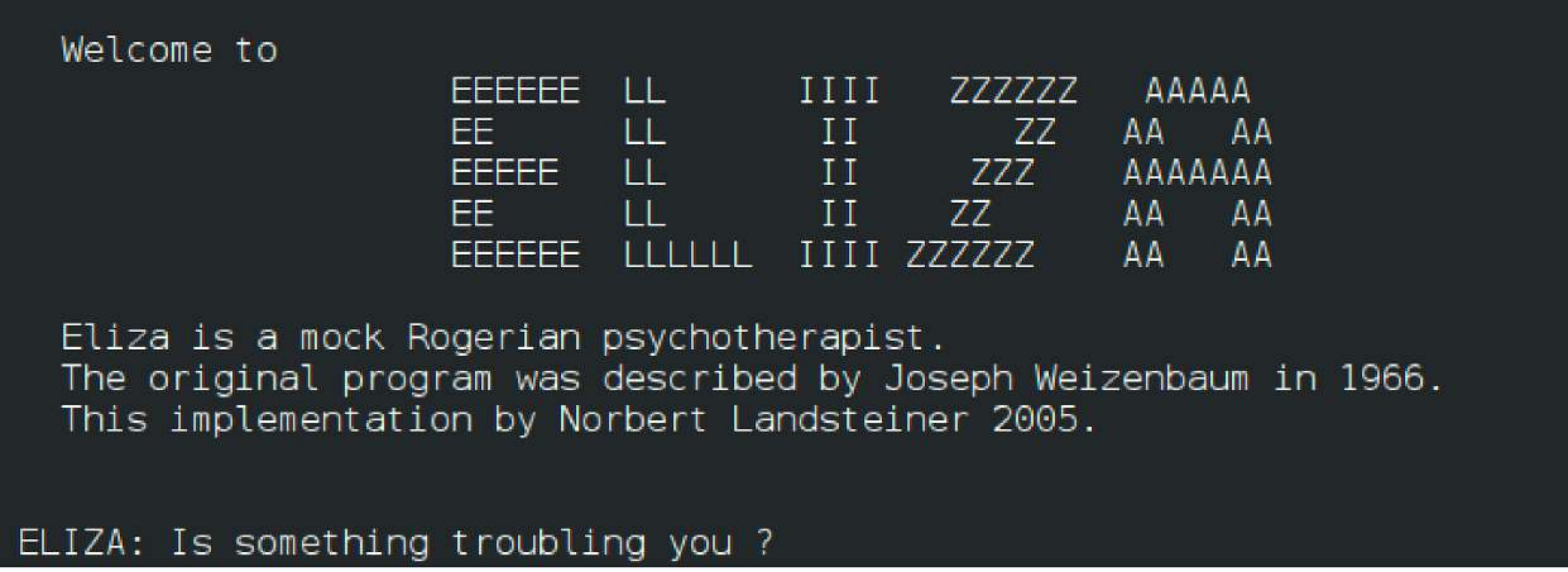
The Enigma machine was a type of encryption device used by the German military during World War II to send secret messages. The machine used a series of rotors to scramble the letters of a message, making it difficult to decipher without knowledge of the specific rotor settings.

During World War II, Turing worked at the Government Code and Cypher School at Bletchley Park in England, where he helped design and build a machine known as the Bombe. The Bombe was a type of electro-mechanical device that used mathematical algorithms to help decrypt messages sent using the Enigma machine.

Turing and his colleagues were able to break the Enigma code, which helped turn the tide of the war in favour of the Allies.

LivAI sepie Funded by the Erasmus+ Programme of the European Union

Turing's work on the Bombe and the breaking of the Enigma code are considered to be major achievements in the field of cryptanalysis, and helped lay the foundation for the development of modern computing and AI.



```
  Welcome to
                    EEEEEE  LL      IIII   ZZZZZZ   AAAAA
                    EE      LL       II       ZZ   AA   AA
                    EEEEE   LL       II     ZZZ    AAAAAAA
                    EE      LL       II    ZZ      AA   AA
                    EEEEEE  LLLLLL  IIII ZZZZZZ    AA   AA

  Eliza is a mock Rogerian psychotherapist.
  The original program was described by Joseph Weizenbaum in 1966.
  This implementation by Norbert Landsteiner 2005.

ELIZA: Is something troubling you ?
```

ELIZA
THE FIRST CHATBOT IN THE HISTORY OF AI

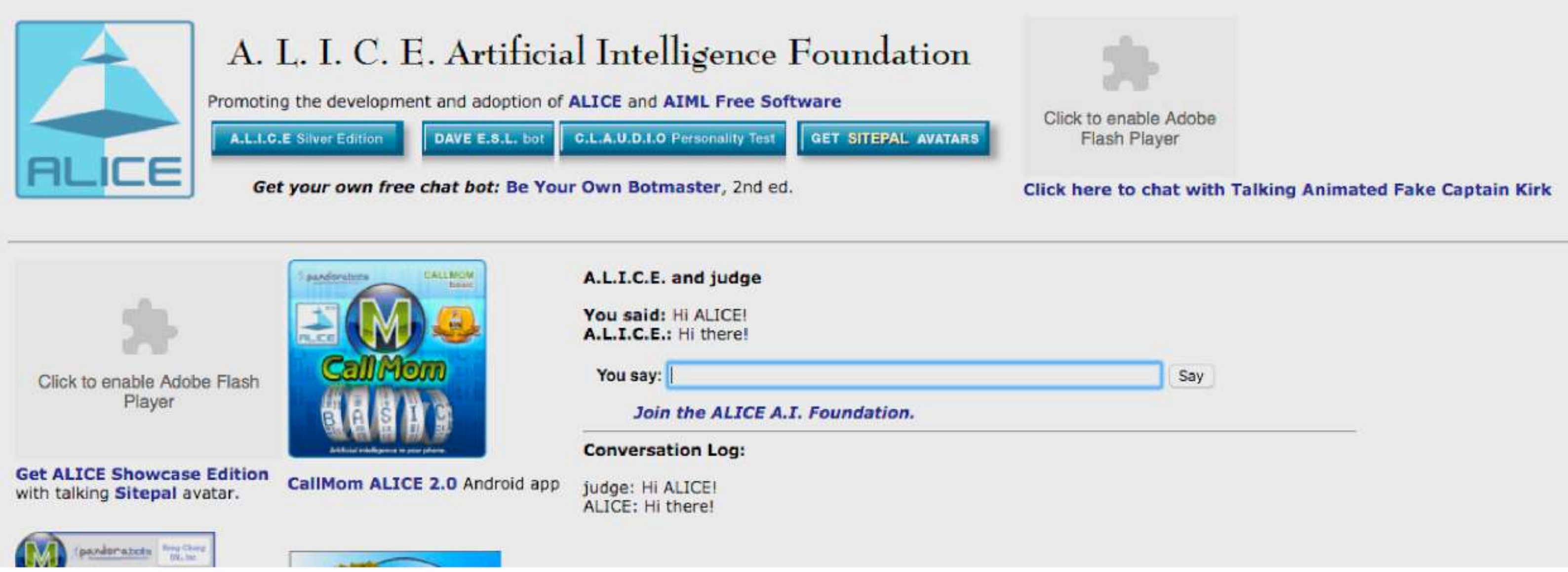
Eliza is one of the earliest examples of a chatbot, developed in the mid-1960s by Joseph Weizenbaum at the Massachusetts Institute of Technology (MIT). It was named after the character Eliza Doolittle in George Bernard Shaw's play Pygmalion.

Eliza was a computer program designed to simulate conversation with humans using natural language processing techniques. It was designed to imitate a psychotherapist and could ask simple questions and respond with pre-programmed replies based on the user's input. The program was based on pattern matching techniques, and it would scan the user's input for certain keywords or phrases and respond with a predetermined response.

Despite its simplicity, Eliza was a ground-breaking development in the field of AI and natural language processing.

LivAI sepie Funded by the Erasmus+ Programme of the European Union

It demonstrated the potential for machines to communicate with humans in a conversational style and paved the way for more advanced chatbots and virtual assistants that we see today.



A. L. I. C. E. Artificial Intelligence Foundation

Promoting the development and adoption of **ALICE** and **AIML Free Software**

A.L.I.C.E Silver Edition
DAVE E.S.L. bot
C.L.A.U.D.I.O Personality Test
GET SITEPAL AVATARS
ALICE
Click to enable Adobe Flash Player
Click to enable Adobe Flash Player
Call Mom
pandorabots

Get your own free chat bot: **Be Your Own Botmaster**, 2nd ed.

Click here to chat with Talking Animated Fake Captain Kirk

Get ALICE Showcase Edition with talking **Sitepal** avatar.

CallMom ALICE 2.0 Android app

A.L.I.C.E. and judge

You said: Hi ALICE!
A.L.I.C.E.: Hi there!

You say: Say

Join the ALICE A.I. Foundation.

Conversation Log:

judge: Hi ALICE!
ALICE: Hi there!

THE CHATBOT ALICE

ALICE (Artificial Linguistic Internet Computer Entity) is a chatbot program developed using AI techniques. It was created by Dr. Richard Wallace in the mid-1990s and was one of the earliest attempts to create a conversational agent that could hold a natural language conversation with a user.

ALICE uses a form of natural language processing to analyse a user's input and generate an appropriate response. The program is based on a rule-based system, which means that it follows a set of pre-defined rules and responds accordingly.

One of the key features of ALICE is its ability to generate responses that are contextually appropriate. It can remember past conversations and use that information to generate more personalised responses.

ALICE has been used in a variety of applications, including customer service and online chat support. It has also been used for research purposes to study human-

-computer interactions and natural language processing.

While ALICE was an early example of a chatbot developed using AI techniques, more advanced and sophisticated chatbots have been developed since then, such as Siri, Alexa, and Google Assistant, which use machine learning and natural language understanding to provide more accurate and personalised responses.

CHESS BATTLE BETWEEN KASPAROV AND DEEP BLUE

The 1997 match between world chess champion Garry Kasparov and the IBM computer Deep Blue is one of the most famous events in the history of AI. Kasparov was the reigning world champion and widely considered the greatest chess player of all time, while Deep Blue was a state-of-the-art computer system designed specifically to play chess.

The first match between Kasparov and Deep Blue took place in 1996, and although Kasparov won the match, it was a wake-up call for the chess world. The computer had shown that it was capable of beating some of the world's best players and was rapidly improving.

In the rematch in 1997, Kasparov was defeated by Deep Blue in a six-game match. It was the first time a computer had beaten a reigning world chess champion in a match. Kasparov accused IBM of cheating, claiming that the computer had received help from human operators during the game.

IBM denied the accusations, and an investigation by the International Chess Federation found no evidence of cheating.

LivAI sepie Funded by the Erasmus+ Programme of the European Union

The match was a major milestone in the development of AI, demonstrating that computers were capable of defeating even the best human players in complex strategic games. It also sparked renewed interest in the development of AI and led to further research and development in the field of machine learning and neural networks.

ALPHA-GO AND THE GO BATTLE

AlphaGo is a computer program developed by the Google DeepMind team that made history by defeating the world champion of the board game Go in a five-game match in 2016. This was considered a major milestone in the field of AI, as the game of Go has a vastly larger and more complex set of possible moves than chess or other popular board games, making it difficult for traditional AI approaches to master.

The achievement of AlphaGo was made possible by a combination of advanced techniques in machine learning, deep neural networks, and reinforcement learning. AlphaGo was trained on a large dataset of expert Go games, and learned to predict the moves of expert players using a neural network. It then improved its performance through a process of reinforcement learning, in which it played millions of games against itself and learned from the outcomes.

The victory of AlphaGo over world champion Lee Sedol was seen as a major breakthrough in AI and brought renewed interest and investment to the field.

Since then, the DeepMind team has continued to develop new and more advanced AI systems, including AlphaZero, which can learn to play multiple games at a superhuman level using only the rules of the game and without any prior human knowledge or data.

VIRTUAL ASSISTANTS

Alexa, Siri, and Google Assistant are all popular virtual assistants that use AI to understand and respond to user requests. They are often referred to as ”smart speakers” and can be used to control various devices in a smart home, provide information, play music, and perform various tasks.

Alexa is the virtual assistant developed by Amazon and used on the Amazon Echo and other devices. It can perform a wide range of tasks, such as playing music, setting alarms, making phone calls, and controlling smart home devices. Alexa is designed to learn from the user’s interactions and become more accurate and helpful over time.

Siri is the virtual assistant developed by Apple and used on iPhones, iPads, and other devices. It can perform a variety of tasks, such as setting reminders, making phone calls, sending text messages, and searching the internet. Siri also has the ability to learn and adapt to the user’s habits and preferences.

Google Assistant is the virtual assistant developed by Google and used on various devices, including Google Home and Android smartphones. It can perform similar tasks as Alexa and Siri, such as playing music, setting alarms, making phone calls, and controlling smart home devices. Google Assistant also has the ability to recognize and respond to voice commands in multiple languages.

One of the key achievements of these virtual assistants is their ability to understand and respond to natural language commands, allowing users to interact with them more easily and intuitively. They have also helped to popularise the idea of smart home devices and the Internet of Things, making it easier for people to control various aspects of their homes with their voices.

However, there are also concerns about privacy and security, as these devices are always listening and collecting data on users' interactions.

LivAI sepie
Funded by the
Erasmus+ Programme
of the European Union

1.2.8 FUTURE PROSPECTS

AI is an ever-evolving field, and there are many potential future developments. One possibility is the development of strong AI, which would have human-level intelligence and the ability to reason, learn, and understand as humans do. With strong AI, we could see many benefits such as improvements in healthcare, education, and science. However, there are also potential risks such as the displacement of jobs and the possibility of the AI becoming uncontrollable or even malicious.

Another potential future development is the creation of superintelligence, which would surpass human intelligence. This would require significant advancements in machine learning and cognitive computing, and could have profound effects on society. While the benefits of superintelligence could be immense, such as solving some of the world’s most pressing problems, there are also significant risks such as the potential loss of control over the AI.

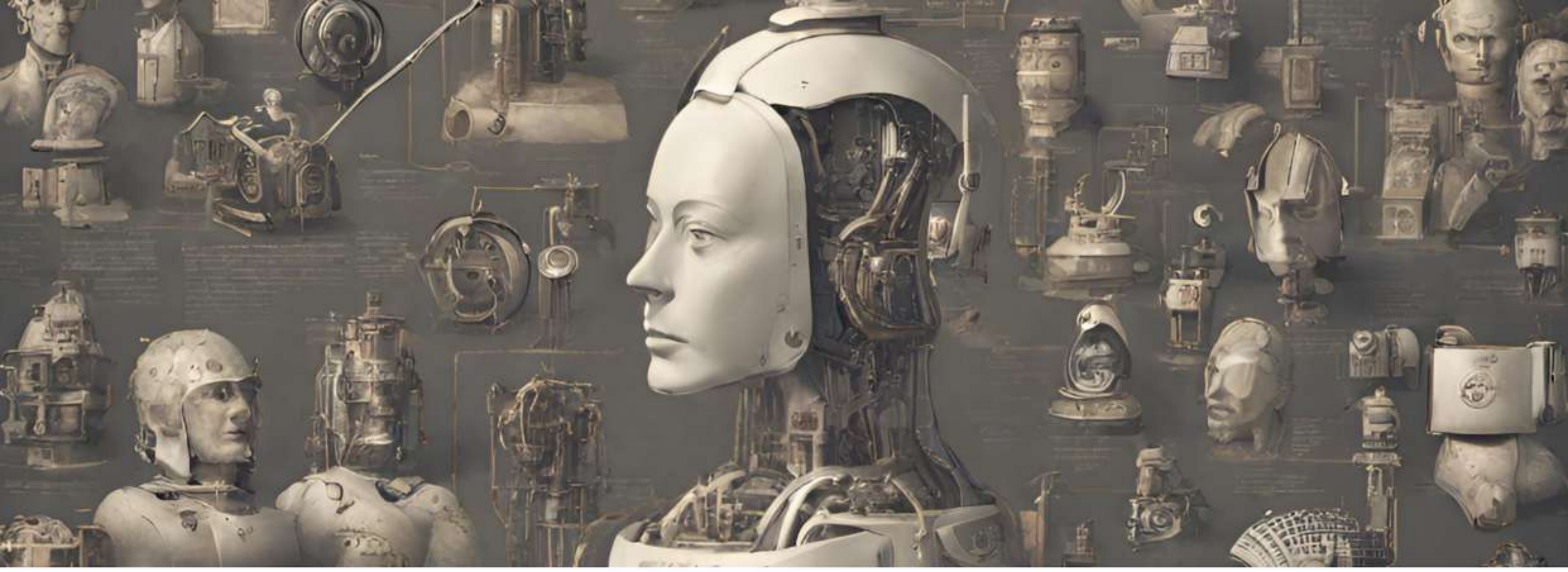
In addition to these concerns, there are also many exciting potential applications of AI, including the continued development of self-driving cars, virtual assistants, and facial recognition systems. These technologies have the potential to revolutionise transportation, communication, and security. However, they also raise important ethical concerns around privacy, surveillance, and bias.

As AI continues to evolve, it is important for society to stay informed and engaged with its development. This includes ongoing discussions about the ethical implications of AI, as well as continued investment in research and development to ensure that AI is used to benefit humanity. By working towards responsible AI development, we can harness the potential benefits of these technologies while minimising the risks.

LivAI sepie Funded by the Erasmus+ Programme of the European Union

1.3 ARTIFICIAL INTELLIGENCE FIELDS

The following is a list of the main AI fields, and some of its applications.



1.3.1 PLANNING AND SCHEDULING

Planning and scheduling are two related fields within AI that deal with the allocation of resources and activities over time to achieve a specific goal or objective.

Planning is the process of coming up with a series of deeds or acts that result in the intended result while taking into account a variety of restrictions, including time, money, and other resources that may be limited. Making a plan that accomplishes the target result in an ideal or nearly ideal way is the aim of planning.

Scheduling, on the other hand, involves the process of determining when and where to execute the actions or steps defined in the plan. The goal of scheduling is to allocate resources and activities in a way that minimises cost, maximises efficiency, and ensures that all deadlines are met.

Both planning and scheduling are critical components of many real-world applications, such as manufacturing, transportation, logistics, and project management. AI techniques, such as search algorithms, constraint satisfaction, optimization, and heuristics, are often used to solve planning and scheduling problems.

Some typical scenarios where planning and scheduling play an important role are as follows:

- **Logistics Planning** - In the logistics industry, planning and scheduling are used to optimise the routing of vehicles, the allocation of resources and personnel, and the scheduling of deliveries to minimise costs and ensure timely delivery. AI-based techniques, such as heuristic search algorithms and constraint programming, are used to generate optimal solutions to complex logistics problems.

- **Project Management** - In project management, planning and scheduling are used to allocate resources, assign tasks to team members, and track project progress. AI-based techniques, such as critical path analysis and resource levelling, are used to create project schedules that ensure that all tasks are completed on time and within budget.

- **Healthcare Planning** - In healthcare, planning and scheduling are used to optimise patient care and resource allocation. AI-based techniques, such as decision support systems and predictive modelling, are used to create patient treatment plans that ensure timely and effective care while minimising costs and resource utilisation.

1.3.2 NATURAL LANGUAGE PROCESSING (NLP)

Natural Language Processing (NLP) is a field of computer science and AI that focuses on making computers understand human language. NLP involves developing algorithms and models that can analyse and interpret text or speech data, and generate responses in a way that is natural and meaningful to humans. NLP enables machines to communicate with humans in a more human-like way, and has many applications, such as chatbots, virtual assistants, language translation, and text summarization.

SOME APPLICATIONS OF NLP INCLUDE:

- **Sentiment analysis**: NLP techniques are used to analyse the sentiment of social media posts, customer reviews, and other text data to determine whether the sentiment is positive, negative, or neutral.

- **Chatbots and virtual assistants**: NLP is used to create conversational agents, such as chatbots and virtual assistants, that can interact with users in natural language to answer questions, provide information, and perform tasks.

- **Machine translation**: NLP techniques are used to automatically translate text from one language to another, such as Google Translate.

- **Text summarization**: NLP is used to automatically summarise long documents or articles, extracting the most important information and presenting it in a concise form.

- **Named entity recognition**: NLP techniques are used to identify and extract named entities, such as people, places, and organisations, from text data.

- **Information extraction:** NLP is used to automatically extract structured information from unstructured text data, such as extracting product names, prices, and descriptions from e-commerce websites.

- **Speech recognition:** NLP techniques are used to transcribe spoken language into text, such as voice assistants like Siri and Alexa.

- **Text classification:** NLP is used to automatically categorise text data into predefined categories, such as spam detection in emails or classifying news articles by topic.

1.3.3 COMPUTER VISION

Computer vision focuses on enabling computers to interpret and understand visual data from the world around us. It involves developing algorithms and models that can analyse and interpret digital images and videos, recognizing patterns, objects, and relationships in the visual data.

SOME APPLICATIONS OF COMPUTER VISION INCLUDE:

- **Object recognition**: Computer vision techniques are used to identify and classify objects in images and videos, which is used in applications such as self-driving cars, robotics, and surveillance systems.

- **Facial recognition**: Computer vision is used to recognize faces in images and videos, which is used in security systems, social media, and entertainment.

- **Image and video analysis**: Computer vision is used to analyse and understand the content of images and videos, which is used in applications such as medical imaging, sports analytics, and content moderation.

- **Augmented reality**: Computer vision techniques are used to overlay digital information onto the real-world environment, creating immersive experiences in fields such as gaming, education, and advertising.

- **Autonomous vehicles:** Computer vision is a critical component in enabling autonomous vehicles to navigate and interact with the world around them.

LivAI sepie Funded by the Erasmus+ Programme of the European Union

- **Quality control and inspection**: Computer vision is used to inspect and analyse products and materials, detecting defects and ensuring quality control in manufacturing and production.

- **Robotics**: Computer vision is used to enable robots to perceive and interact with their environment, which is used in industrial automation, healthcare, and more.

Expert Systems in Artificial Intelligence

1.3.4 EXPERT SYSTEMS

An expert system uses knowledge and inference rules to solve problems and make decisions in a specific domain. It simulates the decision-making ability of a human expert in a particular field by incorporating the expert's knowledge and reasoning processes into a computer program.

Expert systems consist of a knowledge base that stores the expert's knowledge and a set of inference rules that allow the system to reason and make decisions based on that knowledge. The knowledge base is typically organised in a way that reflects the structure of the domain, with rules and facts arranged in a hierarchical or network-like structure. The inference engine uses the knowledge base to reason and make inferences based on the input provided by the user.

Expert systems are particularly useful in domains where there is a large amount of specialised knowledge that is difficult to acquire or where human experts are scarce or expensive.

LivAI sepie
Funded by the Erasmus+ Programme of the European Union

Expert systems can help automate decision-making processes, reduce errors and variability, and provide consistent and reliable advice to users.

SOME APPLICATIONS OF EXPERT SYSTEMS INCLUDE:

- **Medical diagnosis:** Expert systems can assist physicians in diagnosing diseases by providing a knowledge base of symptoms, medical history, and other factors that contribute to a diagnosis.

- **Financial planning:** Expert systems can help individuals and businesses plan their finances by providing advice on investments, taxes, and other financial decisions.

- **Quality control:** Expert systems can assist in quality control by analysing data and identifying potential problems or defects in manufacturing processes.

- **Engineering design:** Expert systems can help engineers design and optimise products by providing knowledge about materials, structures, and other factors that influence product performance.

- **Customer service:** Expert systems can provide automated customer service by answering questions, providing advice, and resolving issues.

1.3.5 FUZZY LOGIC

Fuzzy logic is a mathematical framework that deals with reasoning and decision-making under uncertainty and imprecision. Unlike classical logic, which is based on binary true/false values, fuzzy logic allows for partial truths or degrees of truth, where the truth value can range between 0 and 1.

Fuzzy logic is used in AI and control systems to model and reason about complex systems where the data is uncertain, ambiguous, or imprecise. It provides a powerful tool for handling imprecise and uncertain data by allowing the degree of membership of an element to a set to be a continuous value rather than just true or false.

SOME APPLICATIONS OF FUZZY LOGIC INCLUDE:

- **Control systems:** Fuzzy logic is used in control systems to model and control complex systems that have imprecise or uncertain data. Examples include traffic control systems, and robotics.

- **Natural language processing:** Fuzzy logic is used in natural language processing to handle the ambiguity and imprecision of human language.

- **Medical diagnosis:** Fuzzy logic is used in medical diagnosis to model and reason about the imprecise and uncertain nature of medical data.

- **Financial analysis:** Fuzzy logic is used in financial analysis to model and reason about imprecise financial data.

- **Risk analysis:** Fuzzy logic is used in risk analysis to model and reason about the uncertainty and imprecision of risk data.

- **Traffic control:** Fuzzy logic is used in traffic control systems to manage traffic flow based on real-time data.

LivAI sepie Funded by the Erasmus+ Programme of the European Union

1.3.6 GENETIC ALGORITHMS

Genetic algorithms are a class of optimization algorithms inspired by the process of natural selection in biology. They are a type of metaheuristic algorithm that uses a combination of selection, crossover, and mutation operators to search for the optimal solution to a problem.

In genetic algorithms, a population of candidate solutions is generated and evolved over a number of iterations. Each candidate solution is represented as a string of bits, which is called a chromosome or genotype. The fitness of each candidate solution is evaluated based on how well it solves the problem at hand. The fittest solutions are selected to be used as parents for the next generation, and they undergo genetic operations like crossover and mutation to create a new population of candidate solutions.

The process of selection, crossover, and mutation continues until a stopping condition is met, such as a maximum number of iterations or a desired level of fitness.

The final solution is then the fittest candidate solution from the final population.

The process of selection, crossover, and mutation continues until a stopping condition is met, such as a maximum number of iterations or a desired level of fitness. The final solution is then the fittest candidate solution from the final population.

SOME APPLICATIONS OF GENETIC ALGORITHMS INCLUDE:

OPTIMIZATION PROBLEMS

Genetic algorithms are often used to find the best solution for optimization problems, such as the travelling salesman problem, and the knapsack problem.

DRUG DESIGN

Genetic algorithms can be used to design new drugs by optimising the molecular structure of candidate compounds to maximise their effectiveness while minimising side effects.

ROBOTICS

Genetic algorithms can be used to optimise the control parameters of robots, such as the movement and manipulation of robotic arms.

GAME THEORY

Genetic algorithms can be used to simulate and optimise strategies in game theory problems, such as the prisoner's dilemma or the iterated prisoner's dilemma.

1.3.7 EVOLUTIONARY COMPUTATION

Evolutionary computation is inspired by biological evolution and natural selection. It is a family of algorithms that use iterative optimization techniques to solve complex problems by simulating the natural process of evolution.

Evolutionary computation algorithms typically begin with a population of candidate solutions to a problem, each represented by a set of parameters or variables. The algorithm then applies a selection process, in which the fittest or most successful candidates are chosen to "reproduce" and create new candidate solutions.

The new candidate solutions are created through random mutation or recombination of the parameters from the selected individuals. The process of selection and reproduction is repeated over multiple generations, with each generation resulting in a new population of candidate solutions.

Over time, the population evolves and becomes more diverse, and the algorithm converges towards a set of optimal solutions to the problem.

Evolutionary computation algorithms can be used for a wide range of optimization problems, such as parameter optimization, feature selection, and data clustering

ENGINEERING DESIGN OPTIMIZATION

Evolutionary Computation can be used to optimise the design of complex engineering systems, such as aircraft, automotive engines, and civil structures. By iteratively selecting and breeding candidate designs, Evolutionary Computation algorithms can efficiently search large design spaces to find optimal or near-optimal solutions.

FINANCIAL PORTFOLIO OPTIMIZATION

Evolutionary Computation can be used to optimise investment portfolios by selecting the best mix of assets and their corresponding investment weights. Evolutionary Computation algorithms can take into account various financial risk factors, such as volatility and correlation, to optimise portfolio performance.

IMAGE AND SIGNAL PROCESSING

Evolutionary Computation can be used to optimise image and signal processing algorithms, such as image segmentation, feature extraction, and classification. By evolving the parameters of the algorithms, Evolutionary Computation can optimise their performance for specific image or signal processing tasks.

GAME PLAYING

Evolutionary Computation can be used to evolve strategies for playing complex games, such as chess, poker, and Go. By evolving the parameters of the game-playing agents, Evolutionary Computation algorithms can create intelligent agents that can compete against human players or other agents.

ROBOTICS

Evolutionary Computation can be used to optimise the behaviour of autonomous robots in dynamic and uncertain environments. By evolving the control parameters of the robots, EC algorithms can create robots that can adapt to changing environmental conditions and achieve complex tasks.

1.3.8 SWARM INTELLIGENCE

Swarm intelligence is a subfield of AI that is inspired by the collective behaviour of social animals, such as ants, bees, and birds. It is a type of distributed problem-solving approach that involves a large number of relatively simple agents, called ”swarm members”, that work together to solve a complex problem.

In swarm intelligence, individual members of the swarm are capable of interacting with each other and the environment in a decentralised and self-organised manner. The agents typically have limited capabilities and intelligence, but by coordinating their actions and exchanging information, they are able to collectively solve complex problems that would be difficult for any single agent to solve alone.

Swarm intelligence algorithms often use principles of evolutionary computation, such as natural selection and genetic algorithms, as well as principles of machine learning, such as reinforcement learning and neural networks. The algorithms can be applied to a wide range of problem domains, such as optimization, classification, clustering, and control.

SOME APPLICATIONS OF SWARM INTELLIGENCE INCLUDE:

ROUTING OPTIMIZATION

Swarm intelligence algorithms, such as Ant Colony Optimization, have been used to optimise routing in computer networks, transportation networks, and logistics.

RESOURCE ALLOCATION

Swarm intelligence has been used to allocate resources in complex systems, such as energy grids, water distribution networks, and cloud computing systems.

IMAGE AND SIGNAL PROCESSING

Swarm intelligence has been applied to image and signal processing tasks, such as feature selection, image segmentation, and signal classification.

ROBOTICS

Swarm intelligence has been used to control the behaviour of swarm robots, which are groups of autonomous robots that work together to achieve a common goal.

BIOINFORMATICS

Swarm intelligence has been used in bioinformatics to solve problems such as protein folding, gene expression analysis, and DNA sequence alignment.

FINANCE

Swarm intelligence algorithms have been applied to financial forecasting and prediction, stock market analysis, and portfolio optimization.

SECURITY

Swarm intelligence has been used to detect and prevent network intrusions and cyber attacks, as well as to optimise the placement of security sensors in critical infrastructure.

1.3.9 COGNITIVE COMPUTING

Cognitive computing is an interdisciplinary field of study that combines principles from AI, computer science, neuroscience, and cognitive psychology to create computer systems that can mimic or augment human cognitive abilities.

Cognitive computing systems are designed to learn, reason, and interact with humans in natural ways, using techniques such as machine learning, natural language processing, computer vision, and decision-making algorithms. These systems are able to analyse and interpret complex data and provide insights and recommendations based on that analysis.

LivAI sepie Funded by the Erasmus+ Programme of the European Union

The goal of cognitive computing is to create intelligent systems that can work collaboratively with humans, augmenting our cognitive abilities and enabling us to make better decisions, solve complex problems, and improve our quality of life. Some examples of applications of cognitive computing include personalised medicine, fraud detection, customer service, and autonomous vehicles.

SOME APPLICATIONS OF COGNITIVE COMPUTING INCLUDE:

HEALTHCARE

Cognitive computing can be used to assist with medical diagnoses, personalised treatment plans, and drug discovery. For example, IBM Watson Health is a cognitive computing system that can analyse patient data and suggest treatment plans based on the patient's individual medical history.

FINANCE

Cognitive computing can be used for fraud detection and risk management in the finance industry. For example, the software company Ayasdi uses cognitive computing to detect fraud in financial transactions by analysing patterns in transaction data.

AUTONOMOUS VEHICLES

Cognitive computing can be used to enable self-driving vehicles to perceive and navigate their environment, detect obstacles, and make decisions in real-time. For example, the cognitive computing system developed by Nvidia can enable autonomous vehicles to interpret complex data from sensors and cameras to make real-time decisions while driving.ms, such as the prisoner's dilemma or the iterated prisoner's dilemma.

CUSTOMER SERVICE

Cognitive computing can be used to improve customer service by providing personalised recommendations and resolving customer issues more efficiently. For example, the virtual assistant system Amelia can interact with customers in natural language and provide assistance with their inquiries or problems.

EDUCATION

Cognitive computing can be used to create personalised learning experiences for students, based on their individual learning needs and preferences. For example, Carnegie Learning uses cognitive computing to provide students with personalised feedback and support in their maths studies.

LivAI sepie Funded by the Erasmus+ Programme of the European Union

1.3.10 DECISION SUPPORT SYSTEMS

A Decision Support System is an interactive computer-based information system that supports decision-making activities within an organisation.

A Decision Support System provides users with tools and techniques to help analyse complex problems and make better decisions. It uses a variety of data sources, models, and algorithms to provide insights and recommendations to users, and allows them to interact with the system through a user-friendly interface.

Decision Support Systems are designed to support decision-making activities across various functional areas of an organisation, such as finance, marketing, operations, and human resources. They can be used for a wide range of decision-making activities, such as forecasting, planning, budgeting, and risk analysis.

The main goal of a Decision Support System is to help users make better decisions by providing them with relevant information and tools to analyse and interpret that information. It is not intended to replace human decision-making, but rather to augment it by providing additional support and guidance.

SOME APPLICATIONS OF DECISION SUPPORT SYSTEMS INCLUDE:

HEALTHCARE

Decision Support Systems can be used to support medical diagnosis and treatment planning, as well as patient care management. For example, a DSS can help doctors to identify the best treatment options for patients based on their medical history, symptoms, and lab test results.

FINANCIAL MANAGEMENT

Decision Support Systems can be used for financial planning, budgeting, and forecasting. For example, a Decision Support System can help a financial manager to analyse financial data and make informed decisions on investment opportunities, cash flow management, and financial risk mitigation.

SUPPLY CHAIN MANAGEMENT

Decision Support Systems can be used to optimise supply chain operations, such as inventory management, production planning, and logistics management. For example, a Decision Support System can help a supply chain manager to identify the most cost-effective transportation routes or optimise inventory levels.

MARKETING

Decision Support Systems can be used to support marketing decision-making activities, such as product positioning, target market selection, and pricing strategies. For example, a DSS can help a marketing manager to analyse customer data and identify patterns or trends in their behaviour and preferences.

HUMAN RESOURCES

Decision Support Systems can be used to support human resources decision-making activities, such as recruitment, training, and performance management. For example, a DSS can help a human resources manager to analyse employee data and identify the most suitable candidates for a job position or identify areas for employee training and development.

LivAI sepie
Funded by the Erasmus+ Programme of the European Union

1.3.11 KNOWLEDGE REPRESENTATION AND REASONING

Knowledge Representation and Reasoning is a subfield of AI that deals with the representation of knowledge in a form that can be processed by a computer and the development of algorithms that can reason with this knowledge to solve complex problems.

The goal of Knowledge Representation and Reasoning is to develop algorithms and techniques that can effectively capture and utilise the vast amount of knowledge that exists in various domains and make this knowledge accessible and usable to humans and machines alike.

In Knowledge Representation and Reasoning, knowledge is represented in a structured format, such as a graph or a set of rules, that allows a computer to understand the relationships and dependencies between different pieces of information. This structured representation of knowledge can then be used to reason about new information, make predictions, and solve complex problems.

Knowledge Representation and Reasoning techniques often involve the use of logic, which provides a formal language for expressing and manipulating knowledge. Some of the commonly used logic in KRR include propositional logic, first-order logic, and modal logic.

SOME APPLICATIONS OF KNOWLEDGE REPRESENTATION AND REASONING INCLUDE:

EXPERT SYSTEMS

Knowledge Representation and Reasoning techniques are used to represent and reason about expert knowledge in various domains, such as medicine, finance, and law. Expert systems use this knowledge to provide advice, diagnosis, and decision-making support.

NATURAL LANGUAGE PROCESSING

Knowledge Representation and Reasoning techniques are used to represent and reason about the meaning of natural language text. This allows computers to understand the relationships between words and concepts and to generate meaningful responses to queries and commands.

ROBOTICS

Knowledge Representation and Reasoning techniques are used to represent and reason about the environment and the robot's actions. This allows robots to plan and execute complex tasks, such as navigating through a cluttered environment or assembling a complex structure.

COGNITIVE COMPUTING

Knowledge Representation and Reasoning techniques are used to represent and reason about human cognition and behaviour. This allows cognitive computing systems to understand and mimic human reasoning and decision-making.

DECISION SUPPORT SYSTEMS

Knowledge Representation and Reasoning techniques are used to represent and reason about the rules and constraints that govern a particular decision-making domain. This allows decision support systems to provide recommendations and advice to decision-makers.

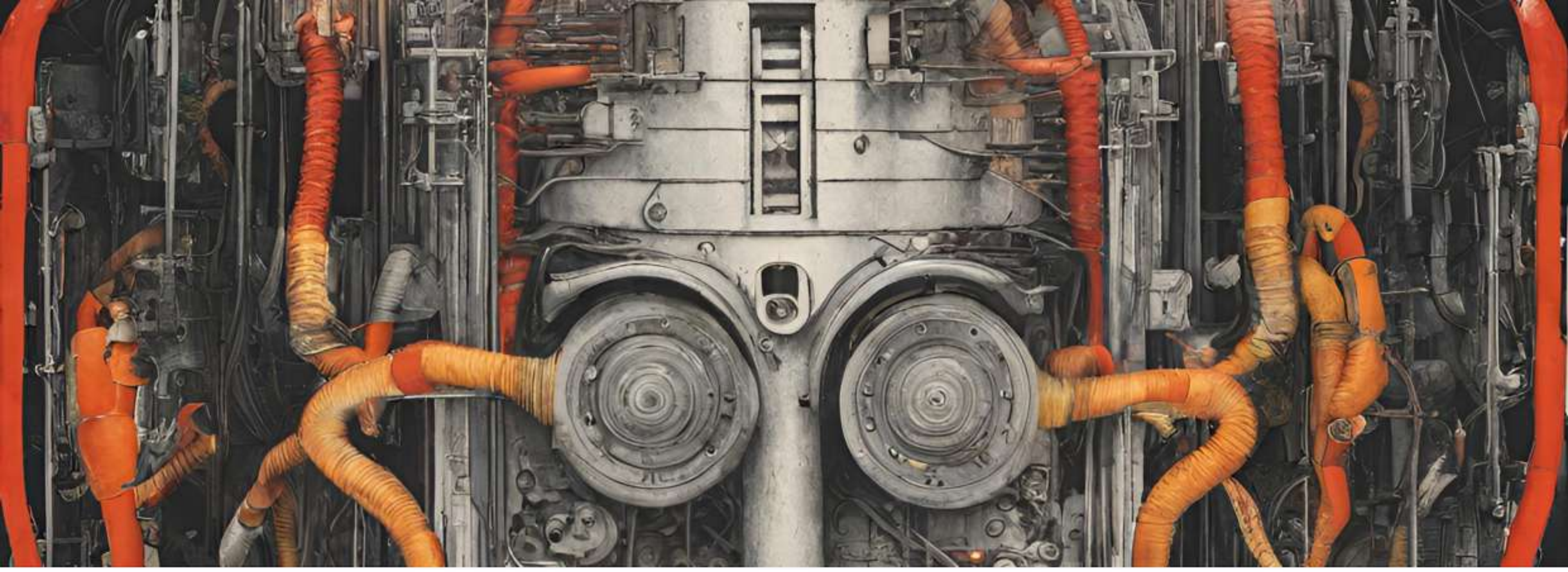
INTELLIGENT TUTORING SYSTEMS

Knowledge Representation and Reasoning techniques are used to represent and reason about the student's knowledge and learning goals. This allows intelligent tutoring systems to provide personalised feedback and guidance to students.

COGNITIVE COMPUTING

Knowledge Representation and Reasoning techniques are used to represent and reason about human cognition and behaviour. This allows cognitive computing systems to understand and mimic human reasoning and decision-making.

Funded by the Erasmus+ Programme of the European Union



1.3.12 MACHINE PERCEPTION

Machine perception is the ability of computers to use various sensors and data processing techniques to gather and interpret information from the physical world. This can involve analysing images, sounds, and other types of data to recognize patterns, identify objects, and make decisions based on the information gathered.

The goal of machine perception is to create computer systems that can perceive and understand the world in a similar way to humans, allowing them to interact with their environment and perform tasks that require sensory input and interpretation.

SOME APPLICATIONS OF MACHINE PERCEPTION INCLUDE:

COMPUTER VISION

Machine Perception is used in computer vision systems to interpret visual data from images and videos, enabling machines to recognize and identify objects, people, and scenes.

SPEECH RECOGNITION

Machine Perception is also used in speech recognition systems to interpret spoken language, enabling machines to understand and respond to human voice commands.

AUTONOMOUS VEHICLES

Machine Perception is a key component of autonomous vehicle systems, allowing vehicles to perceive their environment and make decisions about driving.

ROBOTICS

Machine Perception is also used in speech recognition systems to interpret spoken language, enabling machines to understand and respond to human voice commands.

MEDICAL DIAGNOSIS

Machine Perception is used in medical diagnosis systems to analyse medical images, such as X-rays and MRI scans, to help diagnose diseases and conditions.

LivAI sepie

Funded by the
Erasmus+ Programme
of the European Union

1.4 ARTIFICIAL INTELLIGENCE APPLICATIONS

Speech recognition, also known as automatic speech recognition (ASR), computer speech recognition, or speech-to-text, is a capability that converts spoken English into written language using natural language processing (NLP). Many mobile devices have speech detection built into their operating systems to enable voice search (like Siri) and to increase texting accessibility.

Online virtual agents are replacing human agents in customer support throughout the customer journey. They provide individualised advice, respond to frequently asked questions (FAQs) about subjects like shipping, cross-sell goods, or make size recommendations to users, altering the way we view user interaction on websites and social media. Examples include virtual agent-equipped messaging bots on e-commerce websites, chat programs like Slack and Facebook Messenger, and duties typically carried out by virtual assistants and voice assistants.

Through the use of digital images, videos, and other visual inputs, computer vision technology allows computers and systems to extract meaningful information from those inputs and take appropriate action. It differs from image recognition jobs in that it can make recommendations.

Computer vision uses convolutional neural networks to power picture tagging in social media.

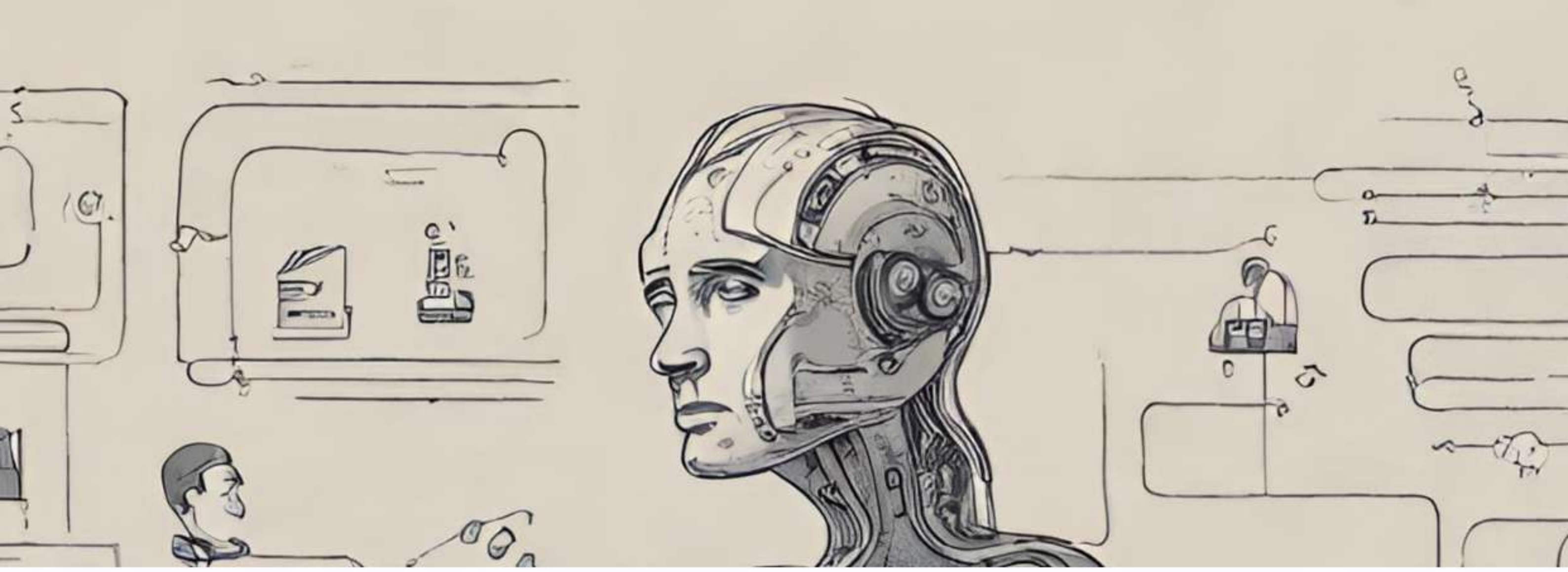
Recommendation engines: By using historical data on consumer behaviour, AI algorithms can help identify data patterns that can be applied to create more successful cross-selling tactics. Online merchants use this to suggest pertinent add-ons to customers during the checkout process.

Automated stock trading: Created to maximise stock portfolios, high-frequency trading platforms powered by AI execute thousands or even millions of transactions every day without the need for a human trader.

1.5 MACHINE LEARNING

Machine learning, which entails teaching computers on massive datasets to recognize patterns and make predictions based on those datasets, is one of the main techniques used to develop AI.

To achieve even higher levels of performance, deep learning, a subset of machine learning, employs artificial neural networks that are modelled after the human brain. Natural language processing, which enables machines to comprehend and react to human language, and computer vision, which enables machines to interpret visual data, are other techniques used in the development of AI.



1.5.1 SUPERVISED MACHINE LEARNING

Supervised machine learning is a type of machine learning where an algorithm is trained on a labelled dataset. The labelled dataset contains input data and corresponding output labels, which are provided to the algorithm during training. The algorithm then learns to map the input data to the output labels based on the patterns and relationships it identifies in the training data. The goal of supervised learning is to create a predictive model that can accurately predict the output labels for new, unseen input data.

The accuracy of the results for a supervised machine learning algorithm depends on the quality and quantity of the input data. The algorithm then uses this labelled dataset to learn patterns and relationships between the input data and the corresponding output.

LivAI sepie Funded by the Erasmus+ Programme of the European Union

If the input data is of high quality, meaning it is representative of the real-world scenario the algorithm is meant to address, and has sufficient quantity, the accuracy of the results is likely to be higher. On the other hand, if the input data is biassed, incomplete, or insufficient, the accuracy of the results may be lower.

It is important to note that the accuracy of the results also depends on the complexity of the problem being solved and the choice of algorithm used. Some problems may be inherently difficult to solve, even with high-quality input data and sophisticated algorithms, while others may be relatively easy to solve. Therefore, it is crucial to carefully consider the problem and choose an appropriate algorithm for the given task.

Some issues, related to the quantity and quality of the training labelled data set, are:

OVERFITTING When the model is too complex or the training data is noisy, the model may memorise the training data instead of learning the underlying patterns. This can result in poor performance on new, unseen data.

UNDERFITTING When the model is too simple or the training data is too sparse, the model may not capture the underlying patterns in the data. This can result in poor performance on both the training and test data.

IMBALANCED DATA

When the training data contains a disproportionate number of examples from one class, the model may perform poorly in the underrepresented class.

MISSING DATA

When the input data contains missing values, it can be difficult to impute the missing values and train a model that accurately captures the underlying patterns.

OUTLIERS

When the input data contains outliers or extreme values, they can skew the model's predictions and reduce its accuracy.

NOISE

When the input data contains random or irrelevant features, the model may learn spurious correlations that do not generalise to new data.

BIAS

When the training data is biassed towards a certain group or population, the resulting model may be biassed towards that group as well.

Addressing these problems requires careful data pre-processing, feature engineering, and model selection. It is also important to have a large and diverse dataset that captures the underlying patterns in the data.

SOME EXAMPLES OF APPLICATIONS OF SUPERVISED MACHINE LEARNING ARE:

- **Image Recognition**: Supervised machine learning algorithms are used to classify images in applications such as object detection, face recognition, and handwritten character recognition.The labelled training data consists of images of handwritten characters or words, and the label is the corresponding word.

- **Spam Filtering**: Spam filtering systems use supervised machine learning to identify and filter out unwanted emails. The training dataset consists of a set of emails, each labelled as ”spam” or ”ham”.

- **Medical Diagnosis**: Machine learning algorithms are used in medical diagnosis to help identify diseases based on patient data such as symptoms, medical history, and lab results. For detecting cancer in medical images such as mammograms, X-rays, and MRI scans, the algorithm is trained to classify images as either malignant or benign based on labelled data, being these labels as ”contains cancer” or ”not contains cancer”, for example.

LivAI sepie Funded by the Erasmus+ Programme of the European Union

- **Fraud Detection:** Supervised machine learning algorithms are used to identify fraudulent transactions in the banking, insurance, and e-commerce industries. For example, the labelled training data consists of a set of banking transactions labelled as ”safe” or ”fraudulent”.

- **Natural Language Processing**: Supervised machine learning algorithms are used to process and analyse large amounts of natural language data, such as text, speech, and audio, for applications such as sentiment analysis, language translation, and speech recognition. For example, in the case of sentiment analysis applied to restaurant user reviews, the labelled training data consists of each of the texts of the review, and the associated label is whether the review is positive or negative.

- **Autonomous Driving**: Supervised machine learning algorithms are used in self-driving cars to identify objects, such as pedestrians, traffic signals, and other vehicles, and to make decisions based on that information.

- **Customer Churn Prediction:** Supervised machine learning algorithms are used to predict customer churn in industries such as telecommunications, banking, and e-commerce. The labelled training dataset is based on the customer’s history, plus a tag indicating whether the customer dropped out or continued to pay for the service offered.

1.5.2 UNSUPERVISED MACHINE LEARNING

Unsupervised machine learning is a type of machine learning where the model is trained on unlabeled data without any specific supervision. In this type of machine learning, the algorithm tries to identify patterns, relationships, and structures in the data by itself without the need for any external guidance or labels. The goal of unsupervised learning is to discover the hidden structures and relationships in the data to make sense of it and find insights that can be useful for solving problems.

SOME EXAMPLES OF APPLICATION OF UNSUPERVISED MACHINE LEARNING ARE:

CLUSTERING

This involves grouping similar data points together into clusters. An example would be clustering customer data into groups based on their shopping habits.

ANOMALY DETECTION

This involves identifying data points that are significantly different from the rest of the data. An example would be detecting fraudulent transactions in financial data.

DIMENSIONALITY REDUCTION

This involves reducing the number of features in a dataset while preserving its essential characteristics. An example would be reducing the number of variables in a stock market dataset to identify the most important factors.

RECOMMENDER SYSTEMS

This involves algorithms used to suggest or recommend products or services to users based on their past preferences, behaviours, and interactions with the system. These systems are widely used in e-commerce, social media, and other online platforms to help users discover new items and enhance their overall experience.

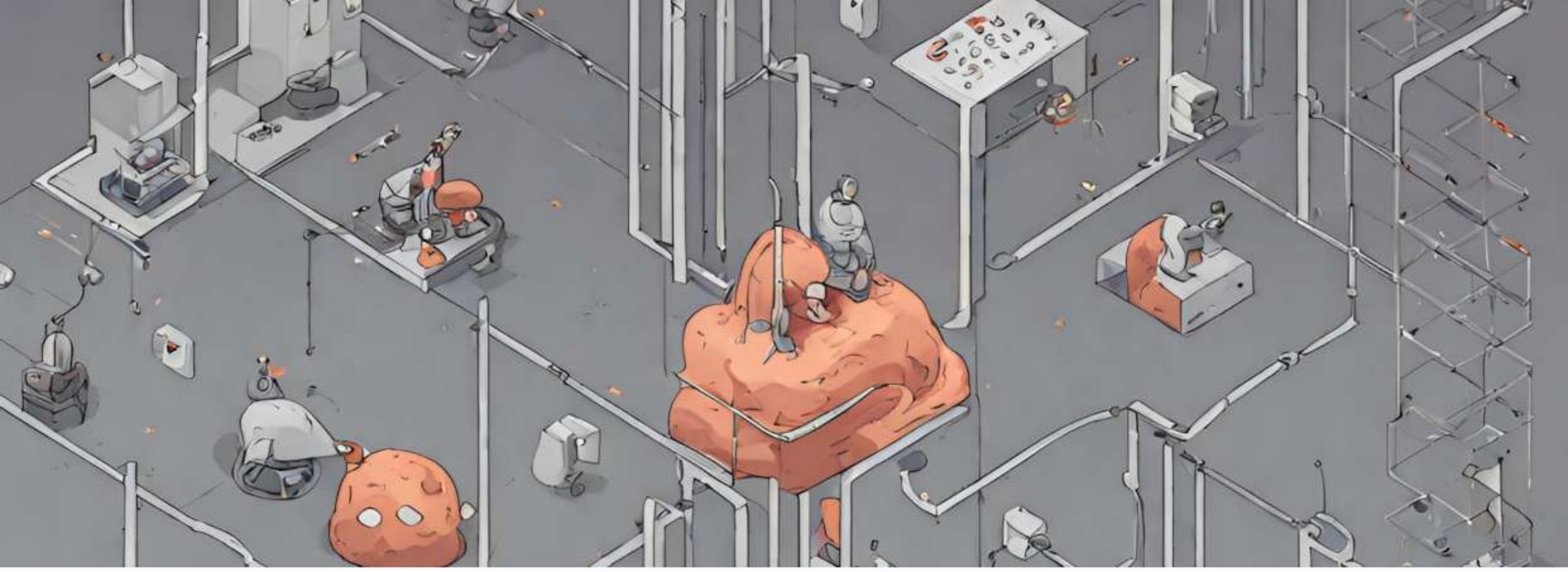
GENERATIVE MODELS

This involves learning the underlying distribution of data and using it to generate new samples. An example would be generating new images based on a dataset of existing images.

ASSOCIATION RULE MINING

This involves identifying patterns and relationships in data. An example would be analysing customer purchase histories to identify product associations.

LivAI
sepie
Funded by the
Erasmus+ Programme
of the European Union



1.5.3 REINFORCEMENT LEARNING

Reinforcement learning is a type of machine learning in which an agent learns how to behave in an environment by performing actions and receiving feedback in the form of rewards or punishments. The goal of the agent is to learn a policy that maximises the total cumulative reward it receives over time. Unlike supervised learning, reinforcement learning does not require labelled data or explicit instructions on what actions to take in a given situation. Instead, the agent must explore the environment and learn from trial and error.

LivAI sepie Funded by the Erasmus+ Programme of the European Union

SOME EXAMPLES OF APPLICATION OF UNSUPERVISED MACHINE LEARNING ARE:

GAME PLAYING

Reinforcement learning has been used to train computers to play complex games like chess, Go, and poker. For example, the program AlphaGo developed by Google DeepMind used reinforcement learning to defeat human champions at the game of Go.

ROBOTICS

Reinforcement learning is used in robotics to teach robots how to perform specific tasks, such as walking or grasping objects. The robot receives rewards or punishments based on its actions, which helps it learn how to perform the task more effectively.

AUTONOMOUS VEHICLES

Reinforcement learning can also be used to train self-driving cars. The car receives rewards for making good driving decisions and penalties for making poor ones. Over time, the car learns how to make the best decisions in different driving situations.

RECOMMENDATION SYSTEMS

Reinforcement learning is used in recommendation systems to learn about users' preferences and make personalised recommendations. The system receives feedback from the user about whether the recommendations are helpful or not, which helps it improve its predictions.

INDUSTRIAL CONTROL SYSTEMS

Reinforcement learning can be used to optimise the performance of industrial control systems, such as power plants or chemical processing plants. The system learns how to make adjustments to maximise efficiency while minimising costs and risks.

LivAI sepie

Funded by the
Erasmus+ Programme
of the European Union

1.5.4 ARTIFICIAL NEURAL NETWORKS

Artificial neural networks are a type of machine learning algorithm that is modelled after the structure and function of the human brain (see Figure 6). Artificial neural networks are composed of interconnected nodes (neurons) that process and transmit information using weighted connections. They are designed to recognize patterns and relationships within input data, and can be trained to make predictions or decisions based on that data.

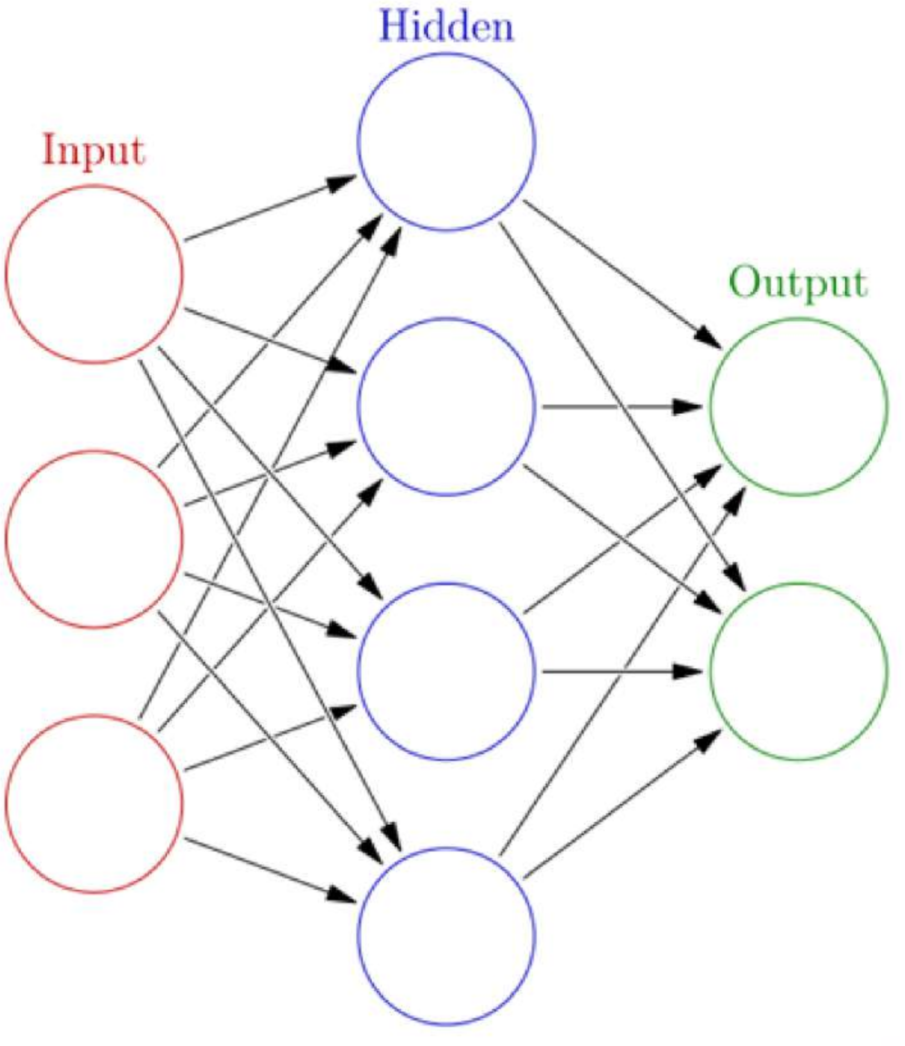


Figure 6:
An illustration of an Artificial neural network. This network has three layers of neurons: the input layer, one hidden layer, and the output layer. It is fully connected, meaning that any neuron in one layer is connected to all other neurons in the following layer.

ARTIFICIAL NEURAL NETWORKS CAN BE USED FOR A WIDE RANGE OF TASKS, SUCH AS:

IMAGE AND SPEECH RECOGNITION

Neural networks are used in computer vision and speech recognition systems, such as facial recognition in security systems, autonomous driving systems, and virtual assistants like Siri and Alexa.

NATURAL LANGUAGE PROCESSING

Neural networks are used to improve natural language processing, enabling machines to understand human language and respond accordingly. Examples of this include language translation systems and chatbots.

MEDICAL DIAGNOSIS

Neural networks can be used to diagnose and predict medical conditions based on patient data, such as symptoms, medical history, and test results.

FINANCIAL ANALYSIS

Neural networks can be used in finance to predict stock prices, detect fraud, and analyse financial data.

ROBOTICS

Neural networks can be used to control robotic systems, allowing them to adapt and learn from their environment.

GAMING

Neural networks are used in gaming to create intelligent opponents that can learn from the player’s actions and adjust their strategies accordingly.

LivAI sepie
Funded by the Erasmus+ Programme of the European Union

PREDICTIVE MAINTENANCE

Neural networks can be used to predict when equipment or machinery is likely to fail, allowing for proactive maintenance and reducing downtime.

MARKETING AND ADVERTISING

Neural networks can be used to analyse customer behaviour and preferences, allowing companies to create more targeted and effective marketing campaigns.

ENVIRONMENTAL MONITORING

Neural networks can be used to analyse environmental data, such as air quality and weather patterns, and make predictions about future conditions.

MATERIALS SCIENCE

Neural networks can be used to predict the properties of materials based on their chemical composition, enabling the design of new materials with specific properties.

1.5.5 DEEP LEARNING

Deep learning is a subfield of machine learning that involves training artificial neural networks with multiple layers to solve complex problems. It uses a hierarchical approach to learning, where each layer of the network learns to represent increasingly abstract features of the data.

The goal of deep learning is to develop algorithms that can automatically learn to recognize patterns and make accurate predictions or decisions, without the need for explicit programming.

LivAI sepie Funded by the Erasmus+ Programme of the European Union

It has been used in a wide range of applications such as:

- **Image and object recognition**: Deep learning algorithms can learn to identify and classify objects in images or videos, making it useful in applications such as self-driving cars, facial recognition, and medical imaging.

- **Natural language processing**: Deep learning can be used to analyse and process natural language, allowing for applications such as speech recognition, language translation, and sentiment analysis.

- **Generative models**: Deep learning can be used to generate new data, such as images, videos, and music, by training generative models.

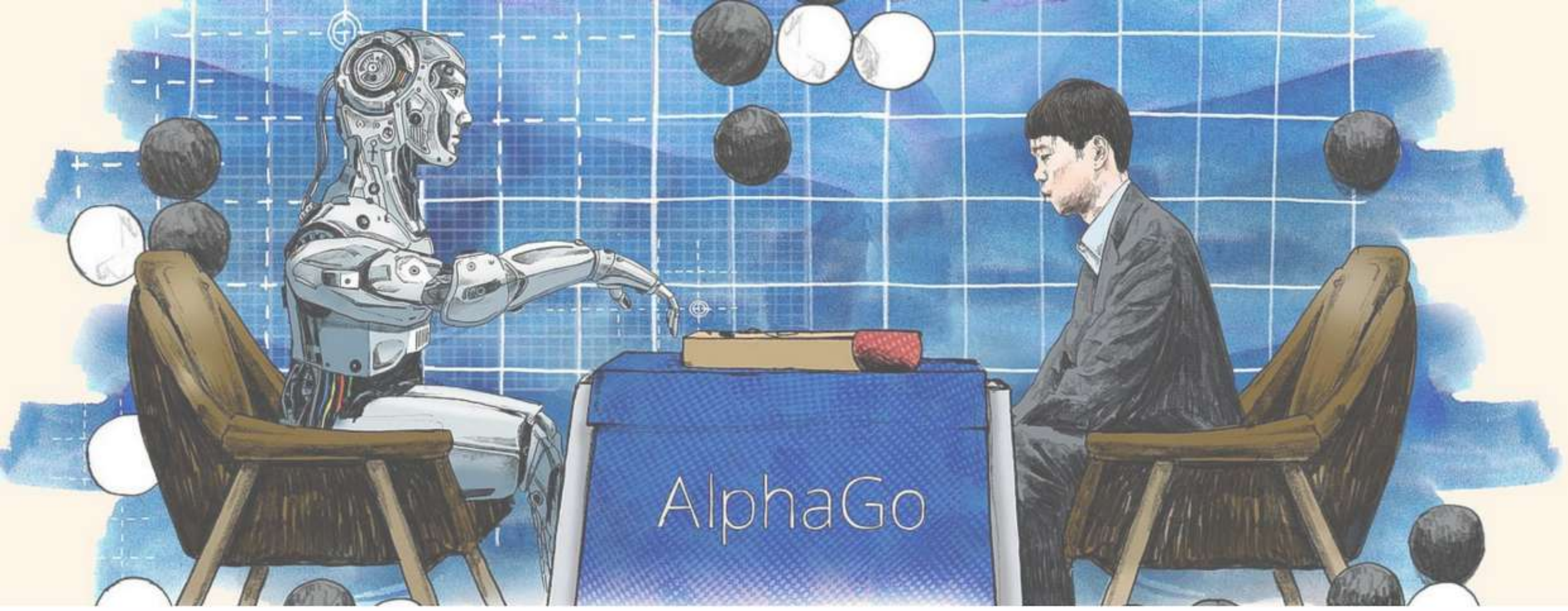
- **Recommender systems**: Deep learning can be used to analyse user data and provide personalised recommendations for products, services, and content.

- **Robotics**: Deep learning algorithms can be used in robotics to enable robots to perceive and interact with the environment, making them more autonomous and capable of performing complex tasks.

- **Financial forecasting**: Deep learning can be used in financial analysis and forecasting, where it can help identify patterns and predict market trends based on large amounts of data.

- **Drug discovery**: Deep learning can be used in drug discovery, where it can help identify potential drug candidates and predict their effectiveness based on molecular structures and biological data.

Recently, deep neural networks have made achievements that match or even surpass the capabilities of humans in some very specific tasks. Some examples of these achievements are ->

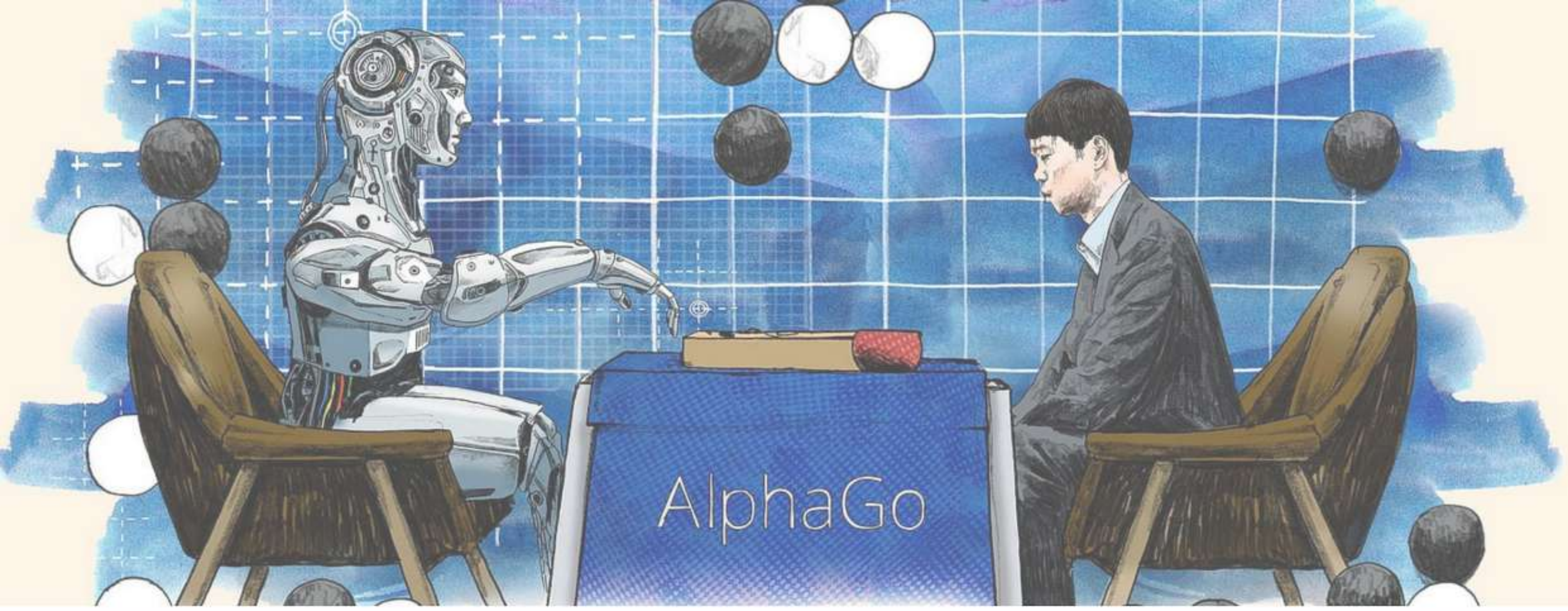


ALPHAGO

AlphaGo is a computer program developed by DeepMind Technologies, which uses deep neural networks to play the board game Go. Deep learning played a key role in the development of AlphaGo, allowing it to learn from millions of previous games and develop strategies that were beyond the reach of previous Go-playing programs.

AlphaGo was first trained on a large dataset of human games, and then further refined through reinforcement learning, in which the program played millions of games against itself and learned from its own successes and failures. The result was a program that was able to defeat the world's top human Go players, including the reigning world champion, in a series of highly publicised matches in 2016.

LivAI sepie Funded by the Erasmus+ Programme of the European Union



The success of AlphaGo was a major breakthrough in the field of AI, demonstrating the potential of deep learning to master complex games and other tasks that were previously thought to be beyond the reach of computers. It also sparked renewed interest in the field of AI and has led to many new developments and applications of deep learning in a wide range of fields.AI

LivAI sepie Funded by the Erasmus+ Programme of the European Union

TESLA AUTOPILOT

Tesla Autopilot is a system that uses various sensors and cameras to assist drivers in steering, accelerating, and braking their vehicles. Deep learning is an important technology behind this system. Tesla Autopilot uses a deep neural network that has been trained on large amounts of data to recognize and classify various objects in real-time, such as other vehicles, pedestrians, and road signs.

The neural network used in Tesla Autopilot is a convolutional neural network, which is a type of artificial neural network that is particularly well-suited to image recognition tasks. The network is trained on a large dataset of labelled images, with each image labelled according to the objects it contains, such as cars, trucks, pedestrians, and cyclists. This training data is used to teach the network how to recognize and classify objects in new images it encounters while driving.

LivAI sepie
Funded by the
Erasmus+ Programme
of the European Union

Once the neural network has been trained, it is integrated into the Tesla Autopilot system, which uses the network's object recognition capabilities to help the car navigate its environment.

For example, the system can detect other vehicles on the road and adjust the car's speed and position accordingly, and it can also detect and respond to traffic lights, stop signs, and other road markings.

Overall, the use of deep learning in Tesla Autopilot is an important example of how AI technologies are being integrated into real-world applications to improve safety and efficiency in a variety of industries.

CHATGPT

ChatGPT, a large language model developed by OpenAI, which uses deep learning techniques to generate natural language responses to user inputs. I have been trained on vast amounts of text data and can generate text in a variety of styles and formats, including conversational dialogue, informative articles, and more.

My abilities allow me to assist with a wide range of tasks, from answering questions and providing information to generating creative writing and even aiding in research and decision-making processes.

1.6 ARTIFICIAL INTELLIGENCE ISSUES AND ETHICAL CONCERNS

Despite the fact that AI has the power to transform many sectors and enhance our quality of life, it also brings up serious ethical and societal issues. For instance, as AI becomes more common in the workforce, there are worries about employment loss and the possibility that machines will eventually outperform humans in many tasks. Creating machines that could possibly outperform human intelligence raises ethical questions as well as worries about the potential misuse of AI for surveillance and other sinister purposes.

It is critical to have a thorough grasp of the risks and advantages of AI as well as to create ethical frameworks and regulations to direct its creation and application in order to allay these worries. This involves making sure AI is created and used in a transparent and accountable manner and that the necessary security and privacy measures are in place.

Machine Learning models do not contain biasses. biasses are introduced through the data sets. An example to show what is meant by the above statement. Statistically we know that in Spain identical job profiles have different salaries depending on the gender of the person.

To put it the other way around, women, in general, earn less in Spain for the same job than men. This is statistically proven, and we all regret that this is the case.

Let's suppose now that we create a Machine learning model that proposes the annual salary that we can propose to a person for a job. Suppose further that we feed this model with a dataset whose samples contain attributes for age, education level, years worked, any other attribute we find interesting, and the person's gender.

By including the person's gender in our model, we are introducing a bias. What will happen when we ask the model to provide us with an annual salary to propose to the person? Will it be the same for women and men? Obviously not, since our model learned from the data that women have a lower average annual salary than men, at least in Spain.

If we want to avoid our models making biassed predictions we must remove possible biasses from the set of data sets with which we feed our Machine Learning models.

1.6.1 REGULATION ON ARTIFICIAL INTELLIGENCE

Regulation on AI is a growing area of concern for governments and policymakers around the world. While AI has the potential to bring many benefits to society, such as improved healthcare and transportation, it also poses risks and challenges that need to be addressed. The main areas of concern for regulation on AI include privacy, bias, safety, transparency, and accountability.

Several countries and regions have already taken steps to regulate AI. In 2018, the European Union (EU) issued guidelines for the ethical development and use of AI. The guidelines include seven principles: human agency and oversight, technical robustness and safety, privacy and data governance, transparency, diversity, non-discrimination and fairness, societal and environmental well-being, and accountability. The EU is also considering regulations that would require AI systems to be transparent and explainable, so that users can understand how they work and what decisions they are making.

In the United States, the Federal Trade Commission (FTC) has issued guidance for companies developing AI systems, encouraging them to consider issues such as bias, transparency, and explainability. The National Institute of Standards and Technology (NIST) has also developed a framework for the responsible use of AI in businesses and organisations.

Other countries, such as Canada and China, have also developed guidelines and regulations for AI development and use. In Canada, the government has established the Canadian Institute for Advanced Research (CIFAR), which supports research on AI and its impact on society. In China, the government has set up guidelines for the development and use of AI, with a focus on promoting innovation and industrial development.

1.6.2 THE EUROPEAN STRATEGY ON AI

The European approach to AI is centred around creating a human-centric and trustworthy framework for the development and deployment of AI. The European Union (EU) has recognized the potential benefits of AI and the need to ensure that its development and use align with ethical principles and respect fundamental rights, including privacy and non-discrimination.

The ”Regulatory framework proposal on AI” aims to establish clear requirements and obligations for developers, deployers, and users of AI, specifically for certain AI applications. Additionally, the regulations strive to alleviate administrative and financial burdens for businesses, especially SMEs.

These regulations are part of a broader package on AI, which includes an updated Coordinated Plan on AI. Collectively, the Regulatory framework and Coordinated Plan aim to safeguard the fundamental rights and safety of individuals and businesses regarding AI. Furthermore, they aim to enhance the adoption, investment, and innovation of AI throughout the European Union.

These regulations define four levels of risk in AI (see Figure 7).

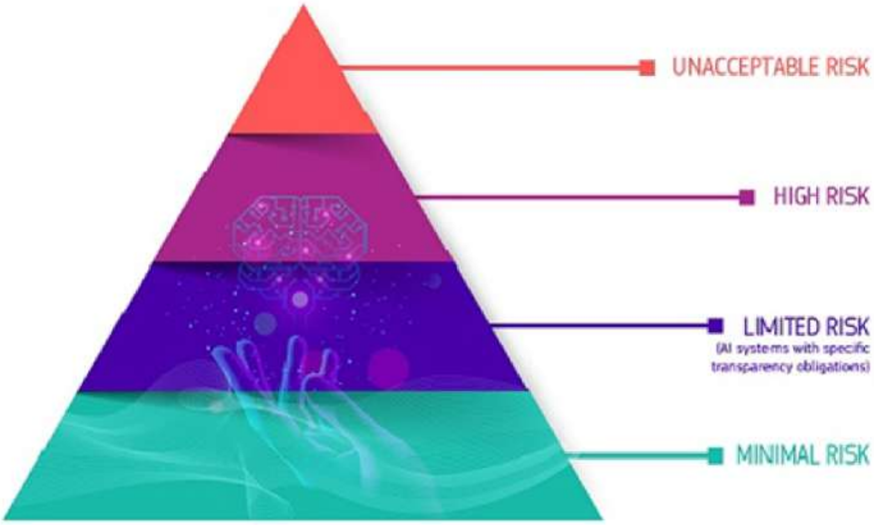


Figure 7: The four levels of risk in AI
(source: https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai)

The European approach to AI is based on ethical principles such as transparency, accountability, and responsibility, and is reflected in the EU's AI regulatory framework, the first of its kind in the world. The framework includes guidelines on how to assess the ethical impact of AI, as well as mandatory requirements for high-risk AI applications, such as facial recognition and self-driving cars.

The AI Act is a proposed European law on AI – the first law on AI by a major regulator anywhere. The law assigns applications of AI to three risk categories. First, applications and systems that create an unacceptable risk, such as government-run social scoring of the type used in China, are banned. Second, high-risk applications, such as a CV-scanning tool that ranks job applicants, are subject to specific legal requirements. Lastly, applications not explicitly banned or listed as high-risk are largely left unregulated.

University of Oxford researchers have created a tool called capAI, a procedure for conducting conformity assessment of AI systems in line with the EU AI Act. CapAI provides organisations with practical guidance on how to translate high-level ethics principles into verifiable criteria that help shape the design, development, deployment and use of ethical AI. This tool can be used to demonstrate that the development and operation of an AI system are trustworthy. The tool is being validated with firms at the moment and the most up-to-date version can be found here.

1.7 SUMMARY

In conclusion, the field of AI is one that is quickly developing and has the potential to drastically alter many facets of our lives. Even though AI development has the ability to have a lot of positive effects, it also raises important ethical and societal issues that need to be carefully considered and resolved.

We can make sure that this potent technology is utilised to benefit humanity, not harm it, by working to create responsible and ethical frameworks for AI creation and use.

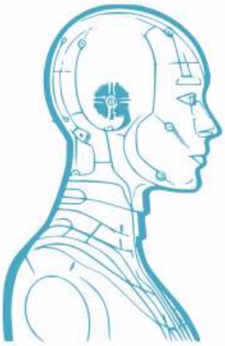
BIBLIOGRAPHY

- Russell S, Norvig P. Artificial intelligence: A modern approach, global edition. 4th ed. London, England: Pearson Education; 2021.
- A. M. Turing, A. M. (1950). Computing machinery and intelligence, Mind, Volume LIX, Issue 236, October 1950, Pages 433–460, , https://doi.org/10.1093/mind/LIX.236.433
- McCarthy, J. (2007). What is artificial intelligence

02

Big Data and Ethics

2.1 BIG DATA DEFINITION



There are numerous attempts found in the literature to define what is Big Data and it is relation to Artificial Intelligence. Big Data refers to large amounts of data produced very quickly by a high number of diverse sources. Data can either be created by people or generated by machines, such as sensors gathering climate information, satellite imagery, digital pictures and videos, purchase transaction records, GPS signals, and more in accordance with DG-Connect of the European Commission (1).

While several definitions of Big Data vary in the literature, they all have in common the fact that they refer to a large amount of data originating from different sources and in different, often unstructured, formats. In an attempt to define the characteristics, IBM defines the main traits of Big Data (5Vs) as follows (2):

- Volume, referring to the scale of data;
- Variety, since data is produced by different data sources in different formats;
- Velocity, which is connected to the analysis of streaming data;
- Veracity, as data is uncertain and needs to be verified before or during use;
- Value, which can be produced by analysing Big Data.

On the other hand, in accordance with the High Level Expert Group on Artificial Intelligence from EC (3), "Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions".

In this sense, Big Data and AI are tightly connected as the analytics performed on top of Big Data, referenced as Big Data Analytics, is in fact the use of processes and technologies, including AI, machine learning (ML) and Deep Learning (DL), in order to combine and analyse a massive number of datasets towards the aim of identifying patterns and developing actionable insights.

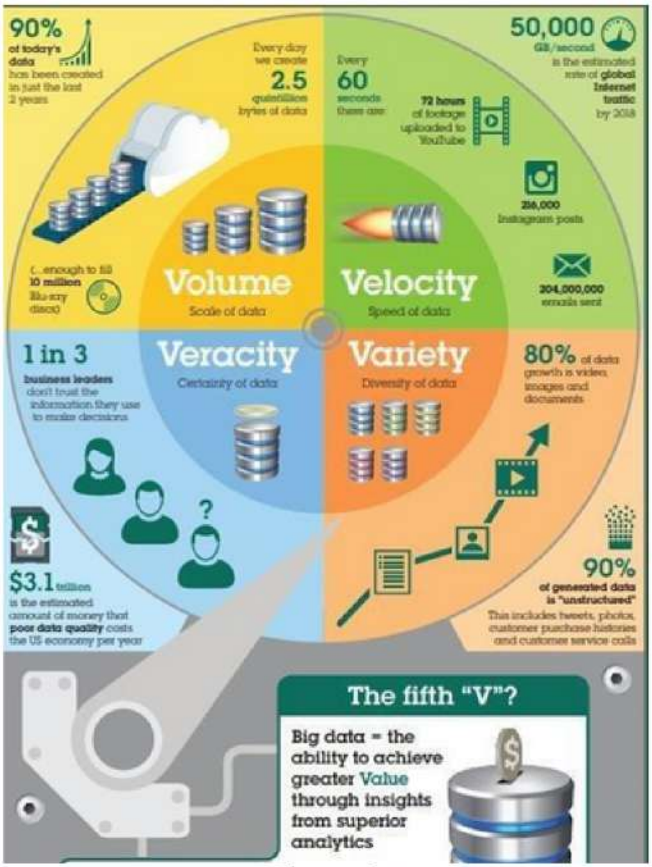

90% of today's data has been created in just the last 2 years
Every day we create 2.5 quintillion bytes of data
Every 60 seconds there are
72 hours of footage uploaded to YouTube
50,000 GB/second is the estimated rate of global Internet traffic by 2018
216,000 Instagram posts
204,000,000 emails sent
(...enough to fill 10 million Blu-ray discs)
Volume
Scale of data
Velocity
Speed of data
1 in 3 business leaders don't trust the information they use to make decisions
Veracity
Certainty of data
Variety
Diversity of data
80% of data growth is video, images and documents
$3.1 trillion is the estimated amount of money that poor data quality costs the US economy per year
90% of generated data is "unstructured"
This includes tweets, photos, customer purchase histories and customer service calls
The fifth "V"?
Big data = the ability to achieve greater Value through insights from superior analytics

Figure 1
Infographic that illustrates the meaning of the five V's related to Big Data (source: https://www.ibm.com/analytics/big-data-analytics)

From the ethical point of view, the ethical challenges and principles applied to Big Data are also applicable in the AI area and vice versa, since those two concepts are tightly connected as explained above.

Funded by the
Erasmus+ Programme
of the European Union

2.1.1 BIG DATA STAKEHOLDERS IDENTIFICATION

The three main categories of Big Data stakeholders, as acknowledged in the literature, are as follows (4):

BIG DATA COLLECTORS

The stakeholders under this category determine which data is collected, which is stored and for how long. They govern the collection, and implicitly the utility, of Big Data.

BIG DATA UTILISERS

The stakeholders under this category reside on the utility production side. While Big Data collectors might collect data with or without a certain purpose, the Big Data utilisers (re-)define the purpose for which data is used, for example regarding:

- Determining behaviour by imposing new rules on audiences or manipulating social processes;
- Creating innovation and knowledge through bringing together new datasets, thereby achieving a competitive advantage.

BIG DATA GENERATORS

The stakeholders under this category are as follows:

- Natural actors that by input or any recording voluntarily, involuntarily, knowingly, or unknowingly generate massive amounts of data.
- Artificial actors that create data as a direct or indirect result of their task or functioning.
- Physical phenomena, which generate massive amounts of data by their nature or which are measured in such detail that it amounts to massive data flows.

In the following section the major categories of the ethical challenges and principles for Big Data are elaborated and organised in the following axes:

FAIRNESS

It includes the principles related to the required fairness in the methods of collection and data retention in Big Data.

PRIVACY AND MISUSE OF PERSONAL DATA

It includes the core principles of privacy by design, data protection by design, data protection by default, as well as guidelines for the privacy impact assessment and the data protection assessment.

SECURITY & INTEGRITY

Under the umbrella of security and integrity fall the core principles of security by design, the cybersecurity compliance and data breaches incident handling.

ACCOUNTABILITY AND LIABILITY

It includes the ethical requirements related to responsibility regarding the functionality and resulting consequences of Big Data systems.

FAIRNESS
SECURITY AND INTEGRITY
ACCOUNTABILITY AND LIABILITY
PRIVACY AND MISUSE OF PERSONAL DATA

Figure 2: Major categories of the ethical challenges and principles for Big Data

Funded by the Erasmus+ Programme of the European Union

2.2 SECURITY AND INTEGRITY

Data security is crucial in ensuring data privacy and data protection. Taking into account available technology and cost of implementation, robust technical and organizational safeguards and procedures (including efficient monitoring of data access and data breach notification procedures) should be implemented to ensure proper data management throughout the data lifecycle and prevent any unauthorised use, disclosure or breach of personal data (5).

The increased volume and variety of information included in Big Data increases the risk of disclosure and its potential effects on individuals. As it is obvious, the risk of individual subjects being identified (or re-identified) after an unintentional, accidental or unauthorised disclosure remains a key ethical issue of Big Data (6). In an attempt to safeguard the security of the stored and used data, many secure data storage and access frameworks have been proposed, implemented and put in place across the different Big Data enabled solutions.

Intensive research and effort have been invested over the years in order to design and develop infrastructures and techniques which promise the secure and safe use of data residing on Big Data solutions. In this sense, security is a very important topic in the age of Big Data and the extended research on this topic results in a constantly evolving large list of state-of-the-art security and privacy guidelines and principles whose combination is attempting to provide solutions for a secure Big Data enabled environment.

Security and Integrity in terms of ethical challenges and principles can be grouped in the following pillars:

- **Security by design** principles that focus on the implementation of robust software solutions;
- **Compliance** to the **cybersecurity** certification framework and legislation set by EC;
- **Data breaches handling** guidelines and obligations.

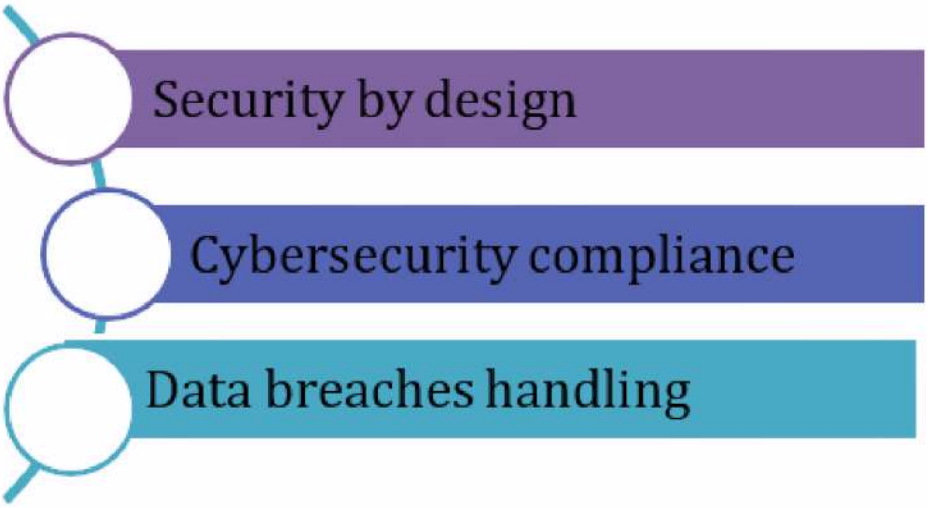
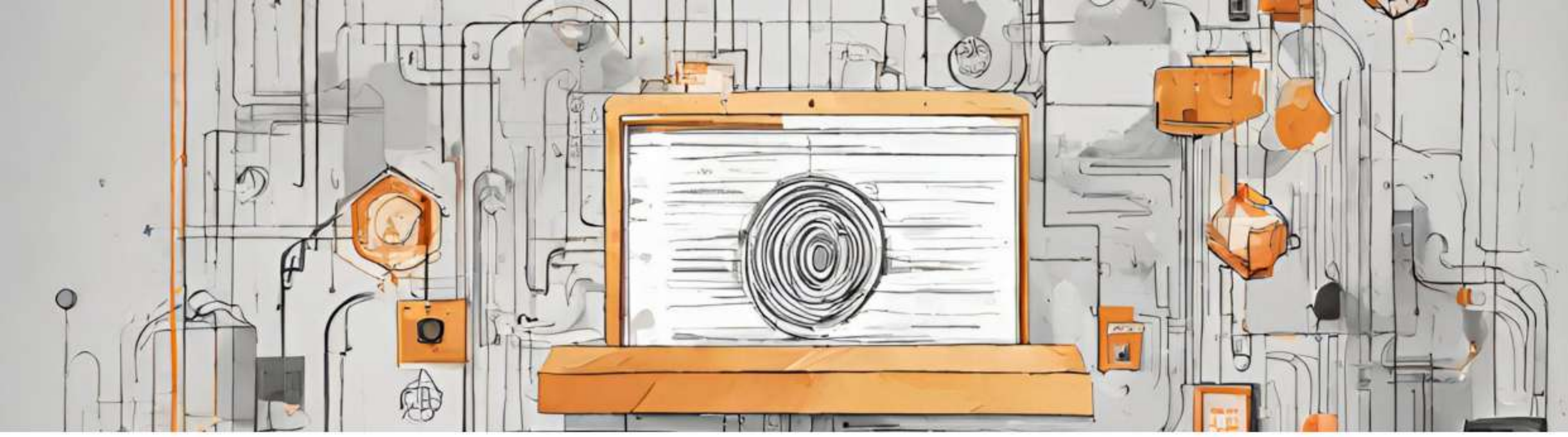


Figure 3: Big Data security and integrity pillars

LivAI sepie Funded by the Erasmus+ Programme of the European Union



2.2.1 SECURITY BY DESIGN

Security-by-design and security-by-default are important guiding principles in cybersecurity. Their ultimate goal is to reduce vulnerabilities in digital systems, services and processes. Security by design is the approach followed during the development process that focuses on making software as secure as possible starting from the early stages of the development process by adopting also the best programming practices.

In accordance with the Open Web Application Security Project (7), their effective application imposes the identification of non-functional security requirements early in the development life cycle of new products and services and their proper prioritisation with respect to functional ones. In the following paragraphs, the main principles of the security-by-design and security-by-default (8) are presented.

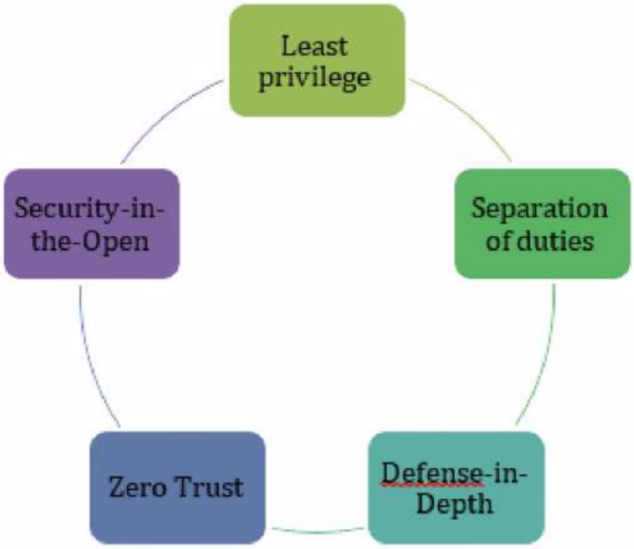
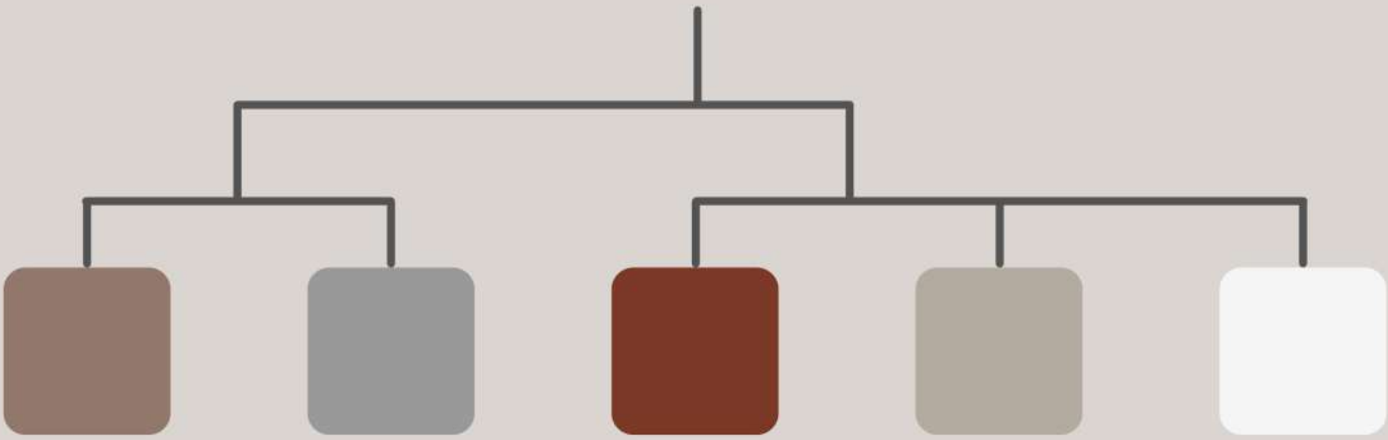


Figure 4: Security by design guiding principles

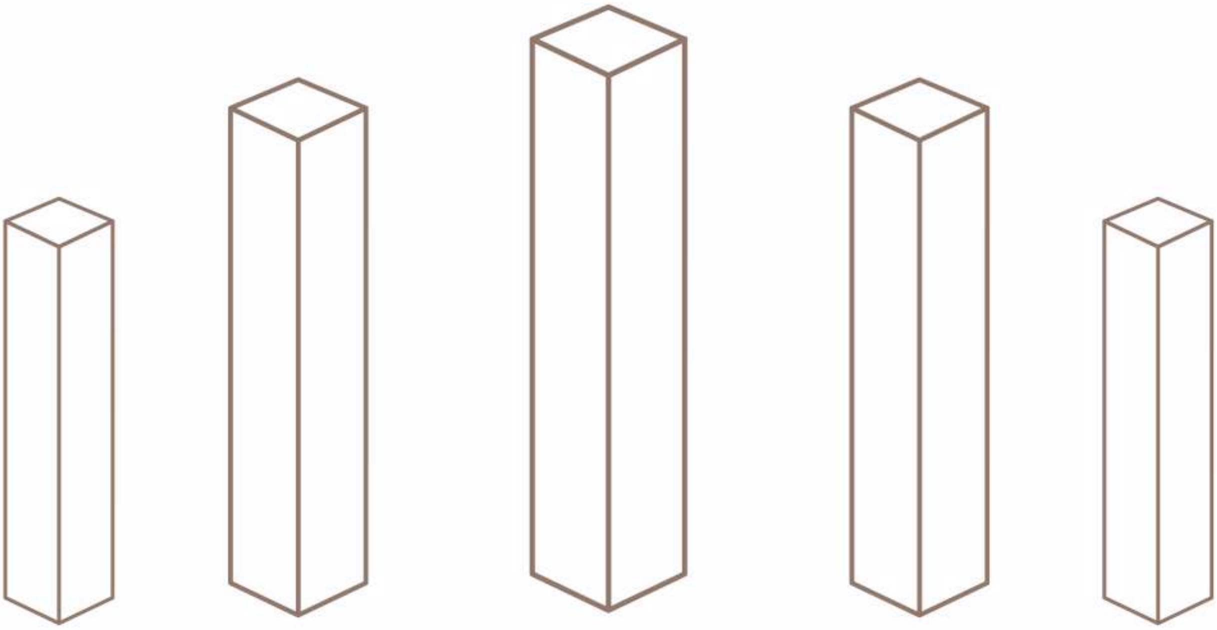
2.2.1.2 THE PRINCIPLE OF LEAST PRIVILEGE

Least Privilege is a security principle that states that users should only be given the minimum amount of access necessary to perform their job. This means that users should only be given access to the resources they need to do their job, and no more. This helps to reduce the risk of unauthorized access to sensitive data or systems, as users are only able to access the resources they need. Least Privilege is an important security principle that should be followed in order to ensure the security of an organization's data and systems.



2.2.1.3 THE PRINCIPLE OF SEPARATION OF DUTIES

Separation of duties is a fundamental principle of internal control in business and organizations. It is a system of checks and balances that ensures that no single individual has control over all aspects of a transaction. This is done by assigning different tasks to different people, so that no one person has control over the entire process. This helps to reduce the risk of fraud and errors, as well as ensures that all tasks are completed in a timely manner. Separation of duties is an important part of any organization’s internal control system, and is essential for maintaining the integrity of the organization’s financial records.



2.2.1.4 THE PRINCIPLE OF DEFENSE-IN-DEPTH

The principle of Defense-in-Depth is a security strategy that involves multiple layers of security controls to protect an organization's assets. It is based on the idea that if one layer of security fails, the other layers will still be able to protect the asset. The layers of security can include physical security, network security, application security, and data security.

The goal of Defense-in-Depth is to create a secure environment that is resilient to attack and can quickly detect and respond to any security incidents. By implementing multiple layers of security, organizations can reduce the risk of a successful attack and minimize the damage caused by any successful attack.

LivAI sepie Funded by the Erasmus+ Programme of the European Union

2.2.1.5 THE PRINCIPLE OF ZERO TRUST

Zero Trust is a security model that assumes that all users, devices, and networks are untrusted and must be verified before access is granted. It is based on the idea that organizations should not trust any user, device, or network, even if they are inside the organization's network. Instead, all requests for access must be authenticated and authorized before access is granted. Zero Trust also requires organizations to continuously monitor and audit user activity to ensure that access is only granted to those who need it. This model is designed to reduce the risk of data breaches and other security incidents by ensuring that only authorized users have access to sensitive data.

LivAI sepie Funded by the Erasmus+ Programme of the European Union

2.2.1.6 THE PRINCIPLE OF SECURITY-IN-THE-OPEN

Security-in-the-Open is a concept that emphasizes the importance of security in open source software development. It focuses on the need for developers to be aware of the security implications of their code and to take steps to ensure that their code is secure. This includes using secure coding practices, testing for vulnerabilities, and using secure development tools. Security-in-the-Open also encourages developers to collaborate with security experts to ensure that their code is secure.

LivAI sepie Funded by the Erasmus+ Programme of the European Union

2.2.2 THE CYBERSECURITY ACT AND THE NEWLY INTRODUCED NIS2 DIRECTIVE

One of the six general ethical principles that any system must preserve and protect based on fundamental rights as enshrined in the Charter of Fundamental Rights of the European Union (EU Charter) is ”Privacy and Data governance” (3) which dictates that people have the right to privacy and data protection and these should be respected at all times. Another report (9) states that strong security measures must be set in place to prevent data breaches and leakages. In addition to this, compliance with the EU Cybersecurity Act and international security standards may offer a safe pathway for adherence to these ethical principles.

The EU Cybersecurity Act introduces an EU-wide cybersecurity certification framework for ICT products, services and processes by setting common cybersecurity standards for connected devices and services, digital products and associated services that are placed on the EU market (10).

LivAI sepie Funded by the
Erasmus+ Programme
of the European Union

Article 51 of the EU Cybersecurity Act (11) states that cyber security certification schemes issued under the framework must achieve a number of cyber security objectives, including:

- **To protect data against** accidental or unauthorised storage, processing, access, disclosure, destruction, loss, alteration or lack of availability during the entire lifecycle of the ICT product, service or process.
- **That authorised** persons, programs or machines are able to access only the data, services or functions to which their access rights refer.
- **To verify** that ICT products, services and processes do not contain known vulnerabilities.
- **To record and make it possible to check** which data, services or functions have been accessed, used or otherwise processed, at what times and by whom.
- **To restore** the availability and access to data, services and functions in a timely manner in the event of a physical or technical incident.
- **That ICT** products, services and processes are secure by design and by default.
- **That ICT** products, services and processes are provided with up-to-date software and hardware that do not contain publicly known vulnerabilities, and are provided with mechanisms for secure updates.

The NIS2 Directive (12) is the EU-wide legislation on cybersecurity. It provides legal measures to boost the overall level of cybersecurity in the EU. In particular, the EU cybersecurity rules introduced in 2016 were updated by the NIS2 Directive that came into force in 2023 modernizing the existing legal framework to keep up with increased digitisation and an evolving cybersecurity threat landscape. The NIS2 Directive expanded the scope of the cybersecurity rules to new sectors and entities, to further improve the resilience and incident response capacities of public and private entities, competent authorities and the EU as a whole.

The NIS2 Directive imposes measures sets measures for a high common level of cybersecurity across the EU and provides legal measures to boost the overall level of cybersecurity in the EU. In particular, Businesses identified by the Member States as operators of essential services in the above sectors will have to take appropriate security measures and notify relevant national authorities of serious incidents. Key digital service providers, such as search engines, cloud computing services and online marketplaces, will have to comply with the security and notification requirements under the Directive.

The following measures are included in NIS2 Directive (13):

- Risk analysis and information system security policies.
- Incident handling (prevention, detection, and response to incidents).
- Business continuity and crisis management.
- Supply chain security.
- Security in network and information systems.
- Policies and procedures for cybersecurity risk management measures.
- The use of cryptography and encryption.

Article 21 of the directive covers the cybersecurity risk management measures and lists the following 10 areas as the minimum recommendation (14):

- Policies on risk analysis and information system security
- Incident handling
- Business continuity and crisis management
- Supply chain security
- Security in network and information systems acquisition, development and maintenance
- Policies and procedures to assess the effectiveness of cybersecurity risk-management measures
- Basic cyber hygiene practices and cybersecurity training
- Policies and procedures regarding the use of cryptography and, where appropriate, encryption
- Human Resource (HR) security, access control policies and asset management
- Multi Factor Authentication, continuous authentication, and secure communications where appropriate

2.2.3 DATA BREACHES

Following the evolution of the cloud and computing offerings of this new era, new technologies are leveraged as more data becomes available. These data are usually stored in (non-) relational databases which are accessible in the cloud and are securely shared among different stakeholders which however increases the risk of disclosure through a potential data breach. Data breach constitutes the intentional or inadvertent exposure of confidential information to unauthorized parties (15).

In this Big Data era, data breaches pose severe legal, financial and reputational threats to the organisations maintaining Big Data solutions as well as to disclosure threats to individuals whose data reside in such solutions. As the volume of data generated and collected is exponentially growing nowadays, the risk and potential consequences of a data breach are higher than ever before and despite the intensive research efforts this matter remains one of the most pressing security concerns.

Hence, Big Data raises a significant number of security questions related to the access of the underlying data, their storage and the usage of the stored data especially when it comes to personal and sensitive data which are potentially included in the collected data. Typically, data breaches are divided into two main adversarial attacks. In the first attack type the attackers aims at gaining access to raw information to either tamper the data analysis process hence its credibility and accuracy or to obtain access to a large volume of sensitive data that will be later used mostly for illegal purposes (such as credit card information). In the second type of attack, the attacker aims at gaining access to processed or analysed information as extracted by multiple data sources. In both cases, the individuals face the exposure of their sensitive information, personal data and other confidential information (e.g. credit card number) to malicious users or the public.

On the other hand, the data collectors face the consequences of the data breach in multiple levels spanning from the reputational penalties in terms of brand image and loyalty, financial damages in terms of loss of market share, customers and intellectual property to legal fines and penalties imposed by the privacy regulations such as GDPR.

LivAI sepie Funded by the Erasmus+ Programme of the European Union

The EU Regulation 1725/2018 dictates that all European institutions and bodies have a duty to report certain types of personal data breaches to the European Data Protection Supervisor (EDPS). EDPS is the European Union's independent data protection authority responsible for supervising the processing of personal data by the European institutions, bodies and agencies (EUIs).

In accordance with the aforementioned regulation, the following requirements are imposed (16):

- **Every EU institution** must do this within 72 hours of becoming aware of the breach, where feasible.
- **If the breach is likely to pose** a high risk of adversely affecting individuals' rights and freedoms, the EU Institution must also inform the individuals concerned without unnecessary delay.
- **EU Institutions must ensure** that they have prevention and detection mechanisms in place for personal data breaches, as well as investigation and internal reporting procedures.
- **EU Institutions must also ensure** that when they identify a personal data breach, they are able to respond effectively to mitigate the negative effects of the breach on the individuals whose data has been compromised.

In accordance with the Guidelines on personal data breach notification for the European Union Institutions and Bodies published by EDPS (17), the approach that should be taken in order to adequately respond to a personal data breach is described in the following guidelines:

- **Not every information security incident** is a personal data breach, but every personal data breach is an information security incident.
- **While assessing each reported incident**, it should be detected if personal data is affected.
- **If personal data is affected**, the security incident will be considered a personal data breach.
- **As soon as there is an indication** that a security incident might affect personal data, the Data Protection Officer (DPO) shall be immediately consulted.
- **Once the security incident is considered** a personal data breach, as the next step, it should be assessed what would be the impact of the incident on individuals; rights and freedoms.
- **Once the security incident is considered a personal data breach, the EUI shall assess** the impact of the breach on the rights and freedoms of data subjects. This assessment shall be as objective as possible. This step is very critical as it will define the notification obligations of the EUI as a controller.

Funded by the
Erasmus+ Programme
of the European Union

- **A EUI shall implement** its own personal data breach management procedure or set of policies that will focus on the impact assessment of every reported personal data breach and the selection of the adequate notification procedure for the EDPS and the data subjects. Roles and responsibilities shall be clearly defined.
- **It is of utmost importance the EUI ensures** a correct assessment of the risks as a trigger for notification to the EDPS and possible communication to a data subject.
- **In cases where there is reported evidence** that a recorded personal data breach creates no risk to the rights and freedoms of data subjects, the controller will not need to notify the EDPS or the data subjects. However, this decision should be taken at the appropriate level and should be well documented, in order to enable the EDPS to verify compliance of the EUI also for data breaches which were not notified.
- **EUI shall integrate**, in the data breach management procedure or in a separate procedure, a step by step guidance or a methodology that will aim at the objective assessment of the level of risk of a personal data breach

- **According to Art. 34 of the Regulation,** a EUI should notify a personal data breach not later than 72 hours to the European Data Protection Supervisor, unless it is unlikely to result in a risk to the rights and freedoms of individuals.
- **Furthermore, according to Art. 35 (1) of the Regulation,** in case the personal data breach result in a ”high risk” to the rights and freedoms of individuals, the EUI should also communicate it to the data subjects.
- **In all cases, the controller must mitigate the effects of any personal data breach and in particular the impact on data subjects.**

- **In addition to this,** in the same report it is clarified that the data breach notification obligation reflects a risk based approach. In this sense, the following should be taken into consideration:
- **Severity of breaches** will need to be assessed on a case-by-case basis. The ”risk to the rights and freedoms of natural persons” should be taken as a basis for consideration when conducting an assessment. The risks identified during a Data Protection Impact Assessment (DPIA) can serve as a starting point.
- **Assessing which data breaches** entail a risk and which entail a high risk is relevant for the notification and communication obligation. In case of a risk which is not high the EUI will only notify the EDPS as the supervisory authority whereas in cases of high risk there is the obligation to communicate also to data subjects.
- **Recitals 46 and 47 of the Regulation** provide that when assessing a risk, consideration should be given to both the likelihood and severity of the risk to the rights and freedoms of data subjects. Then, the risk should be evaluated on the basis of an objective assessment. With an actual data breach, the event has already occurred, and so the focus of the controller is solely on the impact of the breach on individuals.
- **The assessment of the data breach's** impact on the data subject is important as it will also help the EUI to take adequate measures to contain and address the breach.

- **As recommended by the ARTICLE 29 DATA PROTECTION WORKING** PARTY (WP29) in its guidelines, the factors to be taken into account when assessing the risks are:

NATURE, SENSITIVITY, AND VOLUME OF PERSONAL DATA

SPECIAL CHARACTERISTICS OF THE INDIVIDUAL

EASE OF IDENTIFICATION OF INDIVIDUALS

OSEVERITY OF CONSEQUENCES FOR INDIVIDUALS

SPECIAL CHARACTERISTICS OF THE DATA CONTROLLER

TYPE OF BREACH

THE NUMBER OF AFFECTED INDIVIDUALS

- All the above factors need to be carefully assessed each one separately or in combination with the others to indicate the level of the risks to the individuals.

- The risks identified during a DPIA can help the controllers during the process of assessing the risk. It is highly likely that data breaches on processing activities that needed a prior DPIA according to Article 39 of the Regulation, may cause higher risk to the rights and impacts on the individuals.

2.3 PRIVACY AND MISUSE OF PERSONAL DATA

Privacy broadly refers to the possibility to withhold personal information and prevent its use without consent (2). The right to privacy grants the ability to individuals to choose which information about parts of the self can be accessed by others and to control the extent, manner and timing of the use of those parts we choose to disclose (18). EDPS states that respect for human dignity is strictly interrelated with respect for the right to privacy and the right to the protection of personal data.

The impact of Big Data technologies on privacy (and thereby human dignity) ranges from group privacy and high-tech profiling, to data discrimination and automated decision making (19). In the current digital era, the volume of the data collected and stored in Big Data enabled solutions is exponentially increasing hence the need for privacy laws has been identified by many countries that are trying to cope with the changes in technology, as well as the unprecedented data collection and storage possibilities offered today. In addition to this, many organisations and research institutes are investing efforts in the definition and establishment of privacy-preserving guidelines, principles and frameworks.

Privacy in terms of ethical challenges and principles can be grouped in the following pillars:

- **Privacy by Design** defines the principles that are followed to safeguard privacy in software.

- **Data Protection by Design** that defines the specific legal obligations as established by Article 25 of the GDPR.

- **Data Protection by Default** that also defines the specific legal obligations in accordance with Article 25 of the GDPR.

- **Privacy Impact Assessment** that defines the objectives of the system in terms of privacy through a structured process.Data Protection Impact Assessment that defines the process which describes the applied processing procedures, assists in the management of risks related to the rights and freedoms of natural persons from these procedures, assesses them and determines the measures to address them.

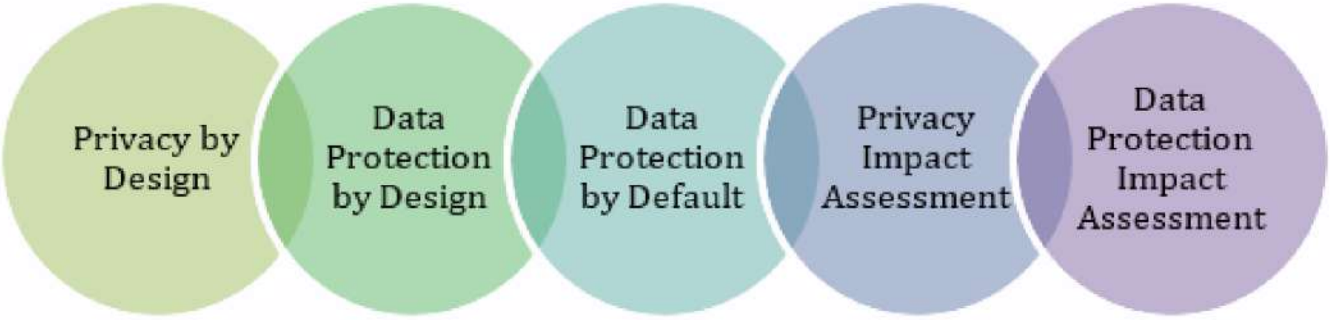

Privacy by Design
Data Protection by Design
Data Protection by Default
Privacy Impact Assessment
Data Protection Impact Assessment

Figure 5: Big Data privacy pillars

2.3.1 PRIVACY BY DESIGN

The International Conference of Data Protection and Privacy Commissioners defines privacy by design ”as a holistic concept that may be applied to operations throughout an organisation, end-to-end, including its information technology, business practices, processes, physical design and networked infrastructure”.

On the other hand, EDPS defines privacy by design as the broad concept of technological measures for ensuring privacy as it has developed in the international debate over the last few decades (20) [++]. The following subsections present the main privacy by design strategies found in literature.

2.3.1.1 PRIVACY DESIGN STRATEGIES

THE MAIN PRIVACY DESIGN STRATEGIES (21) ARE AS FOLLOWS

STRATEGY #1: MINIMISE

STRATEGY #2: HIDE

STRATEGY #3: SEPARATE

STRATEGY #4: AGGREGATE

STRATEGY #5: INFORM

STRATEGY #6: CONTROL

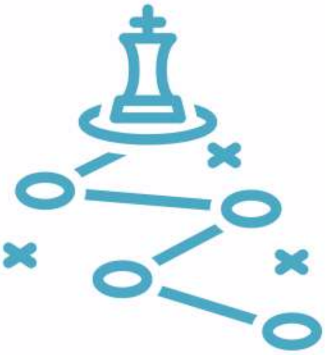
STRATEGY #7: ENFORCE

STRATEGY #8: DEMONSTRATE

LivAI sepie Funded by the Erasmus+ Programme of the European Union

STRATEGY #1

MINIMISE



The most basic privacy design strategy is **MINIMISE**, which states that the amount of personal data that is processed should be restricted to the minimal amount possible. This strategy is extensively discussed by Gürses et al. (22). By ensuring that no, or no unnecessary, data is collected, the possible privacy impact of a system is limited. Applying the MINIMISE strategy means one has to answer whether the processing of personal data is proportional (with respect to the purpose) and whether no other, less invasive, means exist to achieve the same purpose. The decision to collect personal data can be made at design time and at run time, and can take various forms. For example, one can decide not to collect any information about a particular data subject at all.

Alternatively, one can decide to collect only a limited set of attributes. Design patterns include the Common design patterns that implement this strategy are 'select before you collect' (23), the 'anonymisation and use pseudonyms' (24).

LivAI sepie
Funded by the Erasmus+ Programme of the European Union

STRATEGY #2

HIDE

The second design strategy, **HIDE**, states that any personal data, and their interrelationships, should be hidden from plain view. The rationale behind this strategy is that by hiding personal data from plain view, it cannot easily be abused. The strategy does not directly say from whom the data should be hidden. And this depends on the specific context in which this strategy is applied. In certain cases, where the strategy is used to hide information that spontaneously emerges from the use of a system (communication patterns for example), the intent is to hide the information from anybody. In other cases, where information is collected, stored or processed legitimately by one party, the intent is to hide the information from any other party. In this case, the strategy corresponds to ensuring confidentiality.

The HIDE strategy is important, and often overlooked. In the past, many systems have been designed using innocuous identifiers that later turned out to be privacy nightmares. Examples are identifiers on RFID tags, wireless network identifiers, and even IP addresses. The HIDE strategy forces one to rethink the use of such identifiers. In essence, the HIDE strategy aims to achieve unlinkability and unobservability (24).

Unlinkability in this context ensures that two events cannot be related to one another (where events can be understood to include data subjects doing something, as well as data items that occur as the result of an event). The design patterns that belong to the HIDE strategy are a mixed bag.

One of them is the use of encryption of data (when stored, or when in transit). Other examples are mix networks (25) to hide traffic patterns (25), or techniques to unlink certain related events like attribute based credentials (26), anonymisation and the use of pseudonyms. Techniques for computations on private data implement the HIDE strategy while allowing some processing. Note that the latter two patterns also belong to the MINIMISE strategy.

STRATEGY #3

SEPARATE

The third design strategy, **SEPARATE**, states that personal data should be processed in a distributed fashion, in separate compartments whenever possible. By separating the processing or storage of several sources of personal data that belong to the same person, complete profiles of one person cannot be made. Moreover, separation is a good method to achieve purpose limitation. The strategy of separation calls for distributed processing instead of centralised solutions. In particular, data from separate sources should be stored in separate databases, and these databases should not be linked. Data should be processed locally whenever possible, and stored locally if feasible as well. Database tables should be split when possible. Rows in these tables should be hard to link to each other, for example by removing any identifiers, or using table specific pseudonyms. These days, with an emphasis on centralised web based services this strategy is often disregarded.

However, the privacy guarantees offered by peer-to-peer networks are considerable. Decentralised social networks like Diaspora are inherently more privacy-friendly than centralised approaches like Facebook. No specific design patterns for this strategy are known at the moment.

LivAI sepie Funded by the Erasmus+ Programme of the European Union

STRATEGY #4

AGGREGATE

The fourth design pattern, **AGGREGATE**, states that Personal data should be processed at the highest level of aggregation and with the least possible detail in which it is (still) useful. Aggregation of information over groups of attributes or groups of individuals, restricts the amount of detail in the personal data that remains. This data therefore becomes less sensitive if the information is sufficiently coarse grained, and the size of the group over which it is aggregated is sufficiently large. Here coarse grained data means that the data items are general enough that the information stored is valid for many individuals hence little information can be attributed to a single person, thus protecting its privacy.

Examples of design patterns that belong to this strategy are the following: a) aggregation over time (used in smart metering), b) dynamic location granularity (used in location-based services), c) k-anonymity (27), d) differential privacy (28) and other anonymization techniques.

STRATEGY #5

INFORM

The **INFORM** strategy corresponds to the important notion of transparency. Data subjects should be adequately informed whenever personal data is processed. Whenever data subjects use a system, they should be informed about which information is processed, for what purpose, and by which means. This includes information about the ways the information is protected, and being transparent about the security of the system. Providing access to clear design documentation is also a good practice.

Data subjects should also be informed about third parties with which information is shared. And data subjects should be informed about their data access rights and how to exercise them. A possible design pattern in this category is the Platform for Privacy Preferences (P3P) (29). Data breach notifications are also a design pattern in this category. Finally, Graf et al. (30) provide an interesting collection of privacy design patterns for informing the user from the Human Computer Interfacing perspective.

Funded by the Erasmus+ Programme of the European Union

STRATEGY #6

CONTROL

The control strategy states that data subjects should be provided agency over the processing of their personal data. The **CONTROL** strategy is in fact an important counterpart to the INFORM strategy. Without reasonable means of controlling the use of one's personal data, there is little use in informing a data subject about the fact that personal data is collected. Of course the converse also holds: without proper information, there is little use in asking consent.

Data protection legislation often gives the data subject the right to view, update and even ask the deletion of personal data collected about her. This strategy underlines this fact, and design patterns in this class give users the tools to exert their data protection rights. CONTROL goes beyond the strict implementation of data protection rights, however. It also governs the means by which users can decide whether to use a certain system, and the way they control what kind of information is processed about them.

In the context of social networks, for example, the ease with which the user can update his privacy settings through the user interface determines the level of control to a large extent. So user interaction design is an important factor as well.

Moreover, by providing users direct control over their own personal data, they are more likely to correct errors. As a result the quality of personal data that is processed may increase.

Design patterns include the User centric identity management and the end-to-end encryption support control.

STRATEGY #7

ENFORCE

The seventh strategy, **ENFORCE**, states: A privacy policy compatible with legal requirements should be in place and should be enforced. This relates to the accountability principle. The ENFORCE strategy ensures that a privacy policy is in place. This is an important step in ensuring that a system respects privacy during its operation. Of course the actual level of privacy protection depends on the actual policy.

At the very least it should be compatible with legal requirements. As a result, purpose limitation is covered by this strategy as well. More importantly though, the policy should be enforced. This implies, at the very least, that proper technical protection mechanisms are in place that prevent violations of the privacy policy. Moreover, appropriate governance structures to enforce that policy must also be established.

Access control is an example of a design patterns that implement this strategy. Another example are sticky policies and privacy rights management which is a form of digital rights management involving licenses to personal data.

LivAI sepie

Funded by the
Erasmus+ Programme
of the European Union

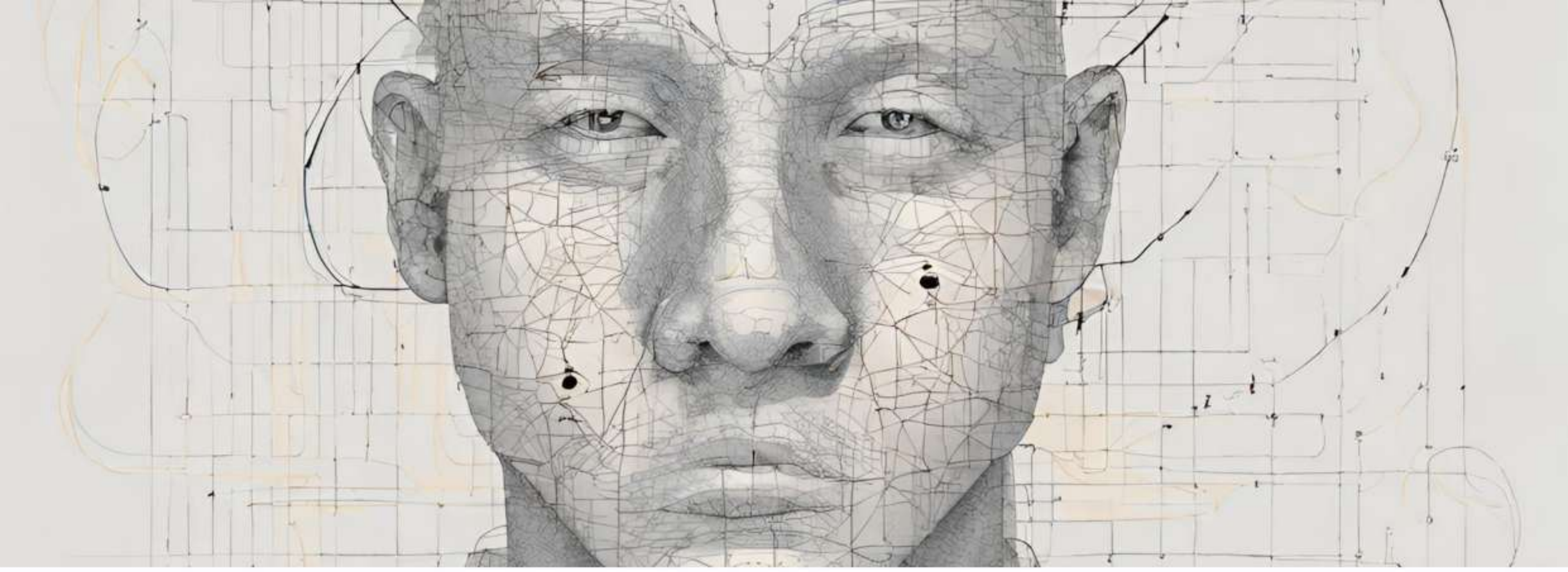
STRATEGY #8

DEMONSTRATE

The final strategy, **DEMONSTRATE**, requires a data controller to be able to demonstrate compliance with the privacy policy and any applicable legal requirements. This strategy supports the accountability principles. This strategy goes one step further than the ENFORCE strategy in that it requires the data controller to prove that it is in control. This is explicitly required in the GDPR regulation.

In particular this requires the data controller to be able to show how the privacy policy is effectively implemented within the IT system. In case of complaints or problems, he/she should immediately be able to determine the extent of any possible privacy breaches. Design patterns that implement this strategy are, for example, privacy management systems (31), and the use of logging and auditing.

LivAI sepie
Funded by the Erasmus+ Programme of the European Union



2.3.2 DATA PROTECTION BY DESIGN

In accordance with a report from ENISA (32), data protection by design is a process involving various technological and organisational components, which implement privacy and data protection principles by properly and timely deploying technical and organisation measures that include also the use of specific Privacy Enhancing Technologies (PETS). As stated by EDPS, data protection by design is defined as the specific legal obligations established by Article 25 of the GDPR (20) as with the full applicability of the General Data Protection Regulation in the EU as of 25 May 2018, data protection by design and by default becomes an enforceable legal obligation.

LivAI sepie Funded by the Erasmus+ Programme of the European Union

In accordance with the report from EDPS (20):

ARTICLE 25

Article 25 of the GDPR ”Data protection by design and by default” provides that the controller shall implement appropriate technical and organisational measures, both at the design phase of the processing and at its operation, to effectively integrate the data protection safeguards to comply with the Regulation and protect the fundamental rights of the individuals whose data are processed. Those measures shall be identified taking into account the state of the art, the cost of implementation and the nature, scope, context and purposes of the processing as well as the risks for the rights and freedoms of those individuals. The Article states that, by default, only personal data that are necessary for each specific purpose of the processing may be processed. The Article concludes that approved certification mechanisms may be used to demonstrate compliance with the set requirements.

ARTICLE 24

The data protection by design and by default requirement of Article 25 complements the controller’s responsibility laid down in Article 24, a core provision of the GDPR. This article defines ”who shall do what” to protect individuals and their personal data and states that a risk-based approach shall be adopted to identify what needs to be done to that purpose. More precisely, it provides for the controller to ”implement appropriate technical and organisational measures to ensure and to be able to demonstrate that processing is performed in accordance...” with the law. These measures shall be designed ”taking into account the state of the art, the cost of implementation and the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for the rights and freedoms of natural persons”.

ARTICLE 32

These include the rules of Article 32, which requires an IT security risk management framework and measures to mitigate risks for the individuals whose data are processed by adequately securing those data. It is useful to remind that, whereas the measures identified in Article 32 are just those targeting one of the data protection principles in Article 5, namely the one called ”integrity and confidentiality”, Article 24 refers to the implementation of all data protection principles and the compliance with the whole of the GDPR.

In the context of the controller’s responsibility to ensure and to be able to demonstrate compliance with the law, Article 25 aims at technical and organisational measures as required by Article 24, stressing some dimensions of their implementation process already implicitly present in Article 24 and adding others, making them all mandatory.

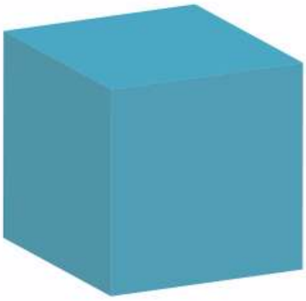
The same report defines the four dimensions of the obligation of data protection by design which are as follows (20)

THE FIRST DIMENSION

Is acknowledging the fact that processing of personal data, partially or completely supported by IT systems should always be the outcome of a design project. Article 25 requires consideration of safeguardsboth at the design and operational phase, thus aiming at the whole project lifecycle and clearly identifying the protection of individuals and their personal data within the project requirements.

THE SECOND DIMENSION

The risk management approach with a view to selecting and implementing measures for effective protection. The assets to protect are the individuals whose data are processed and in particular their fundamental rights and freedoms. In this respect, there is no indication of obligatory measures. Nonetheless, the legislator gives directions on those factors (nature, scope, context and purposes of processing) that the organisation must take into account in the selection of the appropriate measures. At the same time, the organisation is responsible for choosing the safeguards among those available (within the ”state of the art”) and consider their cost among the elements leading to the final decision, weighed against the risks for individuals. These two factors, the state of the art of available technology and the cost of implementation of the measures, must not be interpreted in such a way that the measures chosen do not sufficiently mitigate existing risks and the resulting protection is not adequate.



THE THIRD DIMENSION

The need for these measures to be appropriate and effective. The effectiveness is to be benchmarked against the purpose of those measures: to ensure and be able to demonstrate compliance with the GDPR, to implement the data protection principles and to protect the rights of individuals whose data are processed. In particular, Article 25 provides for those measures to be designed ”to implement data protection principles ... in an effective manner”. These data protection principles, set out in Article 5, can be considered as the goals to achieve. They have been singled out by the legislator as a cornerstone for the protection of individuals when processing their data and are complemented in the GDPR by either more detailed rules (i.e. the information

LivAI sepie
Funded by the
Erasmus+ Programme
of the European Union

to provide to the individuals and their rights as ”data subjects”, which are further elaborated on the ”transparency” principle; or the security obligations of Article 32) or by other accountability instruments, such as the documentation duties of Article 30, which are instrumental to those principles. This means that effectively meeting those principles/goals, as further detailed in the law by other provisions, would ensure the expected protection of personal data.

THE FOURTH DIMENSION

Is the obligation to integrate the identified safeguards into the processing. The GDPR includes some safeguards to protect the individuals whose data are processed through means that are ”external” to the processing itself, such as data protection notices for example. This dimension instead focuses on the need to protect the individuals by directly protecting their data and the way they are managed.

All four dimensions are equally important and become an integral part of accountability and will be subject to supervision from the competent data protection supervisory authorities where appropriate.

2.3.3 DATA PROTECTION BY DEFAULT

Based on the report from EDPS (20), in order to adhere to the data protection by default principle as defined by GDPR organisations must, by default, only process personal data necessary for each specific purpose defined in compliance with the law and transparently notified to the individuals concerned. While it can be argued that this obligation is already implicit in the ”purpose limitation” and ”data minimisation” principles in both the design and operation phases, the explicit rule stresses the importance of taking technical measures to meet the expectations of the individuals whose data are processed, not to have their data processed for other purposes than what the product and service is basically and strictly meant to do, leaving by default any further use turned off, for instance through configuration settings.

The same report as states that some of the added value of the data protection by default provision is also the further elaboration of the principle of data minimisation and the extension to the principle of storage limitation. Article 25(2) explains how the obligation to process by default only personal data that are necessary ”applies to the amount of personal data collected, the extent of their processing, the period of their storage and their accessibility...”. The Article then establishes a precise obligation by instantiating the general principle in one particular use case: organisations shall set up measures to prevent personal data from being made public by default.

In addition, the European Data Protection Board have issued the following guidelines in their report (33):

A ”default” as commonly defined in computer science, refers to the preexisting or pre selected value of a configurable setting that is assigned to a software application, computer program or device. Such settings are also called ”presents” or ”factory presents”, especially for electronic devices. Hence, the term ”by default” when processing personal data, refers to making choices regarding configuration values or processing options that are set or prescribed in a processing system, such as a software application, service or device, or a manual processing procedure that affect the amount of personal data collected, the extent of their processing, the period of their storage and their accessibility.

The controller should choose and be accountable for implementing default processing settings and options in a way that only processing that is strictly necessary to achieve the set, lawful purpose is carried out by default. Here, controllers should rely on their assessment of the necessity of the processing with regards to the legal grounds of Article 6(1). This means that by default, the controller shall not collect more data than is necessary, they shall not process the data collected more than is necessary for their purposes, nor shall they store the data for longer than necessary. The basic requirement is that data protection is built into the processing by default. The controller is required to predetermine for which specified, explicit and legitimate purposes the personal data is collected and processed (per Art. 5(1)(b), (c), (d), (e) GDPR). The measures must by default be appropriate to ensure that only personal data which are necessary for each specific purpose of processing are being processed. The EDPS ”Guidelines to assess necessity and proportionality of measures that limit the right to data protection of personal data” can be useful also to decide which data is necessary to process in order to achieve a specific purpose (34) (35) (36).

If the controller uses third party software or off-the-shelf software, the controller should carry out a risk assessment of the product and make sure that functions that do not have a legal basis or are not compatible with the intended purposes of processing are switched off. The same considerations apply to organisational measures supporting processing operations. They should be designed to process, at the outset, only the minimum amount of personal data necessary for the specific operations. This should be particularly considered when allocating data access to staff with different roles and different access needs.

2.3.4 PRIVACY IMPACT ASSESSMENT

In accordance with ENISA (37), in order to define the objectives of the system in terms of privacy a Privacy Impact Assessment (PIA) should be conducted. PIAs are required in certain situations by the GDPR which stipulates that its results should be taken into account in the privacy by-design process, cf. GDPR, Paragraph 1 of Article 23; "Data protection by design shall have particular regard to the entire lifecycle management of personal data from collection to processing to deletion, systematically focusing on comprehensive procedural safeguards regarding the accuracy, confidentiality, integrity, physical security and deletion of personal data. Where the controller has carried out a data protection impact assessment pursuant to Article 33, the results shall be taken into account when developing those measures and procedures."

LivAI sepie
Funded by the
Erasmus+ Programme
of the European Union

From a technical point of view, the core steps of a PIA are as follows:

- **the identification of stakeholders** and consulting of these stakeholders,
- **the identification of risks** (taking into account the perception of the stakeholders),
- **the identification of solutions** and formulation of recommendations,
- **the implementation** of the recommendations,
- **reviews**, audits and accountability measures.

The inputs of the privacy by design process per se should be the outputs of the second step (risk analysis) and third step (recommendations) and its output contributes to step 4 (implementation of the recommendations). Privacy by design is an iterative, continuous process and PIAs can be conducted at different stages of this process.

Additionally, it should be taken into account that the compliance approach implemented by carrying out a PIA is based on two pillars (38):

Fundamental rights and principles, which are "non-negotiable", established by law and which must be respected, regardless of the nature, severity and likelihood of risks,

Management of data subjects' privacy risks, which determines the appropriate technical and organisational controls to protect personal data.

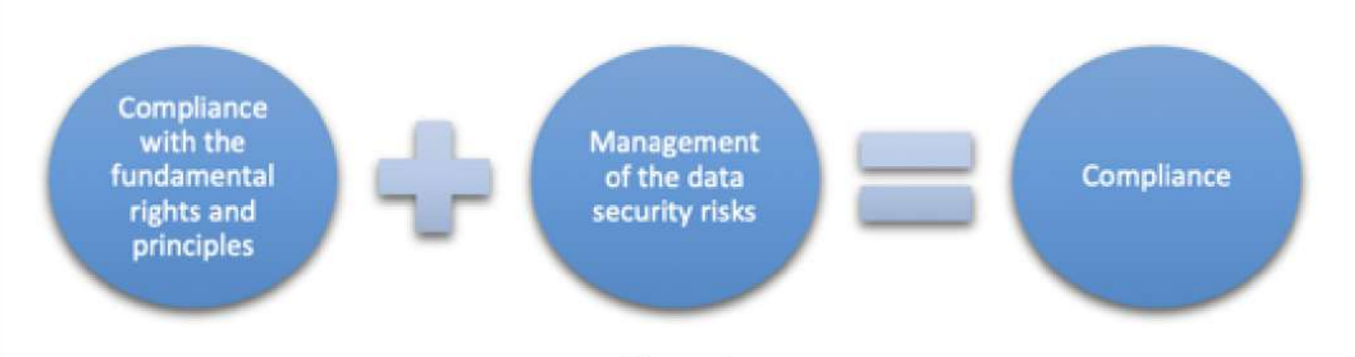

Compliance with the fundamental rights and principles
Management of the data security risks
Compliance

Figure 6
Compliance approach using a PIA (Source: Commission Nationale Informatique & Libertes (CNIL) Privacy Impact Assessment (PIA) Methodology)

Furthermore, to carry out a PIA it is necessary to (38):

- **Define and describe** the context of the processing of personal data under consideration,
- **Analyse** the controls guaranteeing compliance with the fundamental principles: the proportionality and necessity of processing, and the protection of data subjects' rights,
- **Assess privacy** risks associated with data security and ensure they are properly treated,
- **Formally document** the validation of the PIA in view of the previous facts to hand or decide to revise the previous steps.

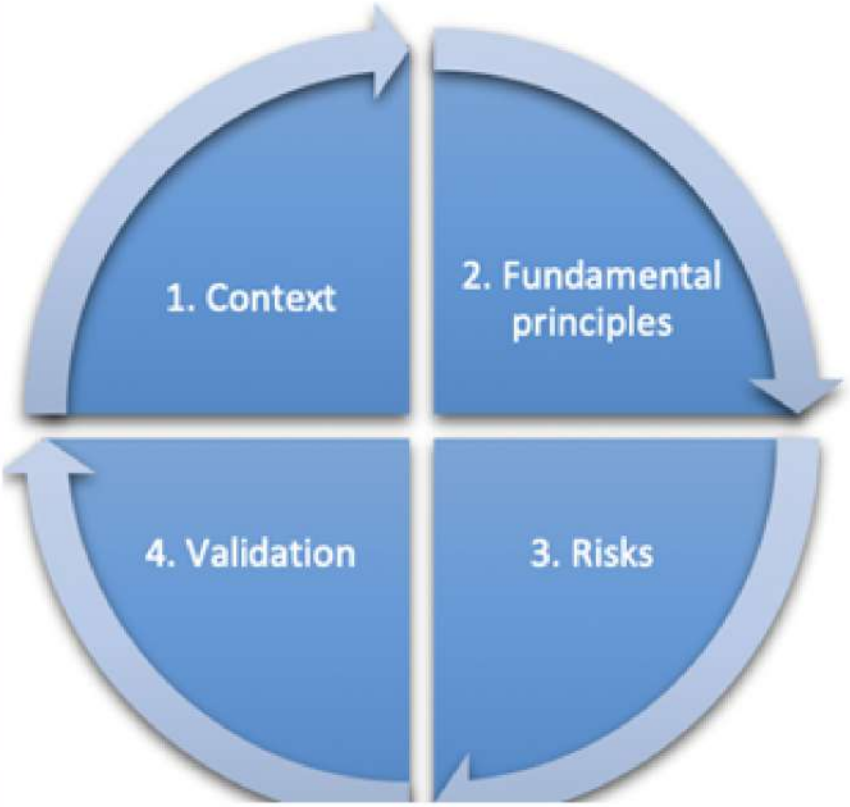


Figure 7
General approach for carrying out a PIA
(Source: Commission Nationale Informatique & Libertes (CNIL) Privacy Impact Assessment (PIA) Methodology)

2.3.5 DATA PROTECTION IMPACT ASSESSMENT (DPIA)

Based on the guidelines issued by the WP29 (39), Data Protection Impact Assessment (DPIA) is a process designed to describe the processing, assess its necessity and proportionality and help manage the risks to the rights and freedoms of natural persons resulting from the processing of personal data by assessing them and determining the measures to address them. DPIAs are important tools for accountability, as they help controllers not only to comply with the requirements of the GDPR, but also to demonstrate that appropriate measures have been taken to ensure compliance with the Regulation (as stated also in article 24 of GDPR). In other words, a DPIA is a process for building and demonstrating compliance.

Under the GDPR, non-compliance with DPIA requirements can lead to fines imposed by the competent supervisory authority. Failure to carry out a DPIA when the processing is subject to a DPIA (Article 35(1) and (3)-(4)), carrying out a DPIA in an incorrect way (Article 35(2) and (7) to (9)), or failing to consult the competent supervisory authority where required (Article 36(3)(e)), can result in an administrative fine of up to 10M€, or in the case of an undertaking, up to 2 % of the total worldwide annual turnover of the preceding financial year, whichever is higher.

ENISA also reports that (39):

- **DPIA is** one of the requirements introduced under the GDPR and can be also perceived as part of the ”protection by design and by default” approach.
- **Further to the emphasis** put by these principles on the engineering of data protection requirements into processing operations, such emphasis is also evident in Article 35 (7)(d) of the GDPR.
- **The legislator explicitly mentions** ”the measures envisaged to address the risks, including safeguards, security measures and mechanisms…” which clearly extends beyond the traditional deployment of technical and organisational measures and calls for a more detailed analysis, selection and operation of techniques able to ensure the required level of protection.

WP29 has also issued guidelines on how to carry out a DPIA as listed below (39):

AT WHAT MOMENT SHOULD A DPIA BE CARRIED OUT?

The DPIA should be carried out ”prior to the processing” (Articles 35(1) and 35(10), recitals 90 and 93). This is consistent with data protection by design and by default principles (Article 25 and Recital 78). The DPIA should be seen as a tool for helping decision-making concerning the processing.

The DPIA should be started as early as is practicable in the design of the processing operation even if some of the processing operations are still unknown. Updating the DPIA throughout the lifecycle project will ensure that data protection and privacy are considered and will encourage the creation of solutions that promote compliance. It can also be necessary to repeat individual steps of the assessment as the development process progresses because the selection of certain technical or organisational measures may affect the severity or likelihood of the risks posed by the processing.

The fact that the DPIA may need to be updated once the processing has actually started is not a valid reason for postponing or not carrying out a DPIA. The DPIA is an ongoing process, especially where a processing operation is dynamic and subject to ongoing change. Carrying out a DPIA is a continual process, not a one-time exercise.

WHO IS OBLIGED TO CARRY OUT THE DPIA? THE CONTROLLER, WITH THE DPO AND PROCESSORS.

The controller is responsible for ensuring that the DPIA is carried out (Article 35(2)). Carrying out the DPIA may be done by someone else, inside or outside the organisation, but the controller remains ultimately accountable for that task.

The controller must also seek the advice of the Data Protection Officer (DPO), where designated (Article 35(2)) and this advice, and the decisions taken by the controller, should be documented within the DPIA. The DPO should also monitor the performance of the DPIA (Article 39(1)(c)).

LivAI sepie Funded by the Erasmus+ Programme of the European Union

If the processing is wholly or partly performed by a data processor, the processor should assist the controller in carrying out the DPIA and provide any necessary information (in line with Article 28(3)(f)).

The controller must **"seek the views of data subjects or their representatives"** (Article 35(9)), "where appropriate". The WP29 considers that:

- those views could be sought through a variety of means, depending on the context (e.g. a generic study related to the purpose and means of the processing operation, a question to the staff representatives, or usual surveys sent to the data controller's future customers) ensuring that the controller has a lawful basis for processing any personal data involved in seeking such views. Although it should be noted that consent to processing is obviously not a way for seeking the views of the data subjects;

- if the data controller's final decision differs from the views of the data subjects, its reasons for going ahead or not should be documented;

- the controller should also document its justification for not seeking the views of data subjects, if it decides that this is not appropriate, for example if doing so would compromise the confidentiality of companies' business plans, or would be disproportionate or impracticable.

LivAI sepie Funded by the Erasmus+ Programme of the European Union

Finally, it is good practice to define and document other specific roles and responsibilities, depending on internal policy, processes and rules, e.g.:

- **where specific business** units may propose to carry out a DPIA, those units should then provide input to the DPIA and should be involved in the DPIA validation process;
- **where appropriate**, it is recommended to seek the advice from independent experts of different professions (lawyers, IT experts, security experts, sociologists, ethics, etc.);
- **the roles and responsibilities** of the processors must be contractually defined; and the DPIA must be carried out with the processor's help, taking into account the nature of the processing and the information available to the processor (Article 28(3)(f));
- **the Chief Information Security Officer (CISO)**, if appointed, as well as the DPO, could suggest that the controller carries out a DPIA on a specific processing operation, and should help the stakeholders with the methodology, help to evaluate the quality of the risk assessment and whether the residual risk is acceptable, and to develop knowledge specific to the data controller context;
- **the Chief Information Security Officer (CISO)**, if appointed, and/or the IT department, should provide assistance to the controller, and could propose to carry out a DPIA on a specific processing operation, depending on security or operational needs.

LivAI sepie Funded by the Erasmus+ Programme of the European Union

WHAT IS THE METHODOLOGY TO CARRY OUT A DPIA? DIFFERENT METHODOLOGIES BUT COMMON CRITERIA.

The GDPR sets out the minimum features of a DPIA (Article 35(7), and recitals 84 and 90):

- ”a description of the envisaged processing operations and the purposes of the processing”;
- ”an assessment of the necessity and proportionality of the processing”;
- ”an assessment of the risks to the rights and freedoms of data subjects”;
- ”the measures envisaged to:
- ”address the risks”;
- ”demonstrate compliance with this Regulation”.

The following figure illustrates the generic iterative process for carrying out a DPIA:

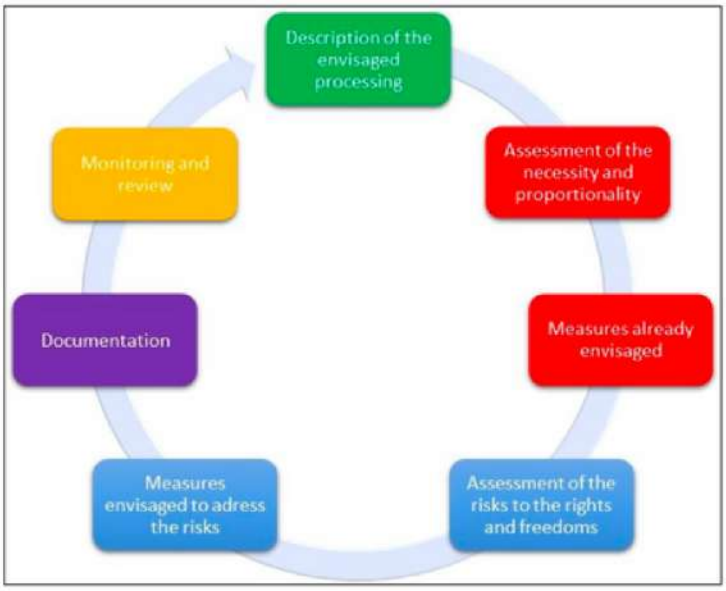

Description of the envisaged processing
Assessment of the necessity and proportionality
Measures already envisaged
Assessment of the risks to the rights and freedoms
Measures envisaged to adress the risks
Documentation
Monitoring and review

Figure 8
DPIA Generic iterative process
(Source: EU Commission. Guidelines on data protection impact assessment (DPIA) (wp248rev. 01)

2.4 ACCOUNTABILITY AND LIABILITY

In accordance with a report from EC (9), accountability for AI applications (and consequently Big Data enabled systems which utilise AI application) entails that the actors involved in their development or operation take responsibility for the way that these applications function and for the resulting consequences. Of course, accountability presupposes certain levels of transparency as well as oversight.

To be held to account, developers or operators of AI systems must be able to explain how and why a system exhibits particular characteristics or results in certain outcomes. Human oversight entails that human actors are able to understand, supervise and control the design and operation of the AI system. Accountability depends on oversight: To be able to take responsibility and act upon it, developers and operators of AI systems must understand and control the functioning and outcomes of the system. Hence, to ensure accountability, developers must be able to explain how and why a system exhibits particular characteristics.

The guidelines issued by EC (9) define a set of general ethical requirements:

- **It MUST** be documented how possible ethically and socially undesirable effects (e.g. discriminatory outcomes, lack of transparency) of the system will be detected, stopped, and prevented from reoccurring.

- **AI systems MUST allow human** oversight and control over the decision cycles and operation, unless compelling reasons can be provided which demonstrate such oversight is not required. Such a justification should explain how humans will be able to understand the decisions made by the system and what mechanisms will exist for humans to override them.

- **To a degree matching** the type of research being proposed (e.g. basic or precompetitive) and as appropriate, the research proposal should include an evaluation of the possible ethics risks related to the proposed AI system. This should also include the risk assessment procedures and the mitigation measures after deployment.

- **Whenever relevant, it should be considered** how end-users, data subjects and other third parties will be able to report complaints, ethical concerns, or adverse events and how these will be evaluated, addressed and communicated back to the concerned parties.

Funded by the Erasmus+ Programme of the European Union

- As a **general principle**, all AI systems should be auditable by independent third parties (e.g. the procedures and tools available under the XAI approach support best practice in this regard). This is not limited to auditing the decisions of the system itself, but also covers the procedures and tools used during the development process. Where relevant, the system should generate human accessible logs of the AI system's internal processes.

2.5 FAIRNESS

Based on the report published by EC (9), fairness entails that all people are entitled to the same fundamental rights and opportunities. This does not require identical outcomes, i.e., that people must have equal wealth or success in life. However, there should be no discrimination on the basis of the fundamental aspects of one's own identity which are inalienable and cannot be taken away. Various legislations already acknowledge a number of them, such as gender, race, age, sexual orientation, national origin, religion, health and disability.

According to the same report, informational fairness requires that the procedure was not designed in a way that disadvantages single individuals or groups specifically. On the other hand, substantive fairness entails that the AI does not foster discrimination patterns that unduly burden individuals and/or groups for their specific vulnerability. Fairness can also be supported by policies which promote diversity. These are policies that go beyond non-discrimination by positively valuing individual differences, including not only characteristics like gender and race, but also people's diverse personalities, experiences, cultural backgrounds, cognitive styles, and other variables that influence personal perspectives.

Supporting diversity means accommodating for these differences and supporting the diverse composition of teams and organisations.

In accordance with Article 5 of the GDPR, one of main principles that must be followed in processing data that directly affects the processing of Big Data and even before, the methods of collection and data retention is lawfulness, fairness and transparency (2). Based on the report from European Parliamentary Research Service (EPRS) (40), the two different concepts of fairness, namely the informational fairness and substantive fairness, are defined and can be distinguished in the GDPR.

Based on the same report by EPRS (40), informational fairness is strictly connected to the idea of transparency and requires that data subjects are not deceived or misled concerning the processing of their data, as is explicated in **Recital (60):**

> *"The principles of fair and transparent processing require that the data subject be informed of the existence of the processing operation and its purposes. The controller should provide the data subject with any further information necessary to ensure fair and transparent processing taking into account the specific circumstances and context in which the personal data are processed."*

Recital (60) also explicitly requires that information is provided on profiling:

> *"Furthermore, the data subject should be informed of the existence of profiling and the consequences of such profiling."*

Informational fairness is also linked to accountability, since it presumes that the information to be provided makes it possible to check for compliance. Informational fairness raises specific issues in connection with AI and Big Data, because of the complexity of the processing involved in AI- applications, the uncertainty of its outcome, and the multiplicity of its purposes. The new dimension of the principle pertains to the explicability of automated decisions, an idea that is explicitly affirmed in the GDPR. Arguably, the idea of transparency as explicability can be extended to automated inferences, even when a specific decision has not yet been adopted.

A specific aspect of transparency in the context of machine learning concerns access to data, in particular to the system's training set. Access to data may be needed to identify possible causes of unfairness resulting from inadequate or biassed data or training algorithms. This is particularly important when the learned algorithmic model is opaque, so that possible flaws cannot be detected through its inspection.

In the same report (40), it is stated that substantive fairness results from Recital (71) which points to a different dimension of fairness, i.e. what we may call substantive fairness, which concerns the fairness of the content of an automated inference or decision, under a combination of criteria, which may be summarised by referring to the standards of acceptability, relevance and reliability:

“In order to ensure fair and transparent processing in respect of the data subject, taking into account the specific circumstances and context in which the personal data are processed, the controller should use appropriate mathematical or statistical procedures for the profiling, implement technical and organisational measures appropriate to ensure, in particular, that factors which result in inaccuracies in personal data are corrected and the risk of errors is minimised, secure personal data in a manner that takes account of the potential risks involved for the interests and rights of the data subject and that prevents, inter alia, discriminatory effects on natural persons on the basis of racial or ethnic origin, political opinion, religion or beliefs, trade union membership, genetic or health status or sexual orientation, or that result in measures having such an effect.”.

THE GUIDELINES ISSUED BY EC (9) DEFINE A SET OF GENERAL ETHICAL REQUIREMENTS:

- **Avoidance of algorithmic bias**: AI systems should be designed to avoid bias in input data, modelling and algorithm design. Algorithmic bias is a specific concern which requires specific mitigation techniques. Research proposals MUST specify the steps which will be taken to ensure data about people is representative of the target population and reflects their diversity or is sufficiently neutral. Similarly, research proposals should explicitly document how bias in input data and in the algorithmic design, which could cause certain groups of people to be represented incorrectly or unfairly, will be identified and avoided. This necessitates considering the inferences drawn by the system which have the potential to unfairly exclude or in other ways disadvantage certain groups of people or single individuals;

- **Universal accessibility:** AI systems (whenever relevant) should be designed to be usable by different types of end-users with different abilities. Research proposals are encouraged to explain how this will be achieved, such as by compliance with relevant accessibility guidelines. To the extent possible, AI systems should avoid functional bias by offering the same level of functionality and benefits to end-users with different abilities, beliefs, preferences, and interests;

- **Fair impacts:** Possible negative social impacts on certain groups, including impacts other than those resulting from algorithmic bias or lack of universal accessibility, may occur in the short, medium and longer term especially if the AI is diverted from its original purpose. This MUST be mitigated. The AI system MUST ensure that it does not affect the interests of relevant groups in a negative way. Methods to identify and mitigate negative social impacts in the medium and longer term should be well documented in the research proposal.

AVOIDANCE OF ALGORITHMIC BIAS:

AI systems should be designed to avoid bias in input data, modelling and algorithm design. Algorithmic bias is a specific concern which requires specific mitigation techniques. Research proposals MUST specify the steps which will be taken to ensure data about people is representative of the target population and reflects their diversity or is sufficiently neutral. Similarly, research proposals should explicitly document how bias in input data and in the algorithmic design, which could cause certain groups of people to be represented incorrectly or unfairly, will be identified and avoided. This necessitates considering the inferences drawn by the system which have the potential to unfairly exclude or in other ways disadvantage certain groups of people or single individuals.

UNIVERSAL ACCESSIBILITY

AI systems (whenever relevant) should be designed to be usable by different types of end-users with different abilities. Research proposals are encouraged to explain how this will be achieved, such as by compliance with relevant accessibility guidelines. To the extent possible, AI systems should avoid functional bias by offering the same level of functionality and benefits to end-users with different abilities, beliefs, preferences, and interests,

LivAI sepie Funded by the Erasmus+ Programme of the European Union

FAIR IMPACTS

Possible negative social impacts on certain groups, including impacts other than those resulting from algorithmic bias or lack of universal accessibility, may occur in the short, medium and longer term especially if the AI is diverted from its original purpose. This MUST be mitigated. The AI system MUST ensure that it does not affect the interests of relevant groups in a negative way. Methods to identify and mitigate negative social impacts in the medium and longer term should be well documented in the research proposal.

BIBLIOGRAPHY

- Guidelines on personal data breach notification [Internet]. European Data Protection Supervisor. [cited 2023Mar20]. Available from: https://edps.europa.eu/data-protection/our-work/publications/guidelines/guidelines-personal-data-breach-notification_en
- Puaschunder JM. Big data ethics. Puaschunder, JM (2019). Journal of Applied Research in the Digital Economy. 2019 Apr 13;1:55-75.
- Da Bormida M. The Big Data World: Benefits, Threats and Ethical Challenges. InEthical Issues in Covert, Security and Surveillance Research 2021 Dec 9 (Vol. 8, pp. 71-91). Emerald Publishing Limited.
- Massimo AT. Opinion 5/2018-Preliminary Opinion on privacy by design.
- Hoepman JH. Privacy Design Strategies–extended abstract. ICT-System Security and Privacy Protection–29th IFIP TC. 2014 Jan;11:2-4.
- Seda Gürses, Carmela Troncoso, and Claudia Diaz. Engineering privacy by design. In Conference on Computers, Privacy & Data Protection (CPDP 2011), 2011.
- Bart Jacobs. Select before you collect. Ars Aequi, 54:1006–1009, December 2005.
- Andreas Pfitzmann and Marit Hansen. Anonymity, unlinkability, undetectability, unobserv- ability, pseudonymity, and identity management – a consolidated proposal for terminology (ver- sion v0.34 Aug. 10, 2010). http://dud.inf.tu-dresden.de/Anon_Terminology.shtml.
- Chaum DL. Untraceable electronic mail, return addresses, and digital pseudonyms. Communications of the ACM. 1981 Feb 1;24(2):84-90.
- Camenisch J, Lysyanskaya A. An efficient system for non-transferable anonymous credentials with optional anonymity revocation. InAdvances in Cryptology—EUROCRYPT 2001: International Conference on the Theory and Application of Cryptographic Techniques Innsbruck, Austria, May 6–10, 2001 Proceedings 20 2001 (pp. 93-118). Springer Berlin Heidelberg.
- Latanya Sweeney. k-anonymity: A model for protecting privacy. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 10(5):557–570, 2002.
- Dwork C. Differential privacy. InAutomata, Languages and Programming: 33rd International Colloquium, ICALP 2006, Venice, Italy, July 10-14, 2006, Proceedings, Part II 33 2006 (pp. 1-12). Springer Berlin Heidelberg.
- Platform for Privacy Preferences (P3P) project [Internet]. W3C. [cited 2023Mar20]. Available from: http://www.w3.org/P3P/
- Manset D. Big data and privacy fundamentals: toward a "digital skin". The Digitization of Healthcare: New Challenges and Opportunities. 2017:241-55.
- Mont MC, Pearson S. An adaptive privacy management system for data repositories. InTrust, Privacy, and Security in Digital Business: Second International Conference, TrustBus 2005, Copenhagen, Denmark, August 22-26, 2005. Proceedings 2 2005 (pp. 236-245). Springer Berlin Heidelberg.
- European Union Agency for Cybersecurity (ENISA) – Data Protection Engineering From theory to Practise, January 2022, Available from: https://www.enisa.europa.eu/publications/data-protection-engineering/@@download/fullReport
- Bincoletto G. EDPB Guidelines 4/2019 on Data Protection by Design and by Default. Eur. Data Prot. L. Rev.. 2020;6:574.

- EDPS. ”Guidelines on assessing the necessity and proportionality of measures that limit the right to data protection”. 25 February 2019. edps.europa.eu/sites/edp/files/publication/19-02-25_proportionality_guidelines_en.pdf
- EDPS. ”Assessing the necessity of measures that limit the fundamental right to the protection of personal data: A Toolkit” https://edps.europa.eu/data-protection/our-work/publications/papers/necessity- toolkit_en
- Article 29 Working Party. ”Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC”. WP 217, 9 April 2014. ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp217_en.pdf
- Danezis G, Domingo-Ferrer J, Hansen M, Hoepman JH, Metayer DL, Tirtea R, Schiffner S. Privacy and data protection by design-from policy to engineering. arXiv preprint arXiv:1501.03726. 2015 Jan 12.
- Commission Nationale Informatique & Libertes (CNIL) Privacy Impact Assessment (PIA) Methodology, February 2018
- EU Commission. Guidelines on data protection impact assessment (DPIA)(wp248rev. 01).
- EU GDPR Recitals https://gdpr-info.eu/recitals/ accessed on 25/06/2023

Practical Applicability to Adult Education

03

3.1 DEFINITION OF ADULT EDUCATION IN A DIGITAL PERSPECTIVE

In this fast-moving, ever-evolving digital era, the concept of learning knows no age limits. In the digital age, adult learning has taken on a new dimension. Education, which was once confined to youth and structured institutions, has expanded its horizons to include a new range of learners in the adult spectrum, breaking down socioeconomic barriers.The arrival of popular AI, data, and machine learning technologies has not only enhanced the accessibility of educational resources but has also paved the way for personalised and adaptive learning experiences.

Adult education refers to the deliberate and self-directed process where individuals, beyond their traditional schooling years, engage in educational activities to acquire new knowledge, skills, attitudes, or values. It is distinct from formal child education and encompasses a wide spectrum of learning, ranging from fundamental literacy and essential job skills to personal enrichment (1). Adult education provides a second chance for those who are poor in society or who have lost access to education for other reasons in order to achieve social justice and equal access to education.

Adult education serves an increasingly diverse student population, including individuals with disabilities, low-income and hard-to-serve adults, and those who have not earned high school diploma or equivalency.

Adult education has an important and different significance than classic youth education. The foundation of adult education is rooted in the belief that adults have the desire and capacity to learn, despite the busyness of life (1). To acquire new skills and shape attitudes for personal and professional growth. Adult education represents an acknowledgment that learning is a lifelong, ever-evolving process.

The New approaches on the digital world emphasise autonomy of adult students, where they take responsibility for their learning journey, with education techniques adapted to meet their specific needs and interests. The combination of a demanding job and study requires more efficient learning modules that make use of big data and artificial intelligence. By understanding students' strategies and needs, educators also benefit from this understanding better how to teach in challenging environments.

Funded by the Erasmus+ Programme of the European Union

One of the targets under the strategic framework for European cooperation and training (ET 2020) is that, at European level, an average of at least 15 % of adults should participate in lifelong learning by 2020.(2) The latest results from the European Union (EU) labour force survey show that in 2018 the participation rate in the EU stood at 11.1 %, 0.2 percentage points above the rate for 2017.(3) The rate has increased gradually since 2015, when it was 10.7 %. On average, across the EU in 2018 the participation rate for adult learning among women was higher (12.1 %) than the rate for men (10.1 %)(4).

In the EU Member States, the highest rates of adult participation in learning were in Sweden (29.2 %), Finland (28.5 %) and Denmark (23.5 %) (5).

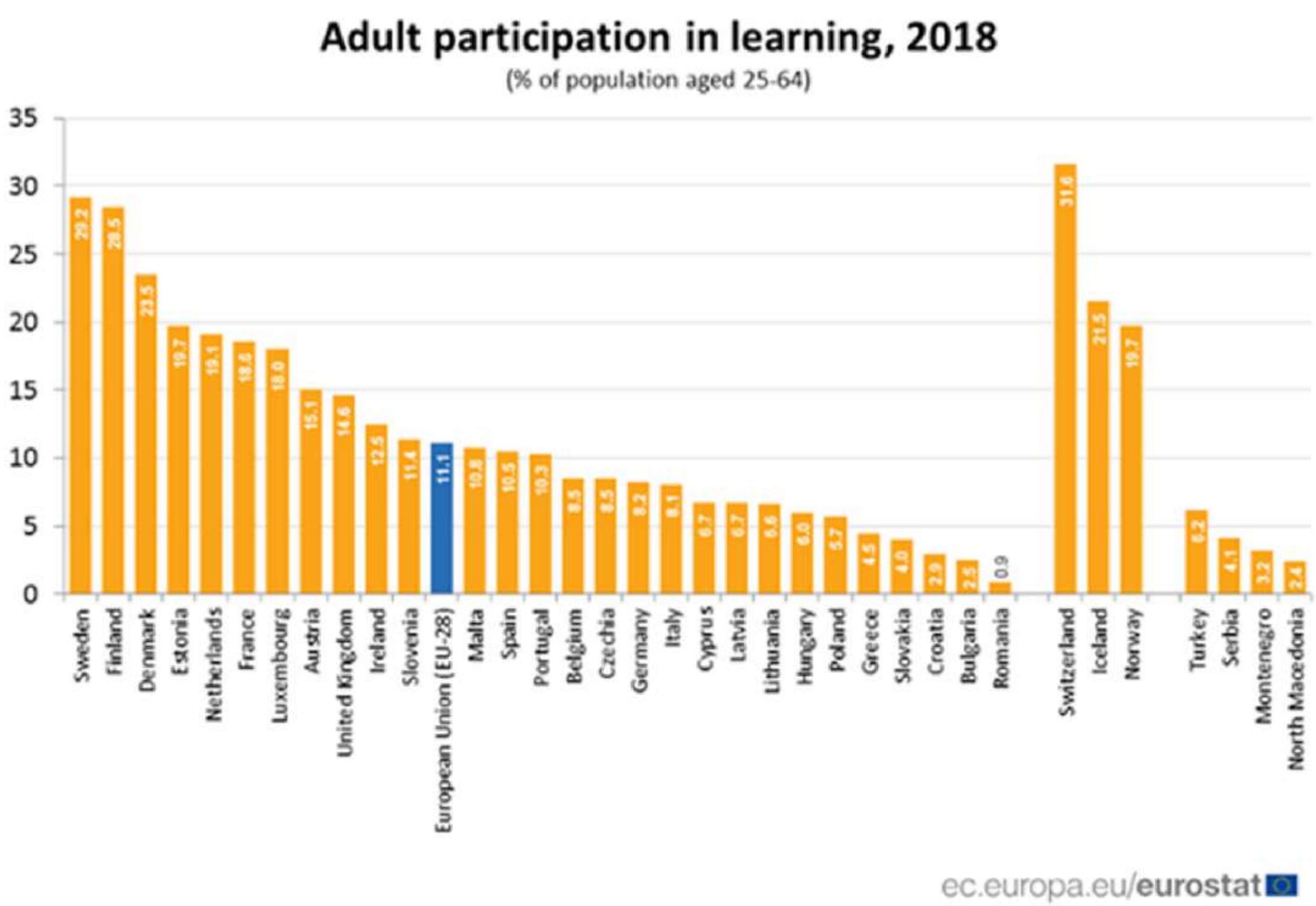


Figure 1: Eurostat

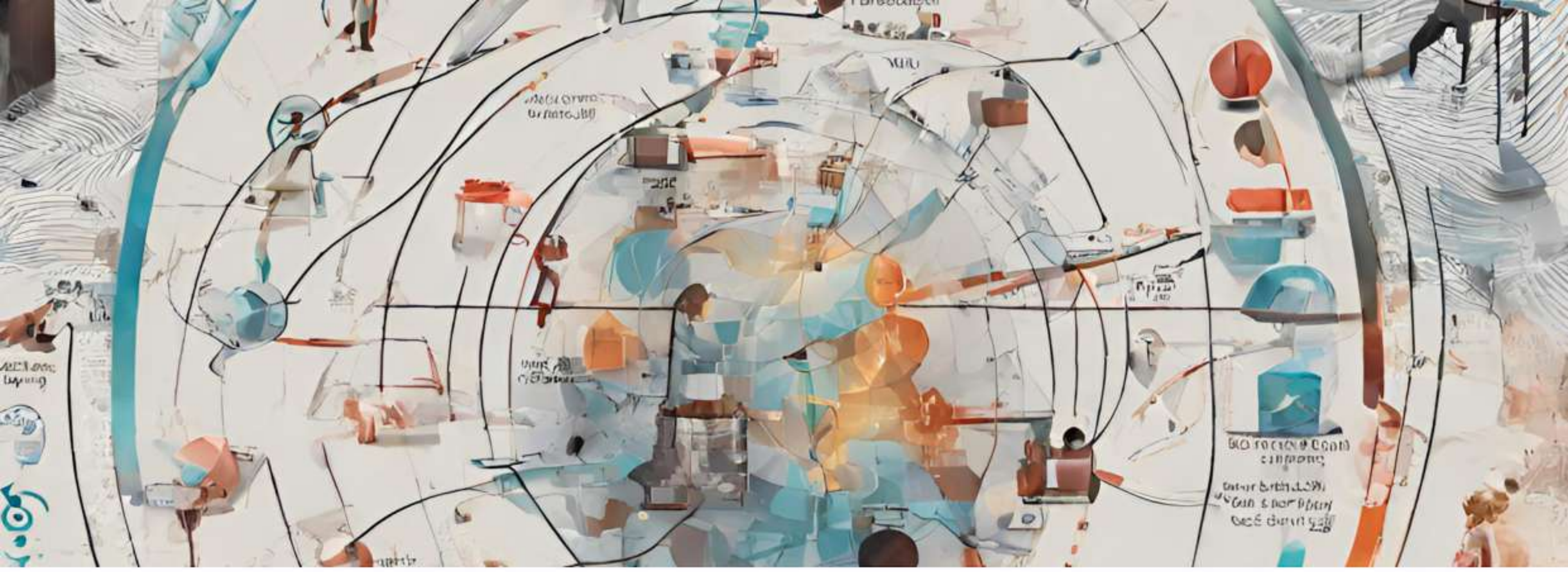
However, education is a rapidly developing field. In recent years, there has been an increasing focus on the digital perspective. With the growing importance of technology in our lives, the education of both young people and adults through digital tools has become increasingly relevant.

The digitalization of education will also help adult learners to develop digital literacy skills, which will enable them to participate fully in the online society and also to be fully adapted to the changes in the labour market. It will help adult learners to improve their employability and promotion.

Adult education through digital tools offer a great amount of flexibility; adult learners can learn online at their own speed and from anywhere, which makes education more accessible for those with familiar and job responsibilities.

However, in the absence of digital skills to take advantage of digital education, educational institutions must provide support and investment to reach a proper digital adult education.

Funded by the Erasmus+ Programme of the European Union



3.1.1 METHODOLOGY

The digitalisation of adult education has revolutionised the way we learn and has created opportunities for people to access education and training in a flexible and convenient manner. With the advancement of technology, learners are no longer limited by geographical location, time constraints or the availability of tutors.

Today, adult learners can choose from a wide range of methodologies that are designed to meet their individual needs and learning-cognitive styles, which are referred to the different ways in which people discover, store, transform and use the information they receive.

Digital tools and artificial intelligences have the ability to recognise, work with and adapt to different expressions of learning with subjects like:

- **Visualisation / verbalisation:** If the preferred mode of representation of ideas and concepts in the mind is visual (images) or if, on the contrary, it is verbal (words, sentences).

- **Focusing / sweeping:** Whether, when faced with a series of tasks, the person prefers to order them one after the other and not start one until the other has been completed, or whether he/she tends to work on all of them for short periods of time.

- **Concreteness / abstraction:** whether the person uses concrete experiences to learn something new or whether he/she prefers to deal with abstract ideas.

- **Independent / sensitive:** A person's tendency to assign his or her own organisation and structure to the information available to perform a task or solve a problem independently of the way it has been presented, in contrast, the tendency to solve the task or problem by handling the available information without detaching it from the context in which it has been presented and without changing its initial structure and organisation.

About the most exploited method of digital education, **e-Learning**. In a simple definition it is a method that integrates the use of digital technologies to deliver educational content and training programs online. This methodology can include online courses, interactive modules, webinars, and other forms of digital media and mobile apps. E-learning is a popular choice for adult learners because it is convenient, flexible, and accessible, stimulating learning from anywhere and anytime.

Blended Learning is a methodology which combines traditional face-to-face instruction with online learning, allowing a more immersive and interactive experience.

Similarly, a growing popularity of the **method of gamification** could be noted in recent years. This method arguably finds its core values in the non-formal learning methodology scheme, but with the integration of technology it grew into a very desired method in formal education as well. In a simple definition, gamification involves the use of game elements and mechanics to motivate learners and engage them in the learning process. This methodology can include gamified quizzes, simulations, and other interactive activities that allow learners to learn in a fun and engaging way.

LivAI sepie Funded by the Erasmus+ Programme of the European Union

Finally, social media and social networks arguably influenced another method called microlearning. Microlearning involves delivering content in small, bite-sized chunks that are designed to be easily digestible, which makes it ideal for adult learners who want to learn on-the-go, in short bursts of time. Microlearning can include short videos, infographics, and other visual aids. As an extension of it, Social Learning uses social media and platforms to facilitate and share knowledge, since it can include online discussion forums, peer-to-peer learning, and collaborative learning activities. That, it is an effective way to increase the engagement and collaboration among learners.

3.1.2 THE AI PERSPECTIVE

The integration of Artificial Intelligence (AI) in adult education is revolutionising the way we learn and teach. AI has the potential to enhance the effectiveness, efficiency and personalization of adult education.

It could be argued that the use of AI chatbots is a very hot topic in 2023, especially after the launch of ChatGPT. AI-powered chatbots are revolutionising the way we communicate and interact with technology. However, chatbots have a significant impact in education, due to their ability to provide instant support; they can be programmed to answer frequently asked questions and it allows them to give information and resources quickly at any time (6). Furthermore, chatbots can give feedback to learners' work, as well as gide learners to improve their skills or knowledge, recommending articles, videos and other material based on learners' interests.

From the other side, AI powered chatbots are challenged and limited to some extent. One of the primary challenges of AI-powered chatbots is that they lack the human touch. Chatbots are programmed to respond to specific queries, and they may not be able to understand the nuances of human language or emotions. According to a study by PwC, 59% of consumers prefer human interactions when dealing with customer service issues (7). This lack of human touch can be frustrating for learners who need more personalised support and assistance.

Chat GPT or the GPT model 3.5 managed to integrate human feedback in the process of training the AI model, and it could be argued that the human touch has been tackled with a strong innovation mechanism, but another challenge that is recognised even by the GPT 3.5 creators is that AI-powered chatbots are limited by their programming and may not understand complex questions or situations. They may provide inaccurate or incomplete information, leading to confusion and frustration for learners.

In a study by Sahoo and Sankaranarayanan, chatbots were found to be less accurate than human agents in answering complex queries (8). Also, AI-powered chatbots require maintenance and upkeep to ensure that they are functioning correctly, which can be time-consuming and expensive, especially for educational institutions with limited resources. They need to be regularly updated to keep up with the latest technology and to improve their functionality.

Moreover, AI-powered chatbots can pose a security and privacy risk since they may collect sensitive information from learners, such as their name, email address, and other personal information. If this information falls into the wrong hands, it can be used for malicious purposes. Educational institutions need to ensure that chatbots comply with data protection laws and that learners' information is stored securely.

Recent advances in AI have brought solutions to these questions, such as Bing's AI "Sydney". A more advanced chatbot that not only answers your questions on any given subject but also engages in human conversation, pretends to show emotions, concerns and ambitions. These developments also bring new ethical and philosophical considerations. Concerns about the replacement of professionals such as teachers, as well as their impact on students' education and cognitive development, are part of the debate.

In a deeper dimension, these new features bring risks not only to the digital security, economic and privacy of users, but to their psychological reach. A position of authority and great responsibility granted to a virtual teacher brings more relevant human security concerns when they acquire the ability to process non-verbal language and read human emotions. There is still much to be debated regarding functionalities and regarding what should be regulated (limited) for the safety of citizens (and students).

LivAI sepie Funded by the Erasmus+ Programme of the European Union

3.2 DATA PIPELINES IN THE FIELD OF ADULT EDUCATION

In recent years, the field of adult education has seen a significant increase in the use of data and analytics to improve the effectiveness of educational programs. The use of data pipelines has been particularly relevant in this context, as they allow for the collection, processing, and analysis of large amounts of data to inform decision-making and improve outcomes.

Data pipelines are sets of processes that move and transform data from various sources to a destination where new value can be derived(9). Data can be sourced from various origins, including APIs, SQL and NoSQL databases, files, and more. However, this data is often not immediately usable for analytical purposes. The responsibility for preparing this data typically falls on data scientists or data engineers, who are responsible for organising and structuring the data.

The raw data typically undergoes a series of processing steps: data ingestion, data transformation, and data storage. First, data is collected from various sources and ideally stored in a cloud data warehouse. Then, data is processed and organised into a usable format, ensuring consistency. Finally, once data has been refined and organised, it can be stored and accessed for various purposes (data analysis, visualisation, and machine learning tasks)(10).

In the context of adult education, **data pipelines** are used to collect data from a variety of sources, like student information systems, learning management systems, and assessment tools. This data is then processed and analysed to provide insights into learning engagement, student performance, and programme processes (11). That will provide real-time data to identify areas where additional support may be needed and facilitate decision-making.

Academic performance can be compared before and after a policy is implemented for improving student outcomes. For example, by linking administrative datasets, statisticians can regularly report trends to decision makers and identify classrooms or schools that perform better than others operating in similar contexts (12). Thus, regular monitoring can reveal changes over time, and evaluate the effectiveness of different educational strategies, providing feedback of the interventions.

In the last years, there has been an emergence of platforms and tools that boost the capabilities of Data pipelines, in terms of facilitating the visualisation, understanding data for decision-making, as well as for direct teaching entertaining techniques:

kidaptive

This platform allows teachers and organisations to view each student's progress, using machine learning algorithms to analyse learning materials and students, to create a complete overview of a student's progress and preferences. This will provide teachers with the basic data needed to understand how each student learns and to develop personalised approaches.

Quizlet

Is an online educational tool employs statistics and machine learning to make the most of user data and content, producing more efficient study techniques. One of its innovations is the Learning Assistant Platform, which identifies challenging terms for students and gives them higher priority in study sessions.

SMART SPARROW

Smart Sparrow is an adaptive courseware platform that uses AI to create personalised learning experiences for students. It allows educators to create custom course content and apps that adapt to individual student needs as they practise. The programme integrates ideas from different fields such as computer science, education, and psychology.

coursera

Is an online platform that uses new technologies to personalise course recommendations based on users interests, course history, and learning goals. It helps students to learn effectively in the long term, to concretise and strengthen their knowledge easily.

Gradescope is a platform that uses machine learning to automatically grade assignments, provide feedback to students, and detect plagiarism. It also offers AI-assisted grading tools with detailed stats for each question and mistake, helping assess how students are doing in specific areas.

Carnegie Learning it is an education service provider that offers maths, literacy, languages, and applied sciences, as well as high dose tutoring. It blends traditional classroom instructions with new techniques like video-streaming and gamification.

Like Carnegie Learning, it is also a software provider that uses AI-powered maths programs for students. It uses machine learning to compile data on factors like lessons completed and time spent per lesson, determining areas where students need extra support. Teachers can also reinforce their knowledge with self-paced professional courses.

Duolingo is alanguage learning platform that uses algorithms to personalise learning content for each student through micro-interval assessments. The result is an entertaining teaching method through the use of gamification.

SMART Learning Suite - Is a subscription-based software that combines various educational techniques to elevate self-learning through formative assessments, game-based learning, and student collaboration, both in and out of the classroom. Through the personalised feedback it fosters students to take ownership of their learning independently and as a group as they transition between classroom and distance learning.

Querium is an AI-powered maths and science learning platform that uses algorithms to personalise instruction for each student. It also uses AI to provide real-time feedback and assessment data to educators.

Cognii is a platform uses natural language processing and machine learning to provide instant feedback on students' written responses, helping them improve their writing skills.

While the benefits are immediate and numerous, the use of data pipelines in adult education also presents some challenges.

One of the biggest challenges is the inadequate security measures for ensuring data privacy. Educational institutions may lack sufficient resources, expertise and clear policies to implement robust data security measures that can respond to the use of innovative technologies. This can leave data vulnerable to hacking or human errors, causing students and school data to fall prey to fraud, theft, or political manipulation, and being vulnerable to advertising, among others.

In addition to the civil security problems involved in data flow and the consequential economic damages to be borne by the institution, we also have to consider the following issue. Leaks in student data can lead to stigmatisation due to their performance (13). Especially worrisome if the data ends up in teachers that can be easily biassed and do misinterpret the reading of data.

To prevent these outcomes and protect students, an effective personal data protection policy is necessary. In the following section, we will explore in depth the European data protection framework and the measures that an effective educational data protection policy must adopt.

3.3 PERSONAL DATA POLICIES

In the previous section we established in general terms the best practices of data pipelines, data analysis and security. We especially emphasised the importance of having a good data policy to structure operations. In this section we will emphasise and deepen the importance of a good data policy. Let's go deeper into data policies in the European Union and Education.

In a simple description, data policies are a set of rules and procedures that regulate the collection, processing, and use of personal data. According to the GDPR (General Data Protection Regulation) definition, ”Personal data means any information relating to an identified or identifiable natural person; [...] who can be identified [...] by reference such as a name, a number, location, an online identifier or more factors” (14). The purpose of personal data policies is to establish an operational governance structure that the organisation must follow in order to protect its users' data, to inform people how their information is used and to make sure that the organisation follows data protection laws.

These policies generally include obtaining consent from data subjects, safeguarding data integrity and security, allowing them access and managing their personal information. Data protection policies will strengthen the security of all information collected and stored by the organisation. It is a good protection against audits and data leakage by establishing strategies, data scope, legal protections and role assignment (15).

The General Data Protection Regulation carries relevant importance at the European Union level and stands as a reference for data protection worldwide. The European Union has taken a leading role in the development of personal data policies since the GDPR came into effect in 2018.

The GDPR is a European directive that encloses the regulations and processes governing the collection, processing, and utilisation of personal data within the European Union. The GDPR provides a framework for the establishment of personal data policies and includes provisions for obtaining informed consent, ensuring data accuracy and security, data breach notification, data portability, and providing individuals with the right to access and control their personal information (16). That means that, an effective DPP that follows the GDPR will allow data subjects to request a copy of their data or request to be deleted.

One of the key challenges in the development of personal data policies is the balance between privacy and the legitimate interests of organisations that collect and process personal data.

On the one hand, individuals have the right to privacy and control over their personal information. On the other hand, organisations need access to personal data to carry out their operations effectively.

In this regard, one of the key innovations of the GDPR is its extraterritorial scope, which means that the regulation applies to any organisation that collects or processes personal data of individuals within the EU, regardless of the organisation's location. This has led to a significant shift in the way that organisations collect and process personal data, as they must now comply with the GDPR's stringent requirements.

GDPR has introduced a number of new rights for individuals, including the right to be informed, the right of access, the right to erasure, and the right to object (16). These rights give individuals greater control over their personal data, and allow them to make informed decisions and security about how their data is used.Also, it has strengthened the accountability of organisations by requiring them to demonstrate compliance with the regulations. organisations must keep detailed records of data processing activities, appoint a Data Protection Officer (DPO), and implement appropriate technical and organisational measures to protect personal data. In the same direction, GDPR has introduced significant financial penalties for organisations that fail to comply with the regulations.

While GDPR has introduced some novel aspects to the field of personal data policies, it is important to recognize that many of its key principles were already established in earlier data protection laws. The European Union has had data protection laws in place since 1995, and the GDPR builds on this existing framework.

Furthermore, some critics argue that GDPR has created a burden for small and medium-sized businesses, who may struggle to meet the complex requirements of the regulation. This is because the cost of complying with the directive is high. When businesses decide to modify how they collect, share, and analyse data, it can result in substantial financial challenges and costs (17). Likewise, if the data that is collected must respond to the requirement of data minimisation, the quality of the services, products and results of the operations of an organisation will be of lower quality (17). Thus, there is a strong incentive for organisations to avoid collecting and processing personal data, which have negative consequences for innovation and economic growth.

Additionally, some experts have raised concerns that GDPR may not be effective in practice. While the regulation has been in force since 2018, there have been relatively few high-profile enforcement actions against organisations that have violated the regulations. "Data brokers are still stockpiling your information and selling it"(18). There has been no effective enforcement of the law, nor is its character strong enough to deter enforcement. Under this scenario, a considerable number of critical voices, officials both from within and outside the European Parliament have called for a reform of the directive.

The negative effects of this regulation take on a greater dimension when we consider its extraterritorial reach. The regulation applies to any business and organisation that requires the data of a European citizen, affecting companies and organisations in third countries. This leads to problems of legitimacy of the GDPR considering international and domestic rules (18). Likewise, it gives greater credibility to its critics as it increases the difficulty of implementing the regulation.

Special attention should be shown to educational institutions with access to adult data, due to the greater impact that fraud and identity theft can have on the financial dimension.

Thus, education organisations should ensure that:

- **The education plan** complies with relevant privacy **laws and regulations**. Being up to date with the General Data Protection Regulation (GDPR) is essential(19). Data collection plan and processing must be done in a transparent and ethical manner under a robust consent management process. This means that students and other stakeholders should be informed of data collection, how it will be used, and which third parties will have access to it . Data should be collected only with the consent of the individuals involved in a confidentiality agreement and only ”adequate, relevant and limited to what is necessary” (20).

- **Also, implementing strong data security** measures is important. That means provide an additional layer of security by using encryption techniques to protect sensitive data, restricting access to data only through authentication mechanisms, regularly monitoring systems or audits for potential threats, store hard copies backups (21), as well as using pseudonymization and anonymization for making it impossible to identify individuals (14). In the event that resources are not available to implement these measures. Private third parties could be used, provided that they meet these minimum criteria.

Likewise, data should be analysed by experts who know how to read the underlying reasons for student performance (paying special attention to socio-economic and socio-emotional layers) and respond appropriately to it. Reduction of data bias will work by ensuring data collection from a representative sample of the population; different backgrounds, ethnicities, and socio-economic status.Once personal information is no longer required for the intended purpose, organisations must take reasonable steps to destroy it. However, if the personal data is a ”Commonwealth record”, this requirement is not necessary (14).

- **In addition, organisations should have** a **Data Breach Response Plan,** which should include notifying affected individuals and regulatory authorities while trying to mitigate the impact as much as possible.

- **Finally, to promote confidence, training employees** on data privacy and security protocols must be considered (14). An effective curriculum should include potential risks, best practices for understanding and interpreting data and how to adapt to new technologies, as well as how to respond to data breaches.

3.4 COMPUTATIONAL THINKING AND GAMIFICATION IN ADULT EDUCATION

The confluence of gamification and computational thinking in education represents a powerful synergy for enhancing human cognition. There is a fascinating duality at play. Just as gamification can catalyse computational thinking, the reverse is equally true. Computational thinking offers the blueprint for designing gamified learning experiences that are not only entertaining but also intellectually enriching, with a purpose beyond teaching limited to one discipline or subject, but to foster interdisciplinary problem solving and innovation. Together, gamification and computational thinking blur the line between play and education, ushering in a whole new era in education.

3.4.1 COMPUTATIONAL THINKING

Computational thinking is a key skill set in today's digital age, involving the use of computer techniques and tools to solve problems and make decisions. This skill is not only important in the field of digital technology, but is also relevant in many other areas, such as science, engineering, finance and medicine. It is essential that students acquire computational thinking skills to be prepared for the challenges and opportunities of today's society.

Computational thinking is a problem-solving technique that involves reducing complex problems into smaller, more manageable parts and using algorithms or logical reasoning to solve those problems (22). It has special importance for computer scientists as it equips professionals to adeptly analyse data. However, beyond this specialised domain, it empowers students to cultivate problem-solving strategies. Likewise, It is an innovative approach that uses digital devices, robots, and computers to practically apply the use of technology to aid in learning.

For example, by using programming to control a robot's movements, utilising software to help on simulating scientific experiments, or using computers to analyse and visualise data in educational contexts.

It is considered an essential knowledge in the 21st century, which should be implemented in classrooms from a very early age, just as it is done with writing, arithmetic, reading, etc (24). Therefore, attention must also be paid to the adult sector so that they can adapt to the rapid changes that are taking place globally, an advance equivalent to the development of scientific thinking.

To continue learning basic computer and programming skills, students and teachers can use digital but equally easy to understand tools such as ”MIT Full STEAM Ahead”. This is an initiative by the Massachusetts Institute of Technology (MIT) designed to promote learning and engagement in science, technology, engineering, arts, and mathematics (STEAM) subjects. The program offers educational resources and hands-on activities in an all ages multidisciplinary approach.

THE MAIN SET SKILLS ASSOCIATED WITH COMPUTATIONAL THINKING ARE [25]:

- **Decomposition:** Breaking down data, processes, or problems into smaller, manageable parts

- **Pattern Recognition:** Observing patterns, trends, and regularities in data.

- **Pattern abstraction:** identifying the essential features of a problem and ignoring irrelevant details.

- **Algorithm Design:** Developing the step by step instructions for solving this and similar problems.

- **Evaluation:** Ensuring that your solution is a good one. For example, by Trial and error, testing different solutions to a problem until the best one is found.

LivAI sepie Funded by the Erasmus+ Programme of the European Union

These skills can be enhanced to find innovative solutions to different problems, through the development of critical thinking, teamwork and the application of computational thinking skills in different contexts and different disciplines.

For adult education, computational thinking will be especially helpful for providing students with the necessary tools to apply the skill set in their day-to-day professional lives. Through practical teaching strategies, by asking the right questions to help find solutions, through group discussions, etc., for information management learning, mathematical problem solving or the automation of everyday tasks. It is in its nature to integrate concepts and techniques from different disciplines simultaneously, encouraging lateral thinking. As well as other advantages such as helping students to ”make powerful connections between the subjects”[25], to facilitate the learning of concepts and to encourage innovation.

When teaching computational thinking, it is advisable to start with ’unplugged’ activities that can facilitate and enhance its learning. On the one hand, to provide a conceptual approach for students unfamiliar with programming and, on the other hand, to develop computational skills in students without access to technology. Hands-on activities without the use of digital tools can teach concepts such as ”algorithms, decomposition and pattern recognition” without long exposure to screens (24). There are different examples such as the ”Three iterative steps process” and the ”Binary numbers activity”.

THE THREE ITERATIVE STEPS PROCESS [26]:

- **Abstraction**: In this first phase we Identify a problem and we break it down into smaller parts. An example could be to design a program that calculates the area of a rectangle.

- **Automation**: In the second phase we design the set of instructions that solves the problem and can be executed by technological tools.In our previous example, we would write a program that could calculate the area with the length and width of a rectangle.

- **Analysis**: In the last stage we test the program and evaluate its effectiveness. So, we could test the program with different measures for the length and width and see if it produces the correct output.

- **The Binary numbers activity**: Involves using cards with dots to teach binary numbers functioning and consist of performing arithmetic operations on them. What is relevant about this game is that it does not require the use of computers and does not require a technical language.

LivAI sepie
Funded by the
Erasmus+ Programme
of the European Union

To continue learning basic computer and programming skills, students and teachers can use digital but equally easy to understand tools such as "**Scratch**" or "**MIT Full STEAM Ahead**". This is an initiative by the Massachusetts Institute of Technology (MIT) designed to promote learning and engagement in science, technology, engineering, arts, and mathematics (STEAM) subjects. The program offers educational resources and hands-on activities in an all ages multidisciplinary approach.

3.4.2 CONNECTION BETWEEN DIGITAL LITERACY AND COMPUTATIONAL THINKING

Digital literacy refers to a person's ability to use digital technologies effectively, understand how they work and evaluate their relevance to a given task. It is a combination of both technical and cognitive abilities that provides people with concepts and training to process data and transform them into information, knowledge, and decisions.

For adults, often disoriented as they find themselves in the midst of rapid technological change, it is important that they quickly learn how to work with the new digital tools in order to exploit them professionally and in their daily lives. Even more important is that they acquire the skills to be able to adapt to new changes without the help of others. This being the case, with digital literacy skills students discern misinformation and have the ability to identify authentic material for both personal and professional research; they will know how to access a wide variety of resources properly, and they will improve writing, reading, listening and speaking skills for online communication and teamwork. Students will find comfort in using technology more often for work.

Thus, the main pillars of digital literacy are: information and data literacy, communication and collaboration, digital content creation; and problem solving [27].

More specifically, digital literacy includes understanding basic web browsing (search engines, bookmarks...), recognizing online threads, email management, social networking and digital communication, computer applications like Word, data knowledge, file management, critical media evaluation, online collaborative work, online research on databases, and e-commerce skills, among others.

The connection between digital literacy and computational thinking is close, as computational thinking is an important component of digital literacy. Moreover, computational thinking is a new form of literacy per se that requires specific social, cognitive, and material features [28] that distinguish it from other types of literacy. That means computational thinking emphasises the importance of integrating computational skills into traditional literacy practices [28].

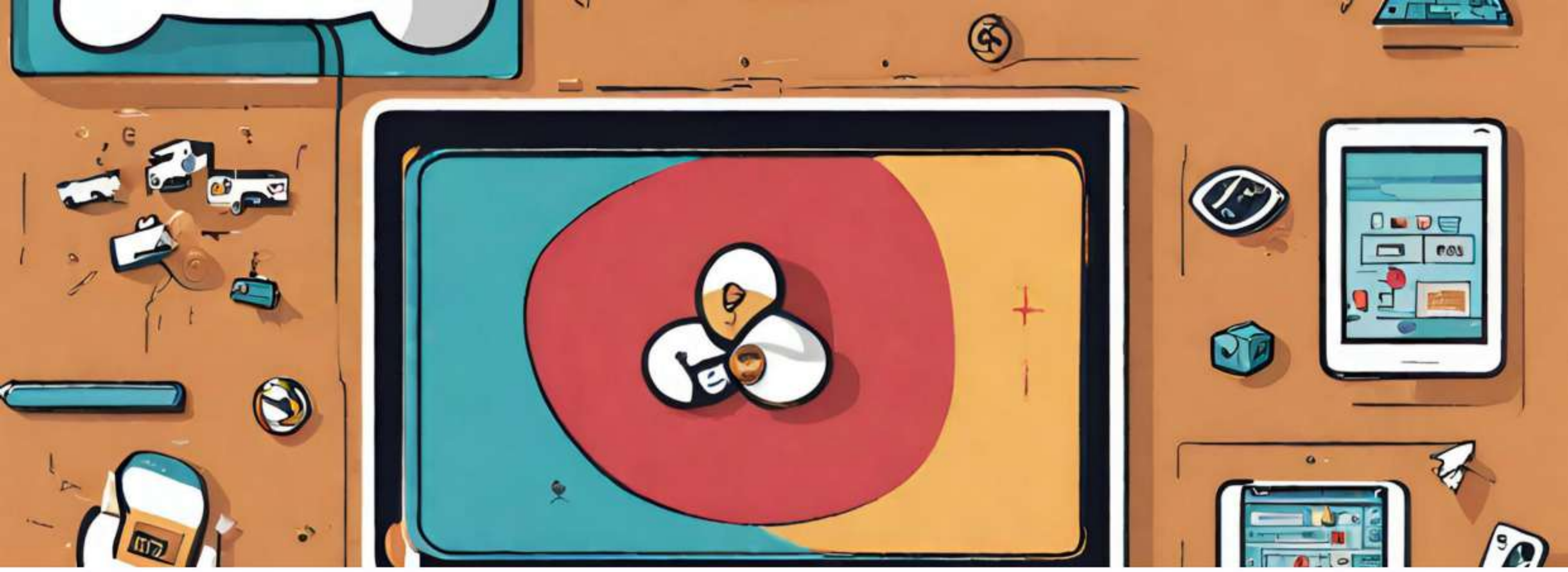
Therefore, digital literacy can help people understand how digital technologies work on a day-to-day basis and computational thinking skills can help people become more proficient in using digital technologies for complex problem solving.

DONE

←

TO DO

→



3.4.3 GAMIFICATION IN ADULT EDUCATION

Gamification in adult education is an innovative technique that uses game elements to motivate and engage adult learners in the learning process. More specifically, gamification is the process of incorporating game elements and mechanics into non-game based learning environments to increase engagement, motivation, and learning outcomes [29]. This technique has been proven to be very effective in adult education, especially when dealing with rapidly developing areas of knowledge that adults cannot keep up with.

The key idea behind gamification is to create a learning environment that is playful, participatory, and dynamic, which improves their knowledge retention and helps them to acquire practical skills. Integrating elements such as point systems, leaderboards, badges, progress metres, levels or other traditionally related to games into ”conventional” learning activities.[30]

This method can be used in a variety of educational contexts, from vocational and technical training to continuing education.

It goes beyond traditional teaching methods by incorporating challenges, rewards, and competitions into the learning process. This not only motivates adult learners but also encourages their active participation. [29]

One of the major advantages of gamification in adult education is its ability to enhance learners' motivation. By infusing game elements into the curriculum, adults feel more connected and engaged with the subject matter. This heightened engagement helps them maintain their interest and motivation. In addition to motivation, gamification also improves knowledge retention by actively involving students in the learning process. It develops cognitive skills that enable adults to respond imaginatively and improvisationally to real life and work problems. This means an increase in response efficiency and learning efficiency compared to traditional methods of rote learning and memorisation. Moreover, research studies have emphasized the capacity of games to spark voluntary knowledge acquisition.

Gamification also encourages collaboration and teamwork for solving problems and achieving common goals. This improves communication, listening, negotiation skills and group decision making in challenging situations. By using gamification techniques, students can learn in a more relaxed and fun way, which helps them to reduce the stress and anxiety associated with studying [31]. Teachers, school administrators, and corporate trainers are catching on to the value of gamification in their teaching and training efforts. As learning and training have moved from physical classrooms to online settings, teachers are getting creative to make the online experience more engaging for students. [32].

Particularly in the realm of corporate training. Numerous companies, organisations, and educational institutions are increasingly recognizing the significance of gamification techniques.

When we are talking about learning tools based on gamification, a basic learning platform for teachers is Kahoot. It is a free platform that allows teachers to create and share customised quizzes, surveys, and discussions with their students. Is easy to use and requires no special training or technical skills.

One of the challenges facing the industry, with training professionals and gamification experts at the forefront, is to break down the current barriers that slow down the expansion of new learning techniques. Among these, bridging the digital divide is a priority.

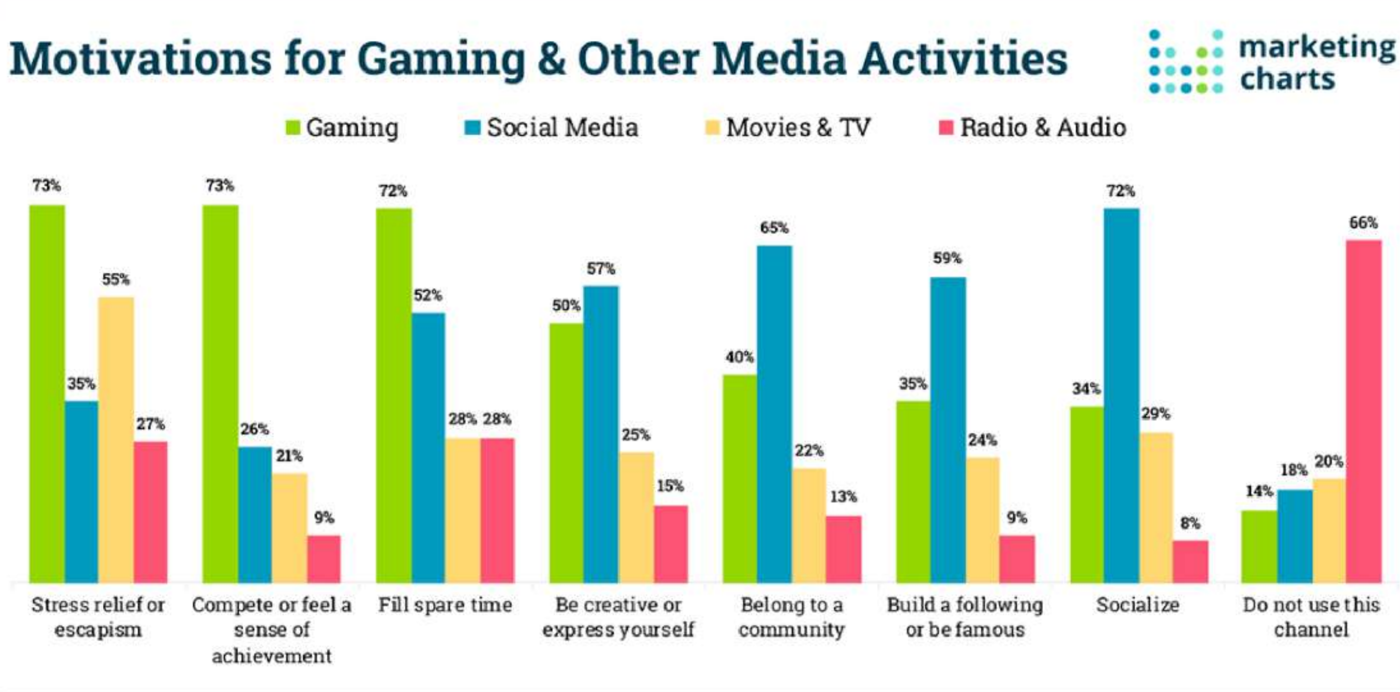
So is the change in the mentality of those responsible for implementing training actions, so that they place all their trust in the indisputable power of games for learning, especially those designed to develop 21st century skills.

Learning is very important for people's personal and professional development. Understanding the content is crucial in that learning process. Only if we feel that we understand what we are learning, will we be truly motivated and committed to achieving the best results.

In order to adapt the method of gamification in adult learning, academic institutions should focus on adult learners' specifics and challenges when designing gamified learning experiences. Also, integrate gamification into existing learning processes, complementing traditional content with gaming to make a more impactful sense of knowledge. Moreover, gamification should be done in a way that is meaningful and respectful of adult learners' time and experience.

Combining gamified learning with data analysis to track how students have interacted with these tools, how much they have learnt and how easy it is for them to acquire knowledge will help educational institutions to implement effective and rapid adult learning strategies.

By combining games that enhance the learning of computational thinking, we will see an exponential growth in the number of adults who are able to use the new tools. We will see that digital creation is not limited to the very young, but adults will also be able to have a greater presence in the development of content and applications on the web.


Motivations for Gaming & Other Media Activities

Published on MarketingCharts.com in March 2022 | Data Source: dentsu

Based on a survey of 1,000 US adults (18+)

LivAI sepie Funded by the Erasmus+ Programme of the European Union

BIBLIOGRAPHY

- Definition of ”Adult education”? [Internet]. LINCS Community | Adult Education and Literacy | U.S. Department of Education. Available from: https://community.lincs.ed.gov/group/111/discussion/definition-adult-education
- Strategic Framework [Internet]. European Education Area. Available from: https://education.ec.europa.eu/about-eea/strategic-framework
- CEICdata.com. European Union Labour Force participation rate [Internet]. 2018. Available from: https://www.ceicdata.com/en/indicator/european-union/labour-force-participation-rate
- Eurostat. 11.1 % of adults participate in lifelong learning. Eurostat [Internet]. 2019 May 17; Available from: https://ec.europa.eu/eurostat/web/products-eurostat-news/-/DDN-20190517-1
- Eurostat. 11.1 % of adults participate in lifelong learning. Eurostat [Internet]. 2019 May 17; Available from: https://ec.europa.eu/eurostat/web/products-eurostat-news/product/-/asset_publisher/VWJkHuaYvLIN/content/DDN-20190517-1/pop_up
- Mallow J. ChatGPT For Students: How AI Chatbots Are Revolutionizing Education [Internet]. eLearning Industry. 2023. Available from: https://elearningindustry.com/chatgpt-for-students-how-ai-chatbots-are-revolutionizing-education
- PricewaterhouseCoopers. Building customer loyalty and retention - PwC Customer Loyalty Survey 2023 [Internet]. PwC. Available from: https://www.pwc.com/us/en/services/consulting/business-transformation/library/building-customer-loyalty-guide.html
- Kirkner RM. Chatbots as accurate as ophthalmologists in giving advice. Medscape [Internet]. 2023 Sep 9; Available from: https://www.medscape.com/viewarticle/995856
- Densmore J., 2021. Data Pipelines Pocket Reference: Moving and Processing Data for Analytics. O’Reilly Media, Inc. 2021 March: 978-1-492-08783-0
- What is a data pipeline | IBM [Internet]. Available from: https://www.ibm.com/topics/data-pipeline
- Sharma, Kshitij & Papamitsiou, Zacharoula & Giannakos, Michail. (2019). Building Pipelines for Educational Data using AI and Multimodal Analytics: a ”grey-box” approach. British Journal of Educational Technology. 50. 10.1111/bjet.12854.
- Global Partnership for Education. How to unleash the power of data to transform education policies | Global Partnership for Education [Internet]. Global Partnership for Education. Available from: https://www.globalpartnership.org/blog/how-unleash-power-data-transform-education-policies
- Australian Government. (2018, June 5). Guide to securing personal information. OAIC. https://www.oaic.gov.au/engage-with-us/consultations/personal-information/guide-to-securing-personal-information-update
- Art. 4 GDPR – Definitions - General Data Protection Regulation (GDPR) [Internet]. General Data Protection Regulation (GDPR). 2018. Available from: https://gdpr-info.eu/art-4-gdpr/

- Data protection Policy: key elements to include & best practices [Internet]. Cloudian. 2023. Available from: https://cloudian.com/guides/data-protection/data-protection-policy-9-things-to-include-and-3-best-practices/amp/
- Keeping Up with Data Protection Regulations | Cloudian [Internet]. Cloudian. 2023. Available from: https://cloudian.com/guides/data-protection/data-protection-regulations/amp/#2
- A case against the General Data Protection Regulation | Brookings [Internet]. Brookings. 2022. Available from: https://www.brookings.edu/articles/a-case-against-the-general-data-protection-regulation/
- Burgess M. How GDPR is failing. WIRED [Internet]. 2022 May 23; Available from: https://www.wired.com/story/gdpr-2022/
- Azzi A. The challenges faced by the extraterritorial scope of the General Data Protection Regulation [Internet]. 2018. Available from: https://www.jipitec.eu/issues/jipitec-9-2-2018/4723
- Ashbel A. A beginner's guide to data privacy laws and compliance in the education industry [Internet]. 2020. Available from: https://bluexp.netapp.com/blog/ccs-blg-data-privacy-laws-and-compliance-in-the-education-industry
- General Data Protection Regulation (GDPR) – Official Legal text [Internet]. General Data Protection Regulation (GDPR). 2022. Available from: https://gdpr-info.eu/
- Arrington K. Computational Thinking: Essential Tips & Learning Resources to Master Algorithms-Authentic Jobs [Internet]. Authentic Jobs. 2023. Available from: https://authenticjobs.com/computational-thinking-tips-resources-algorithms/
- Zamansky M. How an unplugged approach to computational thinking can move schools to computer science. EdSurge [Internet]. 2019 Dec 19; Available from: https://www.edsurge.com/news/2019-12-19-how-an-unplugged-approach-to-computational-thinking-can-move-schools-to-computer-science
- Hunsaker E. Computational thinking [Internet]. Cc_By. 2020. Available from: https://edtechbooks.org/k12handbook/computational_thinking
- Repenning, Alexander (4 September 2016). ”Computational Thinking Tools”. IEEE Symposium on Visual Languages and Human-Centric Computing. Retrieved 7 April 2021.
- University of San Diego - Professional & Continuing Education. A Teacher's Guide to Digital Literacy & Digital Literacy Skills in the classroom [Internet]. University of San Diego - Professional & Continuing Education. 2023. Available from: https://pce.sandiego.edu/digital-literacy/#What-is-Digital-Literacy
- Jacob S, Warschauer M. Computational thinking and literacy. Journal of Computer Science Integration [Internet]. 2018 Aug 24;1(1). Available from: https://doi.org/10.26716/jcsi.2018.01.1.1
- Hogle PS. Use gamification to boost motivation for learning [Internet]. Available from: https://www.ottolearn.com/post/136-use-gamification-to-boost-motivation-for-learning

- Justas. Gamification and Adult Education - DNS The Necessary Teacher Training College DNS The Necessary Teacher Training College [Internet]. DNS the Necessary Teacher Training College. 2022. Available from: https://www.dns-tvind.dk/gamification-and-adult-education/
- Understanding the ROI of gamification in adult learning [Internet]. Available from: https://www.agilemeridian.com/blog/understanding-the-roi-of-gamification-in-adult-learning-2
- Katriina. Playing for adults – five examples of game-based learning tools - Elm [Internet]. Elm. 2021. Available from: https://elmmagazine.eu/future-of-adult-education/playing-for-adults-five-examples-of-game-based-learning-tools/

04

Ethical System Frameworks and Open Source Platforms

that are ready to use

4.1 OPEN-SOURCE DEFINITION

Open-source does not just mean access to the source code. The distribution terms of open-source software must comply with the following criteria:

i. Free redistribution

The licence shall not restrict any party from selling or giving away the software as a component of an aggregate software distribution, containing programs from several different sources. The licence shall not require a royalty or other fee for such sale.

ii. Source Code

The program must include source code and must allow distribution in source code as well as in compiled form. Where this is not the case though, there must be a well-publicised means of obtaining the source code for not more that a reasonable reproduction cost, indicatively via simple downloading from the internet.

iii. Derived works

The licence must allow modifications and derived works and must allow them to be distributed under the same terms as provided for in the licence of the original software.

iv. Integrity of the Author's Source Code

The licence may restrict source-code from being distributed in modified form only if the licence allows the distribution of "patch files" with the source code for the purpose of modifying the program at build time. Software built from modified source code must be directly provided for by the licence. What is more, different versions of the software must carry different numbers accordingly.

v. No discrimination against persons and groups or fields of endeavour.

vi. Distribution of licence

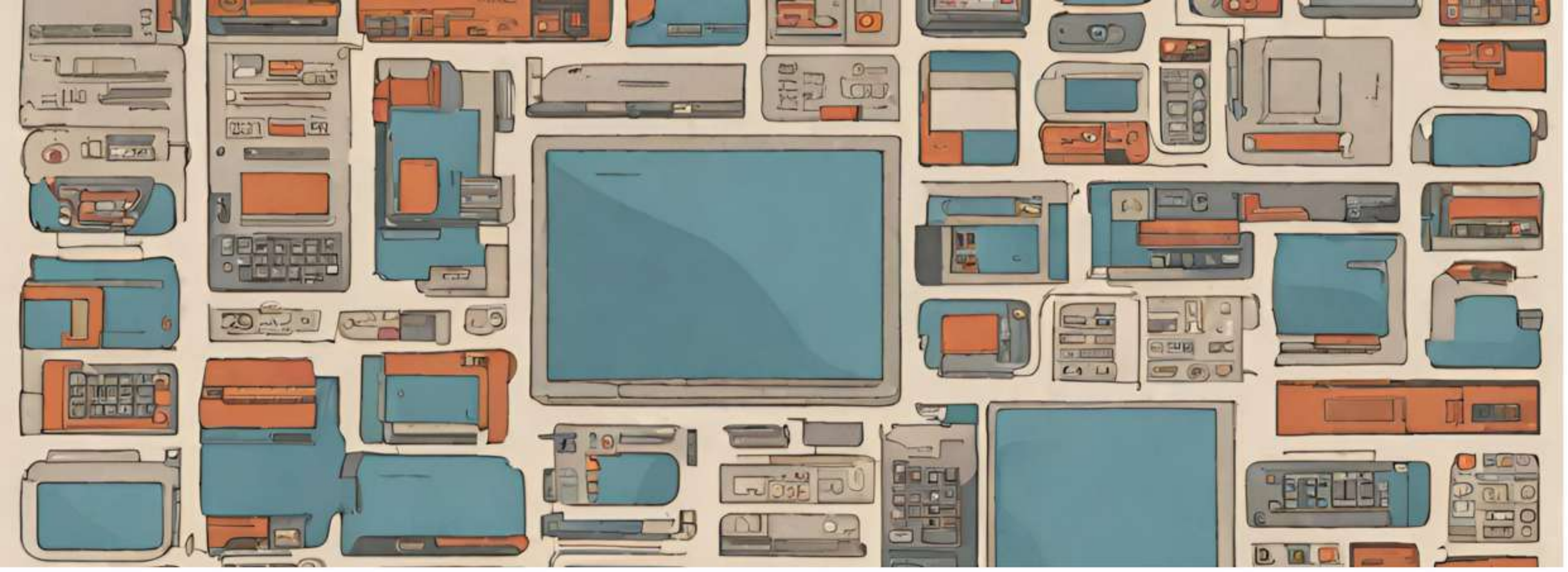
The rights attached to the program must apply to all to whom the program is redistributed without the need of a new licence.

vii. Licence must not be specific to a product

The rights attached to the program must not depend on the program being part of a particular program distribution.

viii. The licence must not place restrictions on other software that is distributed along with the original software.

4.2 OPEN SOURCE AND ACCOUNTABILITY



4.2.1 THE ROLE OF OPEN SOURCE SOFTWARE IN ADDRESSING ACCOUNTABILITY BARRIERS IN COMPUTING

In her article entitled Computing and Accountability, Helen Nissenbaum cites four barriers to accountability:

1. The problem of many hands,

2. Bugs,

3. Computer as scapegoat and

4. Ownership without liability.

She asserts that these barriers can lead to ”harm and risks for which no one is answerable and about which nothing is done” (Nissenbaum, 1994). We will examine how OSS may have addressed barriers 1 and 2. Number 3 is a general issue and number 4 does not apply because there is no software ownership per se in open source. In open source, if a developer were to write irresponsible code, others contributing to the open source software would be unlikely to accept it. So, in this case, there is built-in individual accountability.

If a developer were part of a large company, where all programming parts contribute to a large commercial venture, it then would fall on both the company and the individual to accept responsibility for the problematic software product. Often this is not done. So the many hands problem referred to by Nissenbaum in Computing and Accountability can be reduced in OSS because parts of code can be ascribed to various developers, and their peers hold them accountable for their contributions.

Nissenbaum argues that accepting bugs as a software fact of life has issues regarding accountability (Nissenbaum, 1994). The open source approach to software development treats the bug problem with a group effort to detect and fix problems. Torvalds states, ”given enough eyeballs, all bugs are shallow” (Raymond, 2001, p. 315). The person that finds a bug in OSS may not be the person to fix it. Since many adept developers examine OSS code, bugs are found and corrected more quickly than in a development effort where only a few developers see the code. In this group effort, accountability is not lost in the group, but is instead taken up by the entire group. The question of whether this group accountability is as effective as individual responsibility is, again, empirical.

The examples of Apache and Linux (Webcab solutions, 2003) offer at least anecdotal evidence that some OSS demonstrate high reliability. Don Gotterbarn is also concerned about issues of professional accountability in OSS (Wolf et al, 2002).

In addition to worries about sufficient care in programming and maintaining OSS, Gotterbarn points out that an OSS licensing agreement forces the authors of the software to relinquish control of the software. If someone puts OSS to a morally objectionable use, then the developers have no right to withdraw the software from that use. Gotterbarn's objection has some theoretical interest, for the OSS licensing agreements clearly state that no one who follows the OSS rules can be blocked from using the software. But if we accept the idea that software developers have a moral duty to police the use of the software they distribute, especially when the software is utility software, we fall into a practical and theoretical thicket.

How is a vendor to know the eventual use of software, especially when the software is utility software (such as an operating system or a graphics package)?

Are software developers empowered to judge the ethics of each customer or prospective customer?

These responsibilities are overreaching ethically, and far too ambitious in a practical sense. Furthermore, the relinquishment of control argument has practical significance only if existing competing software models include effective control over the use of software. (That is, should OSS be held to a higher standard than commercial software in relation to ethical responsibility for downstream use?)

LivAI sepie Funded by the Erasmus+ Programme of the European Union

4.3 ETHICAL TOOLS AND OPEN-SOURCE PLATFORMS

As AI continues to advance and become increasingly integrated into our daily lives, concerns around the ethical implications of these technologies have grown. There is a growing recognition that the development and deployment of AI systems need to be guided by ethical principles to ensure that they are safe, transparent, accountable, and aligned with human values. This has led to the development of various tools and open platforms that can help organisations and individuals create and implement ethical frameworks for AI.

The development of ethical frameworks for AI is crucial as it provides guidance on how to build and deploy AI systems that are trustworthy, safe, and respectful of fundamental human rights. Ethical frameworks can help to prevent unintended negative consequences, such as biassed or discriminatory outcomes, and provide a clear understanding of the ethical implications of AI systems. They can also help to build trust and legitimacy with stakeholders, including users, customers, and regulatory bodies.

Tools and open platforms have been developed to support the development and implementation of ethical frameworks for AI. These tools range from checklists and guidelines to more advanced frameworks that incorporate machine learning algorithms to help identify and mitigate ethical risks in AI systems. Open platforms provide a collaborative environment where individuals and organisations can share knowledge, resources, and best practices around AI ethics.

One of the key benefits of using tools and open platforms for AI ethical frameworks is that they provide a structured approach to addressing ethical issues in AI. They help organisations to systematically identify and address potential ethical risks associated with AI systems, while also providing a framework for ongoing monitoring and evaluation. Additionally, they can help to reduce the burden on individual organisations to develop their own ethical frameworks from scratch and provide access to shared resources and expertise.

4.3.1 ETHICAL OS

Ethical OS (https://ethicalos.org/) is a valuable toolkit for organisations that are developing emerging technologies such as AI. It is designed to help organisations anticipate and address ethical risks that may arise during the development process. The toolkit consists of a set of questions and scenarios that organisations can use to identify ethical considerations and take appropriate measures to mitigate risks.

The importance of ethical considerations in technology development has become increasingly evident in recent years, with several high-profile cases highlighting the potential risks associated with the misuse of emerging technologies. Ethical OS provides a structured approach to identifying and addressing ethical considerations, which can help organisations avoid potential legal and reputational risks.

LivAI sepie Funded by the Erasmus+ Programme of the European Union

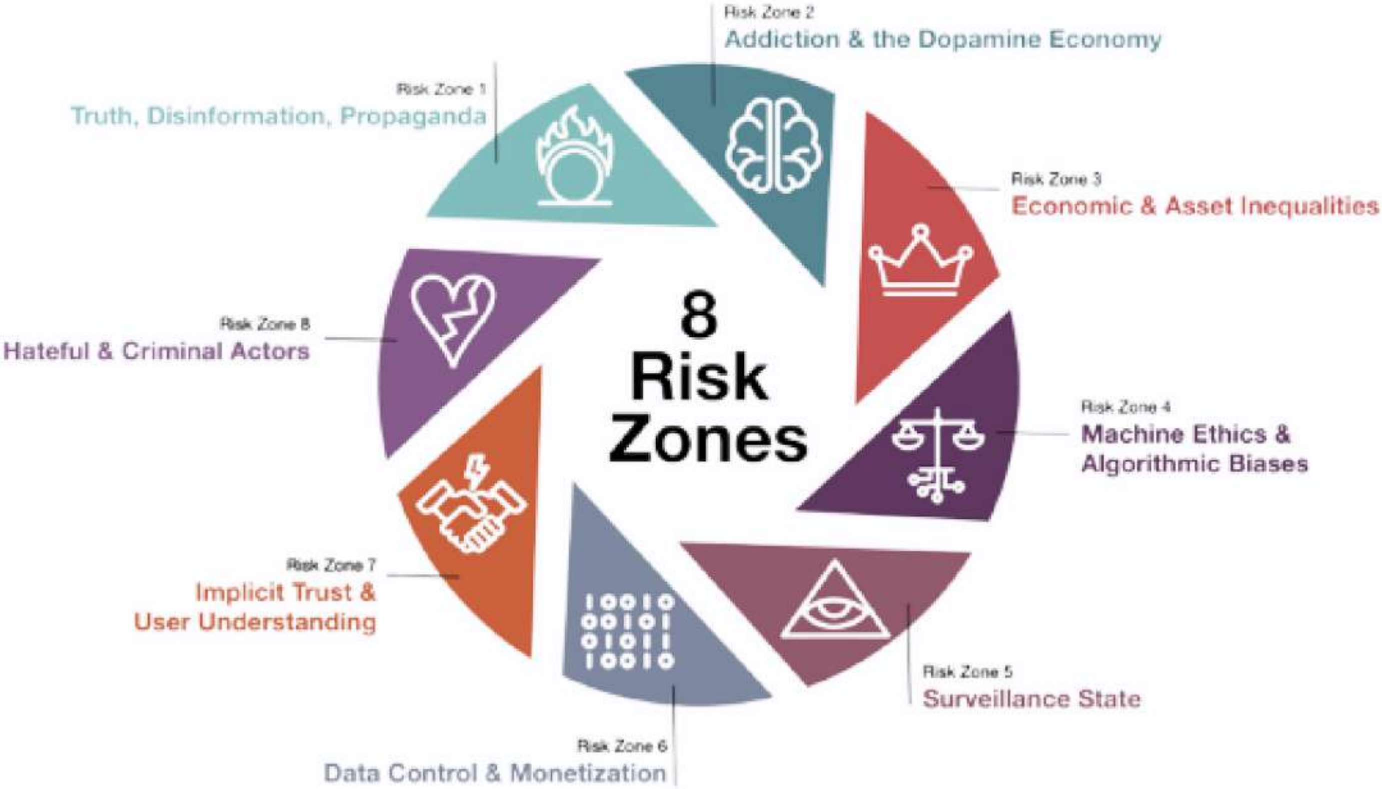

8 Risk Zones
Risk Zone 1
Truth, Disinformation, Propaganda
Risk Zone 2
Addiction & the Dopamine Economy
Risk Zone 3
Economic & Asset Inequalities
Risk Zone 4
Machine Ethics & Algorithmic Biases
Risk Zone 5
Surveillance State
Risk Zone 6
Data Control & Monetization
Risk Zone 7
Implicit Trust & User Understanding
Risk Zone 8
Hateful & Criminal Actors

Figure 1: Ethical OS checklist
(source: https://ethicalos.org/)

One of the key strengths of Ethical OS is its flexibility. The toolkit can be used by a wide range of organisations, from startups to large corporations, and can be applied to various stages of technology development. This makes it a versatile tool for organisations that are looking to build ethical considerations into their development processes. Another strength of Ethical OS is its focus on scenario planning. By asking organisations to consider hypothetical scenarios, the toolkit encourages proactive thinking and can help organisations anticipate potential ethical risks before they occur. This can help organisations take pre-emptive action to mitigate risks, rather than reacting to ethical concerns after they have already arisen.

Overall, Ethical OS is a valuable toolkit for organisations that are looking to build ethical considerations into their technology development processes. By providing a structured approach to identifying and addressing ethical risks, Ethical OS can help organisations avoid potential legal and reputational risks, and build trust with their stakeholders.

4.3.2 AI ETHICS LAB-TOOLBOX

The Toolbox (https://aiethicslab.com/toolbox/) is a collection of resources developed by the AI Ethics Lab to support organisations in their efforts to develop ethical and responsible AI systems. The toolbox is designed to provide practical guidance and actionable tools to help organisations navigate the complex landscape of AI ethics. The AI Ethics Lab toolbox includes several components, including frameworks, guidelines, and case studies. The frameworks and guidelines are designed to help organisations identify and address ethical considerations at various stages of AI development. They cover a range of topics, such as data ethics, transparency and explainability, fairness and bias, and accountability and governance.

LivAI sepie Funded by the Erasmus+ Programme of the European Union

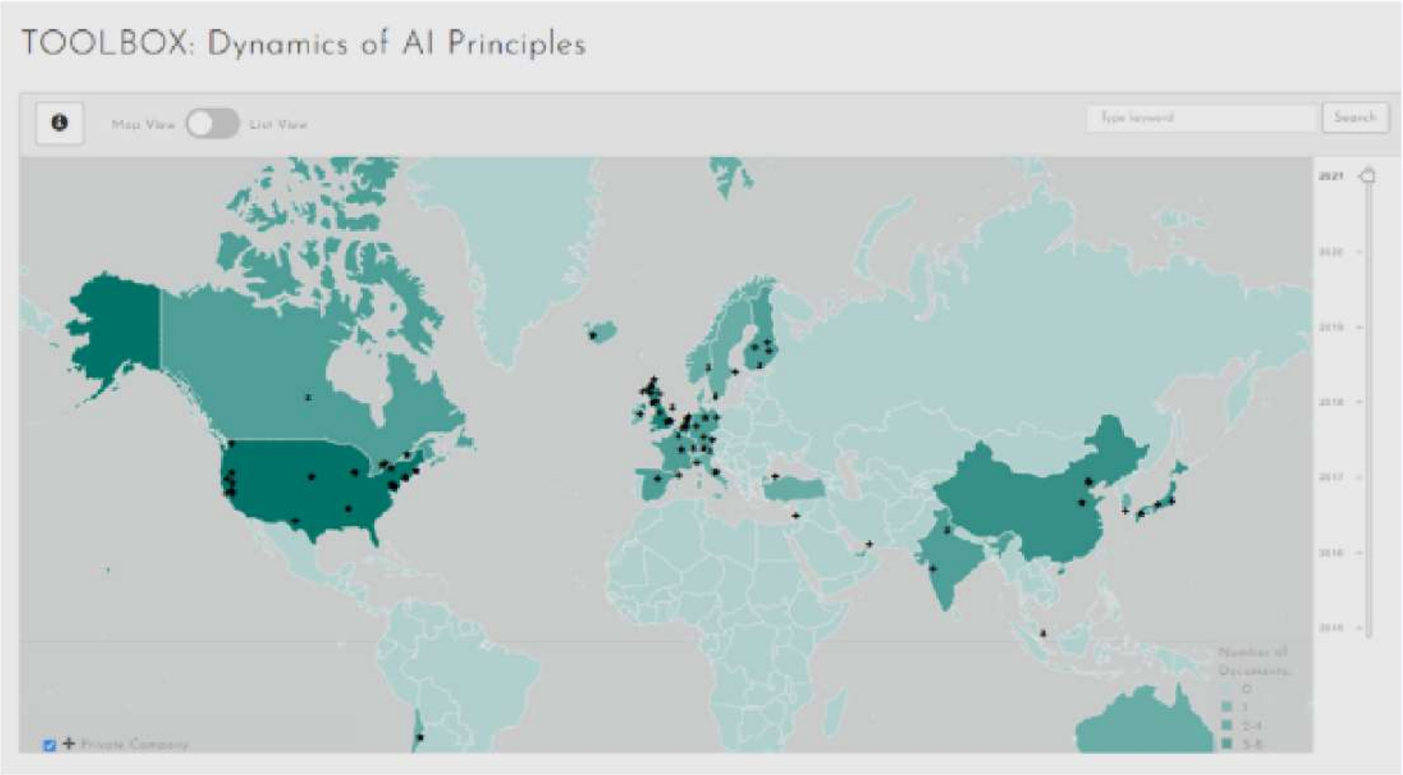


Figure 2:
TOOLBOX (source: https://aiethicslab.com/big-picture/)

The case studies provide real-world examples of ethical challenges that organisations have faced in developing AI systems, along with insights into how those challenges were addressed. The case studies are intended to help organisations learn from the experiences of others and to provide practical guidance on how to navigate ethical challenges. The toolbox is available on their website and is free to access. It includes a variety of resources, such as toolkits, checklists, and worksheets, that can be downloaded and used by organisations. It was designed to be flexible and adaptable to different organisational contexts and can be used by a range of stakeholders, including researchers, developers, policymakers, and business leaders.

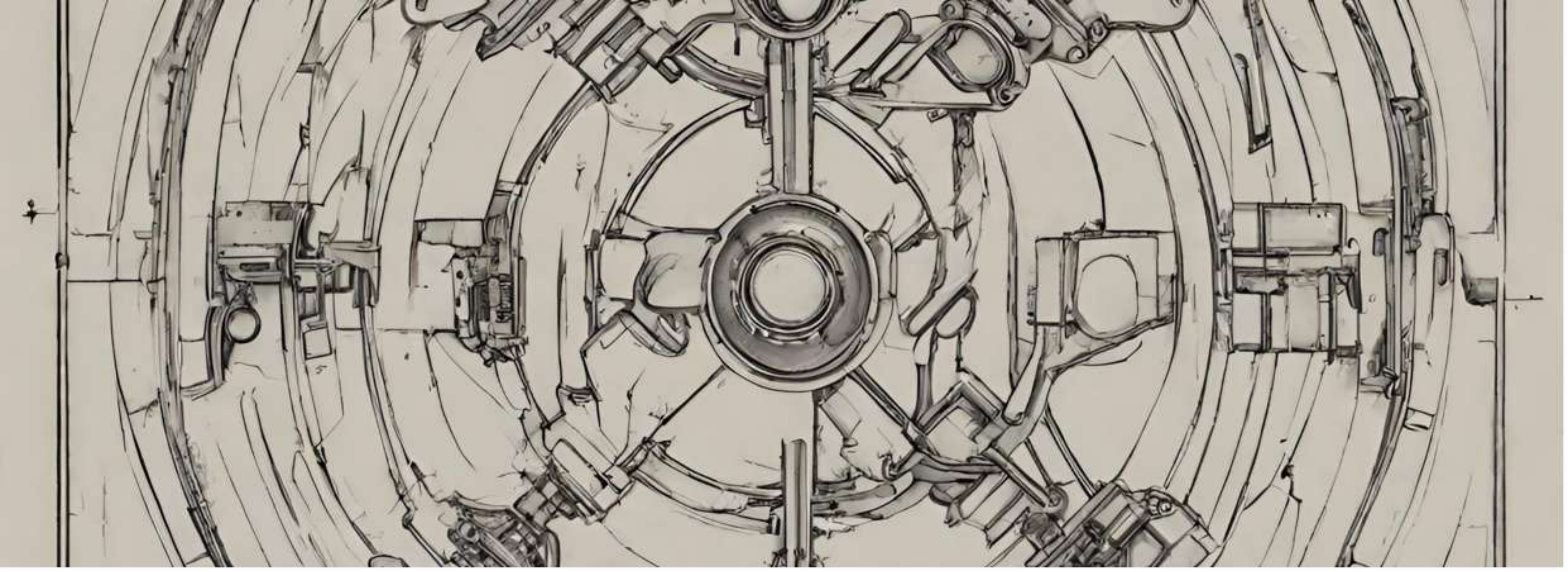
Overall, the AI Ethics Lab toolbox provides a valuable set of resources for organisations looking to develop ethical and responsible AI systems. Its focus on practical guidance and actionable tools makes it a useful resource for organisations at all stages of AI development.

4.3.3 ETHICS & ALGORITHMS TOOLKIT

The Ethics & Algorithms Toolkit (https://ethicstoolkit.ai/) is a resource developed by the Ada Lovelace Institute to help organisations design and deploy algorithmic systems in an ethical manner. The toolkit is intended to provide practical guidance and resources for organisations looking to implement ethical AI and algorithmic systems. The toolkit is organized into four main sections: Plan, Define, Develop, and Test. Each section includes a set of activities and resources to help guide organisations through the process of designing and deploying ethical algorithmic systems.

The Plan section includes guidance on scoping and framing the algorithmic system, identifying ethical considerations, and establishing governance and oversight structures. The Define section focuses on defining the system requirements and establishing ethical design principles. The Develop section covers the actual development process, including testing, evaluation, and ongoing monitoring. The Test section provides guidance on testing the system for fairness, transparency, and other ethical considerations.

The toolkit includes a range of resources, including checklists, templates, case studies, and best practice guides. It is designed to be used by a range of stakeholders, including developers, designers, policymakers, and regulators. The Ethics & Algorithms Toolkit webpage provides a comprehensive overview of the toolkit, including information on its development, how to use it, and additional resources. The webpage also includes a series of case studies and examples of how the toolkit has been used in practice by organisations across different sectors.



4.3.4 PROBAST TOOL

ProBASt (Probabilistic Binary Assessment) (https://www.probast.org/probast/) is a tool developed by the European Commission's Joint Research Centre (JRC) to help assess the risks of artificial intelligence (AI) systems. It uses probabilistic modelling to evaluate the likelihood and severity of risks associated with AI systems. The tool allows users to input information about the AI system being evaluated, such as its purpose, inputs, outputs, and potential impacts. The tool then generates a set of probabilistic models that can be used to assess the likelihood of different risk scenarios, such as a system failing to recognize certain inputs or producing biassed outputs.

One of the key features of ProBASt is its ability to generate quantitative risk assessments. Users can input their own data on the likelihood and severity of different risk scenarios, or use default values provided by the tool. The tool then calculates the overall risk of the AI system based on these inputs.

ProBASt also includes a user-friendly interface that allows users to visualise and explore the results of their risk assessments. This includes a dashboard that displays key metrics such as the overall risk level, as well as more detailed information on specific risk scenarios.

Overall, ProBASt provides a powerful tool for assessing the risks of AI systems and can help stakeholders make informed decisions about the development, deployment, and regulation of these technologies. The image below shows an example of the ProBASt interface, including a summary of risk scenarios and a visualisation of the overall risk level.

4.3.5 IEEE GLOBAL INITIATIVE

The IEEE Global Initiative is a platform for the development and implementation of ethical guidelines for the design and use of autonomous and intelligent systems. The initiative is focused on developing standards, certifications, and educational resources for engineers, technologists, and policy-makers to ensure that AI is developed and used in a responsible and ethical way. The IEEE Global Initiative was launched in 2016 and has since developed several initiatives and resources for the ethical development of AI. One of its main initiatives is the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, which has developed a comprehensive set of ethical principles and guidelines for the design and use of AI.

The ethical principles developed by the initiative include transparency, accountability, and privacy, among others. These principles are intended to guide the development and implementation of AI systems that are safe, reliable, and trustworthy.

The initiative also includes working groups that are focused on specific issues related to the ethical development of AI, such as the use of AI in healthcare and the development of standards for autonomous systems.

It has also developed a certification program for AI practitioners, which includes a set of ethical guidelines and standards that practitioners must adhere to. This certification program is designed to ensure that AI practitioners have the knowledge and skills necessary to develop and use AI in an ethical and responsible way.

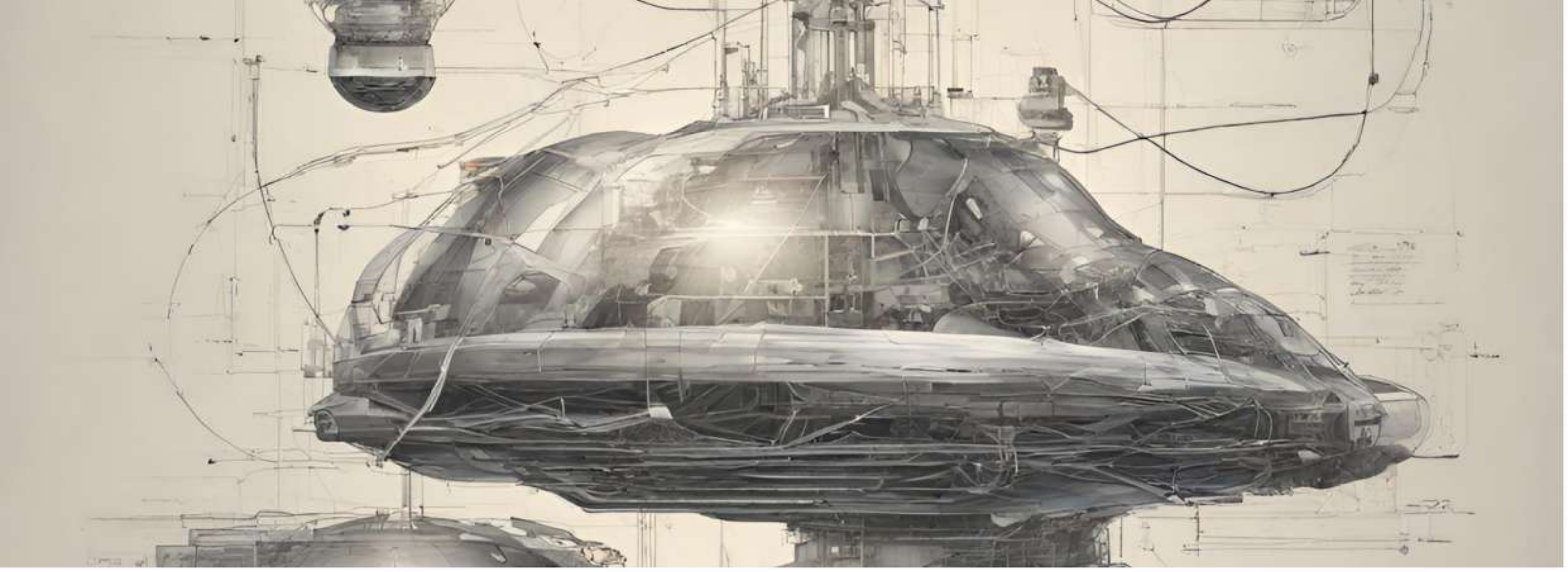
In addition to its initiatives and resources, the IEEE Global Initiative also hosts events and conferences focused on the ethical development of AI. These events bring together experts from academia, industry, and government to discuss the latest developments and best practices in the ethical development of AI. Overall, the IEEE Global Initiative is a valuable platform for the development and implementation of ethical guidelines for the design and use of AI. Its focus on transparency, accountability, and privacy, among other principles, is crucial for ensuring that AI is developed and used in a responsible and ethical way.

The advancement of artificial intelligence (AI) has led to concerns about the ethical implications of these technologies. To ensure that AI systems are safe, transparent, accountable, and aligned with human values, ethical frameworks need to be developed and implemented.

Tools and open platforms have been developed to support this effort, providing a structured approach to identifying and addressing ethical issues in AI. Ethical OS, AI Ethics Lab Toolbox, and Ethics & Algorithms Toolkit are examples of such resources that offer practical guidance and actionable tools to help organisations navigate the complex landscape of AI ethics. These resources are flexible, adaptable to different organisational contexts, and can be used by a range of stakeholders, including researchers, developers, policymakers, and business leaders.

Overall, these tools and platforms help to build trust and legitimacy with stakeholders while preventing unintended negative consequences, such as biassed or discriminatory outcomes, in AI systems.

4.4 ETHICAL ISSUES IN OSS



4.4.1 ETHICAL CONCERNS RELATED TO OPEN-SOURCE RELEASE OF INNOVATIVE TECHNOLOGIES

It is understood that open sourcing tools might both increase the probability of socially beneficial uses of the technology, but also use the technology for unintended use.

ABSTRACTIVE SUMMARIZATION

Summarization is a complex part of the project. The nature of this technology and its application is to filter the relevant part and enable the users to engage with a vast quantity of content pieces. To ensure best outcomes, we will provide transparent documentation and evaluate if and how the perception of content changes when summarised with the tool. During evaluation we will not only investigate how well the summarization works, but also how summarization tools will be used by the media professionals, and what effect the summarization technologies have on information gathering. Abstractive summarization rewrites the original text in a compressed way, this rewriting might lead to an incorrect layout of the original facts. One of our major efforts during the research is to minimise this problem but, as always, users should be aware and warned about the possible problem. We will also look for biases that might arise from the automatic distillation of large-scale news content and for ways to address this aspect of implementation of machine learning assisted summarization.

OPEN-SOURCE USE

By definition, open-source licensing cannot limit any type of usage even if it is unethical. The provider of open-source software can also not be liable for any unethical use of their product. It is generally accepted that the benefits of open-source release outweigh the risks. And while the current definition of an open-source licence excludes any limitations of usage, there is no doubt that there are voices and initiatives in the open-source community campaigning for a change in the definition to the one that focuses on ethical considerations. In this context, there will be an endeavour in closely monitoring any developments that might occur in this area.

COMMERCIAL USE

The main ethical implications from the commercial use of an open access platform appear to arise from the possible use of personal data to target individuals.

4.4.2 USING ‘OPEN SOURCE’ DATA

OSS data are publicly available but even though there are limits to their use when using ’open source’ personal data about identifiable people it should be made sure that the data processing is fair to the data subject and that their fundamental rights are respected. In case data from social media networks are obtained there is no obligation to have the data subjects’ explicit consent. However, it is important to evaluate if these people intended to make their information public through the privacy settings or limited audience to which the data were made available. It should also be clear that the intended use of the data within LivAI complies with the terms and conditions published by the data controller. The DPO’s contribution is necessary under these circumstances.

LivAI sepie
Funded by the
Erasmus+ Programme
of the European Union

4.4.3 SOFTWARE QUALITY

Quality software, in the traditional sense, is software that meets requirement specifications, is well tested, well documented, and maintainable (Schach, 2002). Advocates of OSS claim that its developers/users are motivated to do quality work because not only are they developing software for their own use, but their reputations among their peers also are at stake. Critics of OSS claim that volunteers will not do professional quality work if there is no monetary compensation. They also claim that documentation and maintenance are non-existent.

While it is true that documentation and maintenance are concerns, OSS advocates assert that OSS meets users' requirements, is tested by its developers and is constantly being upgraded. Documentation evolves as more and more users become interested in the software and use it. For example, books on Linux can be found everywhere. The question of whether OSS is of higher or lower quality than comparable commercial software is essentially an empirical rather than philosophical question.

LivAI sepie Funded by the Erasmus+ Programme of the European Union

The answer to this question is not readily available, but we can cite some preliminary anecdotal evidence on this issue.

The Apache web server is an OSS that competes with commercial web servers. The web server market is a potentially lucrative one, and we expect commercial software developers to compete for that market with high quality software. Yet despite commercial alternatives, according to third party observers (Netcraft, 2002) the OSS Apache server is by far the most used web server.

According to an August, 2002 survey, 63% of web servers on the Internet are Apache. At least in this market segment, it appears that OSS is sufficiently high quality for most users. Of course, Apache is free and other servers aren't; the cost motivation might explain some of Apache's popularity. But if the Apache server were of significantly lower quality than commercial alternatives, then it would be surprising to see its widespread use. This raises the question of whether market-dominance and popularity should be a benchmark for software quality. Does the fact that Microsoft Windows runs on some 90% of home computers assure us of its quality?

We would argue that popularity and quality might be linked if it can be shown that there is a level of expertise about software quality in the people making the choices. System administrators have more expertise than an average user of a home computer system.

Therefore, when a majority of these professionals choose an OSS alternative, it deserves notice. The Apache example illustrates an important distinction among OSS users. Initially, first adopters of OSS are its developers and as the code becomes more known, OSS gains users who were not involved in the development.

These users adopt the OSS for many reasons, but some of these new users (particularly non-programmers), appreciate the product, though they may not understand or care about the process that developed it. All users of OSS gain if the software delivers needed functionality. If an OSS project pleases its developers, but does not gather a following outside the developing community, that may be fine with the developers; if a commercial project only pleases its developers, it is a financial failure. The OSS model has different kinds of successes, and fewer outright failures. The rewards for developers in an OSS project are likely to be less tangible than rewards for a successful commercial product, but that does not make the rewards less real. The public has potential gains in the OSS movement that do not require large investments by the public. Another distinction between OSS projects and commercial projects is the lack of a release date. While open source developers anticipate frequent releases, there are no release deadlines. The announcements of a release day by a commercial vendor impose pressure on developers to cut corners, thus increasing the possibility of errors in the software.

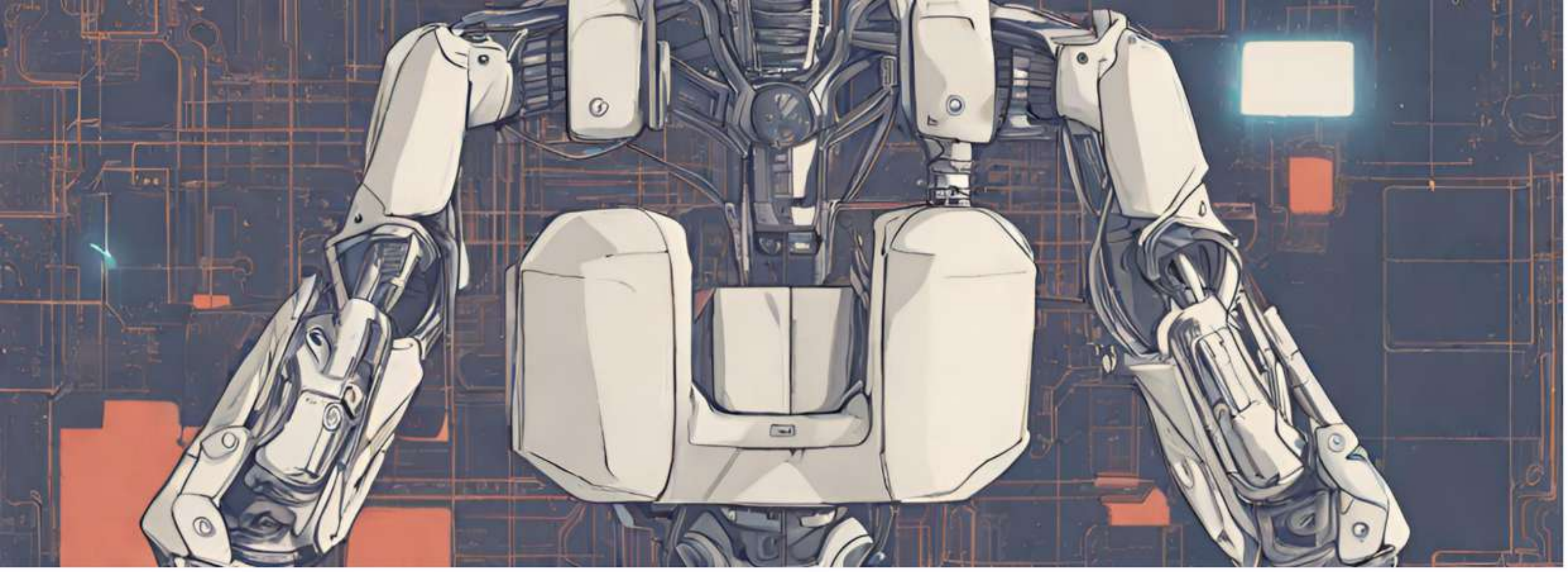
LivAI sepie Funded by the Erasmus+ Programme of the European Union

Furthermore, such a deadline has a tendency to impose on the autonomy of the developer. Finally we note that both open source and proprietary developers share the same professional ethical responsibility to develop solid, well-tested code. The social pressure in the open source community to avoid code forking provides incentives for project leaders to ensure that the code is the best it can be. On the other hand, when an open source developer believes there is too much risk associated with a particular piece of code, he/she can rewrite it and release it. While there is a reputation risk in doing so, there is the opportunity to publicly demonstrate that the forked product is superior. In a proprietary model, however, a developer's main avenue of recourse is to 'blow the whistle' on his/her manager or employer. To do so entails grave personal risk to one's livelihood, professional standing, lifestyle and family.

4.5 ETHICS GUIDELINES FOR TRUSTWORTHY AI BY THE EUROPEAN COMMISSION

4.5.1 THE AIM OF THE ETHICS GUIDELINES

The aim of the Ethics Guidelines for Trustworthy AI is to promote trustworthy AI. Trustworthy AI has three components, which should be met throughout the system's entire life cycle: (1) it should be lawful, complying with all applicable laws and regulations (2) it should be ethical, ensuring adherence to ethical principles and values and (3) it should be robust, both from a technical and a social perspective as , even with good intentions, AI systems can cause unintentional harm. Each component in itself is necessary but not sufficient for the achievement of Trustworthy AI. Ideally, all three components work in harmony and overlap in their operation.



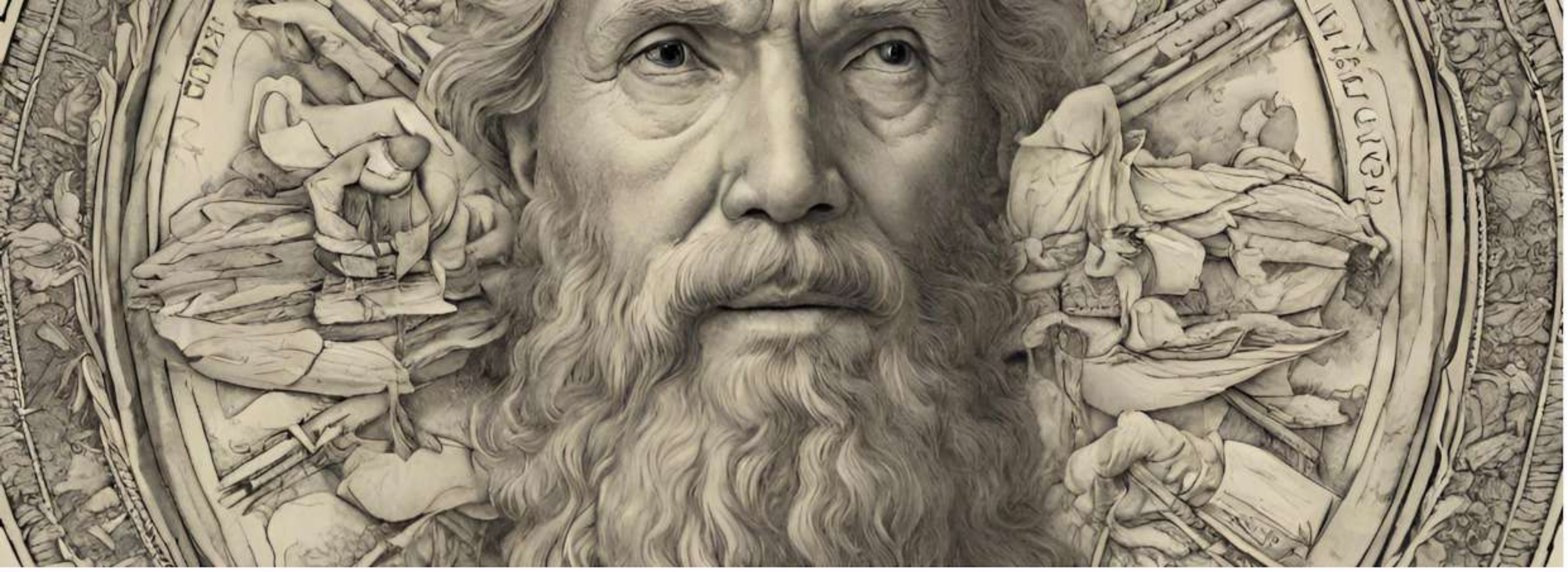
4.5.2 LAWFUL AI SYSTEMS

Lawful AI systems do not operate in a lawless world. A number of legally binding rules at European, national and international level already apply or are relevant to the development, deployment and use of AI systems today. Legal sources include, but are not limited to: EU primary law (the Treaties of the European Union and its Charter of Fundamental Rights), EU secondary law (such as the General Data Protection Regulation, the Product Liability Directive, the Regulation on the Free Flow of Non-Personal Data, anti-discrimination Directives, consumer law and Safety and Health at Work Directives), the UN Human Rights treaties and the Council of Europe conventions (such as the European Convention on Human Rights), and numerous EU Member State laws. Besides horizontally applicable rules, various domain-specific rules exist that apply to particular AI applications.

LivAI sepie Funded by the Erasmus+ Programme of the European Union

4.5.3 ROBUST AI

Even if an ethical purpose is ensured, individuals and society must also be confident that AI systems will not cause any unintentional harm. Such systems should perform in a safe, secure and reliable manner, and safeguards should be foreseen to prevent any unintended adverse impacts. It is therefore important to ensure that AI systems are robust. This is needed both from a technical perspective (ensuring the system's technical robustness as appropriate in a given context, such as the application domain or life cycle phase), and from a social perspective (in due consideration of the context and environment in which the system operates). Ethical and robust AI are hence closely intertwined and complement each other.



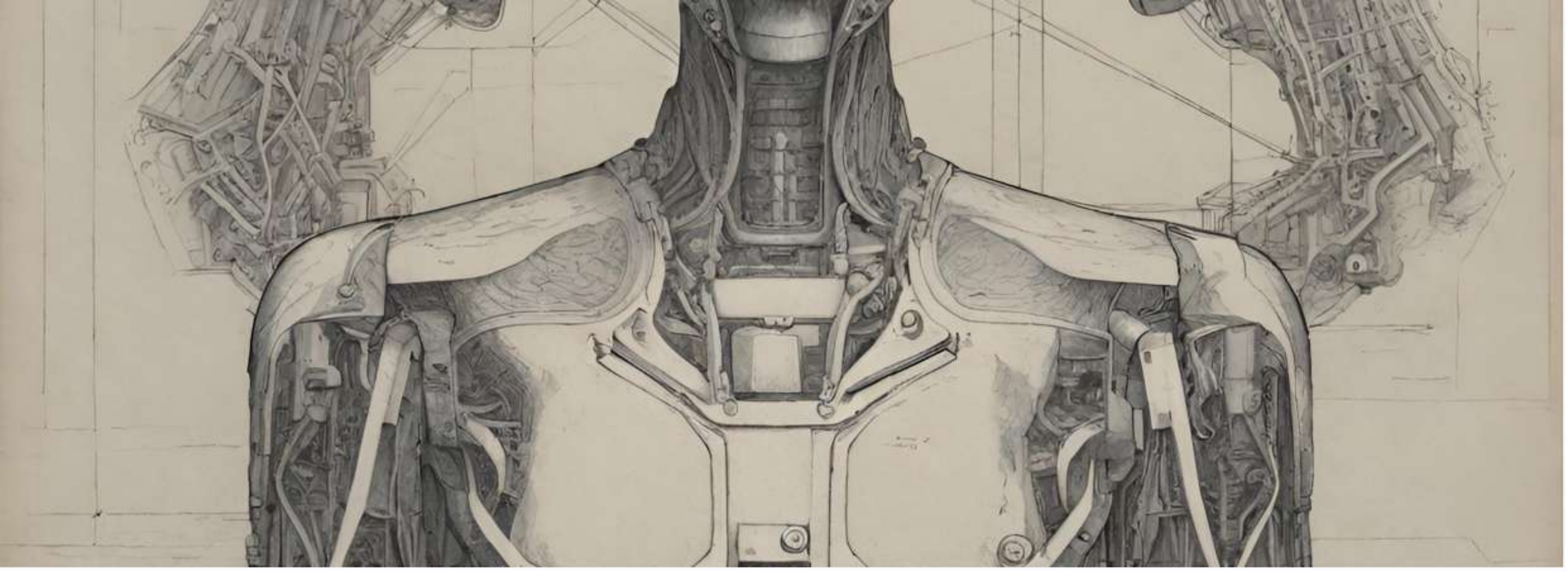
4.5.4 ETHICAL PRINCIPLES

In the Context of AI Systems many public, private, and civil organisations have drawn inspiration from fundamental rights to produce ethical frameworks for AI systems. In the EU, the European Group on Ethics in Science and New Technologies ("EGE") proposed a set of 9 basic principles, based on the fundamental values laid down in the EU Treaties and Charter. These ethical principles can inspire new and specific regulatory instruments, can help interpreting fundamental rights as our socio-technical environment evolves over time, and can guide the rationale for AI systems' development, deployment and use − adapting dynamically as society itself evolves. AI systems should improve individual and collective wellbeing.

This section lists four ethical principles, rooted in fundamental rights, which must be respected in order to ensure that AI systems are developed, deployed and used in a trustworthy manner. They are specified as ethical imperatives, such that AI practitioners should always strive to adhere to them.

Without imposing a hierarchy, we list the principles here below in a manner that mirrors the order of appearance of the fundamental rights upon which they are based in the EU Charter. These are the principles of: (i) Respect for human autonomy (ii) Prevention of harm (iii) Fairness (iv) Explicability.

Many of these are to a large extent already reflected in existing legal requirements for which mandatory compliance is required and hence also fall within the scope of lawful AI, which is Trustworthy AI's first component. Yet, as set out above, while many legal obligations reflect ethical principles, adherence to ethical principles goes beyond formal compliance with existing laws.



4.5.4.1 THE PRINCIPLE OF RESPECT FOR HUMAN AUTONOMY

The fundamental rights upon which the EU is founded are directed towards ensuring respect for the freedom and autonomy of human beings. Humans interacting with AI systems must be able to keep full and effective self-determination over themselves, and be able to partake in the democratic process. AI systems should not unjustifiably subordinate, coerce, deceive, manipulate, condition or herd humans. Instead, they should be designed to augment, complement and empower human cognitive, social and cultural skills. The allocation of functions between humans and AI systems should follow human-centric design principles and leave meaningful opportunity for human choice. This means securing human oversight over work processes in AI systems. AI systems may also fundamentally change the work sphere. It should support humans in the working environment, and aim for the creation of meaningful work.

LivAI sepie Funded by the Erasmus+ Programme of the European Union

4.5.4.2 THE PRINCIPLE OF PREVENTION OF HARM

AI systems should neither cause nor exacerbate harm or otherwise adversely affect human beings. This entails the protection of human dignity as well as mental and physical integrity. AI systems and the environments in which they operate must be safe and secure. They must be technically robust and it should be ensured that they are not open to malicious use. Vulnerable persons should receive greater attention and be included in the development, deployment and use of AI systems. Particular attention must also be paid to situations where AI systems can cause or exacerbate adverse impacts due to asymmetries of power or information, such as between employers and employees, businesses and consumers or governments and citizens. Preventing harm also entails consideration of the natural environment and all living beings.

4.5.4.3 THE PRINCIPLE OF FAIRNESS

The development, deployment and use of AI systems must be fair. While we acknowledge that there are many different interpretations of fairness, we believe that fairness has both a substantive and a procedural dimension. The substantive dimension implies a commitment to: ensuring equal and just distribution of both benefits and costs, and ensuring that individuals and groups are free from unfair bias, discrimination and stigmatisation. If unfair biases can be avoided, AI systems could even increase societal fairness. Equal opportunity in terms of access to education, goods, services and technology should also be fostered. Moreover, the use of AI systems should never lead to people being deceived or unjustifiably impaired in their freedom of choice.

LivAI sepie Funded by the Erasmus+ Programme of the European Union

4.5.4.4 THE PRINCIPLE OF EXPLICABILITY

Explicability is crucial for building and maintaining users' trust in AI systems. This means that processes need to be transparent, the capabilities and purpose of AI systems openly communicated, and decisions – to the extent possible – explainable to those directly and indirectly affected. Without such information, a decision cannot be duly contested. An explanation as to why a model has generated a particular output or decision (and what combination of input factors contributed to that) is not always possible. These cases are referred to as 'black box' algorithms and require special attention. In those circumstances, other explicability measures (e.g. traceability, auditability and transparent communication on system capabilities) may be required, provided that the system as a whole respects fundamental rights. The degree to which explicability is needed is highly dependent on the context and the severity of the consequences if that output is erroneous or otherwise inaccurate.

4.6 DATA PROTECTION

The scope of this section is to identify the ethics framework within which the LivAI research is conducted. Apart from assessing the legality of the conducted research, the ethics framework is important also because many ethical concerns related to the processing of personal data are already codified in the European data protection legislation. In this section, attention is drawn first to the data protection and privacy legal framework. As it was established in the previous section, different work packages in LivAI make use of personal data. These data are either collected directly from data subjects through interviews, are harvested online, or belong to existing databases that are available to the scientific community. As a result, the GDPR is applicable.

In LivAI, part of the open source data used qualifies as personal data. 'Personal data' are defined in the GDPR as any information relating to an identified or identifiable natural person ('data subject'). An identifiable natural person, on the other side, is the one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.

Furthermore, data collected during personal interviews conducted in the LivAI framework fall under the same definition of personal data. The anonymization of data that is envisaged in the LivAI architecture for data collected from open sources, as well as for migrant interviews and the pseudonymisation of data for expert interviews does not change this qualification for as long as the anonymization and pseudonymization activities are reversible or there is the possibility to identify an individual by the connection of different databases. As a result, the general data protection framework established by the GDPR is applicable.

Neither the GDPR nor the Police Directive regulate explicitly the use of open source data. They create, however, a general legal framework within which the processing of open source personal data takes place. In June 2019 the Open Data Directive was adopted. It addresses the re-use of materials held by public sector bodies in the Member States, at national, regional and local levels, such as ministries, state agencies and municipalities, as well as organisations funded mostly by or under the control of public authorities and also the research data pursuant to the conditions set out in article 10 thereof.

It focuses on the economic aspects of the reuse of information rather than on access to information by citizens and encourages the Member States to make as much information available for re-use as possible. The scope of this Directive is thus limited, and does not cover the open source data that will be used in the

LivAI framework. For complying with the data protection framework, first attention must be paid to the principles of lawful data processing established in article 5 GDPR. From the LivAI architecture as well as from the data knowledge and management plan, it is clear that all the work packages have considered the data processing principles as their priority.

The principles for lawful data processing, namely: (a) lawfulness, fairness and transparency; (b) purpose limitation; (c) data minimization; (d) accuracy; (e) storage limitation; (f) integrity and confidentiality; and (g) accountability are said to be followed and fulfilled during all the stages of the project and from all the members of the consortium. Furthermore, for personal data to be processed, compliance with the principle of lawfulness is very important. This principle requires compliance with one of the conditions established in article 6 GDPR. Depending on the provenience of the data, different lawfulness grounds might come into play.

Firstly, and importantly for data gathered from interviews, the consent of data subjects is required. The work packages that are collecting data from interviews have already prepared or are in the phase of preparing their consent forms. These consent forms must comply with the conditions for lawful consent that are established in Article 7 GDPR, with the right to withdraw consent and, not underestimate the vulnerability of the interviewed party. Secondly, for the open-source data that are collected from the internet, the condition of consent cannot be established. Making personal data available does not automatically qualify as giving the consent to whomever has access to these data to process them as deemed necessary. In this situation, the processing of the data for research purposes can be considered as lawful under the justification of performance of a task carried out in the public interest.

As it is evident from the questionnaires as well as from the description of data processed by different work packages, some of the processed data qualify as sensitive. A personal image, for example, might reveal the religion of the data subject or his ethnic origin.

Even though the consortium is not actively searching or processing such data, they might be part of the created databases and might also be revealed during interviews or be part of posts data subjects have made in social media platforms. According to the GDPR, processing of sensitive data should not take place unless falling under specific situations for which such processing can be justified. The work packages do not intend to further process sensitive data during the LivAI research, nor to engage in individual profiling, but their processing can be justified on the basis of article 9(2)(e) GDPR on data that are made manifestly public.

The processing of sensitive data for research purposes can be justified also under the provision of article 9(2)(j) GDPR that allows processing of sensitive data for archiving purposes in the public interest and for scientific reasons. With regards to the right to information of the data subjects, there are different approaches taken during the LivAI research. With regard to data that are collected directly from data subjects, data subjects are informed about their right at the moment in which they give their consent for data processing. With regards to the accountability of the members of the consortium, they all remain responsible in case of data breaches before the data protection authorities of their countries.

All the consortium members have also indicated Data Protection Officers already instituted in their organisations.

The Data Protection Impact Assessment DPIA is a process intended to evaluate the data-protection impacts of a project, policy, programme, product or service and, in consultation with relevant stakeholders, to ensure that remedial actions are taken as necessary to correct, avoid or minimise the potential negative impacts on the data subjects.

LivAI sepie Funded by the Erasmus+ Programme of the European Union

4.7 PRIVACY

While the processing of personal data during the LivAI research appears to comply with data protection rules, attention might also be paid to the protection of the right to privacy. The information obtained due to data processing may go far beyond what the individuals have made public on social media and thus severely interfere with their private sphere.

This will be done following the data protection rules and the results of the analyses will only serve the purpose of shaping the methodology of the necessary risk assessment analyses and all content data will be deleted once it is understood what types of data are available and how they can be integrated in the risk assessment framework. It cannot be excluded however, that interferences into the private sphere of individuals might occur also on the basis of the data analyses conducted by other WPs.

4.8 IP LAWS

Data in open sources can be subject to intellectual property rights, most importantly, copyright and database rights. For LivAI, reproduction or extraction of open-source data on social media can, consequently, only be done with a licence from the rights holder. A licence can be both explicit and implicit. In the case where data is extracted from social media platforms such as Facebook or Twitter, an implicit licence can be assumed based on the method of extraction. Twitter, for example, receives a non-exclusive licence to the Tweets of its users.

The extraction of the social media data is primarily done through the API (Application Programming Interface) of the platforms. Using this method, the social media platform remains in control of the method of data extraction, which means the platform controls the manner in which the intellectual property is transferred to the users of the API. For the LivAI research, for example, Twitter gives an explicit licence to use the intellectual property on its platform (i.e., the Tweets) for purposes restricted by its API. In addition, even where an explicit licence is not issued by the platform, an implicit licence could be derived from the platform's characteristics.

For LiVAI, the copyright protection of social media posts is thus not an issue where an explicit licence is issued through the platform's API agreement. Most platforms allow access to its posts through such an agreement. In case such an agreement is not published, an implicit licence can be further derived from the characteristics of the platform (on a case-by-case basis). The same argument applies to databases that are protected through intellectual property. Consequently, the LivAI research is conducted in compliance with the relevant IP laws.

4.9 CONCLUSION

OSS is no longer an academic curiosity. We have demonstrated that certain OSS products are making a significant niche for themselves in computing environments.

Both Apache and Linux are increasing in popularity. The OSS model is distinct from commercial software development from several viewpoints: as a software engineering process, as an economic plan, and as a marketing strategy. In both models, however, developers have certain obligations and responsibilities to their users. In our analysis we have found that the authors of OSS have complex motivations, some laudatory, and others less so. OSS has produced some successes, and the public has benefited from these. There are questions about reliability and professionalism, but evidence against the quality of OSS is not, as yet, convincing to us. It does not appear likely that OSS will displace commercial software in the foreseeable future, and we have not uncovered any ethical imperative that it should. Yet, OSS has distinct economic advantages for many especially in the academic arena. It can help bridge the digital divide and can involve growing numbers of people in computing, both as developers and users.

Developers of OSS strive to be the best they can to contribute to the sustainable whole and thus secure their reputation ethically among their peers. OSS and commercial software can coexist, each giving the public the goods it desires. Both advocates and critics of OSS have an ethical obligation to respect each other and to avoid inaccurate and mean- spirited accusations. All software developers have ethical obligations for quality and openness (Software engineering code of ethics, 1999). OSS is a novel development of traditional ideas of sharing academic intellectual property, but OSS exists in a world dominated by commercial enterprise.

As such, OSS challenges the status quo in a way that can be a constructive check on excesses of traditional free enterprise systems. In a time when many for-profit corporations have disappointed the public with their lack of ethical behaviour, OSS has the potential to be a positive ethical force in the world of computing. Hackers who get involved in OSS development can contribute to the sustainable whole and, thus ethically secure their reputation among their peers. This is a way to publicly excel at hacking without illegal and unethical harm to others.

BIBLIOGRAPHY

- Etzioni A, Etzioni O. Incorporating ethics into artificial intelligence. J Ethics. 2017;21(4):403−18. Available from: http://dx.doi.org/10.1007/s10892-017-9252-2.
- Ethical issues in Open Source Software, Sacred Heart University, Frances Grodzinsky, Keith W. Miller, Marty J. Wolf, 2003. https://silo.tips/download/ethical-issues-in-open-source-software.
- Ethics Guidelines for trustworthy AI in the European Union https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html
- [ONLINE] Datatilsynet report: Artificial intelligence and privacy, available at https://www.datatilsynet.no/globalassets/global/english/ai-and-privacy.pdf,last accessed on 12/5/2021
- HLEG, A. (2019). A definition of AI: main capabilities and disciplines. Brussels. https://ec. europa.eu/digital-single.
- [ONLINE] European Commission, ”Ethics guidelines for trustworthy AI”, available at https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai , last accessed on 12/5/2021
- Mitrou, L. (2018). Data Protection, Artificial Intelligence and Cognitive Services: Is the General Data Protection Regulation (GDPR)'Artificial Intelligence-Proof'?. Artificial Intelligence and
- Cognitive Services: Is the General Data Protection Regulation (GDPR)'Artificial Intelligence Proof.
- [ONLINE] European Commission, ”Guidelines 4/2019 on Article 25 Data Protection by Design and by Default”, available at https://edpb.europa.eu/our-work-tools/ourdocuments/guidelines/guidelines-42019-article-25-data-protection-design-and_en , last accessed on 12/5/2021
- [ONLINE] CIPL. ”CIPL Recommendations on Adopting a Risk-Based Approach to Regulating Artificial Intelligence in the EU”, https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_riskbased_approach_to_regulating_ai__22_march_2021_.pdf , last accessed on 12/5/2021
- Frances S. Grodzinsky, Marty J. Wolf, Keith William Miller ”Ethical Issues in Open Source Software” Article in Journal of Information Communication and Ethics in Society November 2003. https://www.researchgate.net/publication/241209540_Ethical_issues_in_open_source_software

More info:
livai.uji.es

Funded by the
Erasmus+ Programme
of the European Union

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

Project no: 2022-1-ES01-KA220-ADU-000087381