



Trustworthy **AI**

# Trustworthy AI Card Deck



Co-funded by the  
Erasmus+ Programme  
of the European Union

**Copyright note:**

This material is presented to ensure timely dissemination of scholarly and technical work. Copyright and all rights therein are retained by authors or by other copyright holders. All person copying this information are expected to adhere to the terms and constraints invoked by each author's copyright. In most cases, these works may not be reposted without the explicit permission of the copyright holder. This material is presented to ensure timely dissemination of scholarly and technical work. Copyright and all rights therein are retained by authors or by other copyright holders. All persons copying this information are expected to adhere to the terms and constraints invoked by each author's copyright. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International Licence.



**Copyright holder:** Stichting ALLAI Nederland

Section 1

# Introduction

# Trustworthy AI

AI has the potential to strengthen society and support humans on different fronts. However, we need to carefully implement AI-systems to guard our fundamental rights, health and safety. Developing, deploying and using AI in a human-centric manner can bring us AI that is safe, secure, fair and inclusive. To achieve this, the High-Level Expert Group on Artificial Intelligence developed the Ethics Guidelines for Trustworthy AI<sup>1</sup>. As a basis for these Guidelines the group chose 3 pillars that should underpin AI for it to be trustworthy:

1. it should be lawful, meaning that it should comply with all applicable laws and regulations
2. it should adhere to ethical principles and values, and
3. it should be both technically and socially robust and not cause unintentional harm

For pillars 2 and 3, the group developed **7 requirements for Trustworthy AI** that should guide the development, deployment and use of AI in Europe:

1. **Human Agency & Oversight:** AI systems should support human autonomy and allow them to make informed decisions. In order to achieve this, AI systems should act as enablers to a democratic, equitable society by supporting user's agency, fostering fundamental rights and allowing for human oversight. Students should be taught about appropriate levels of human agency and autonomy, human control and overall human dignity.



2. **Technical Robustness & Safety:** This principle requires that AI systems should be developed with a preventative approach to risks and in a manner such that they reliably behave as intended while minimizing unintentional and unexpected harm and preventing unacceptable harm. Additionally, the physical and mental integrity of humans should be ensured. Students should be taught about how to recognize and ensure accuracy and reliability of AI systems. Further, students should know how to balance technical robustness and ethical constraints.



3. **Privacy & Data governance:** Privacy is a fundamental right particularly affected by AI systems. Prevention of harm to privacy must be a priority. This requirement logically necessitates adequate data governance. This covers the quality and integrity of the data used, its relevance in light of the domain in which the AI systems will be deployed, its access protocols and the capability to process data in a manner that protects privacy. Students should be taught about how to collect and recognize high quality data, how to handle it sensitively, maintain privacy and prevent biases in the data and models built from it.



4. **Transparency:** This requirement has two parts. It encompasses transparency of elements relevant to AI systems. This includes the data collected, training and working of the system, explanations of its outcomes and the relevant business models. It also encompasses the obligation to be transparent about the use of AI systems and not use them covertly. Students should know how to recognize transparent systems, and also be provided with the skills to develop explainable AI. This involves teaching students on how to properly document and communicate about data usage, as well as decisions taken in the design process.



<sup>1</sup> Ethics Guidelines for Trustworthy AI, High-Level Expert Group on Artificial Intelligence, 2019





5. **Diversity, Non-Discrimination & Fairness:** The outcomes of AI-systems should be non-discriminatory and free from unacceptable bias. Another important part of this requirement is equal access through inclusive design processes. Inclusion and diversity must be enabled throughout the entire AI-system's life cycle. This includes the consideration and involvement of all affected stakeholders throughout the process. Students should be taught about the importance and added value of interdisciplinary expertise when developing, deploying and using AI systems. They should also be taught about the potential discriminatory effects of choices made throughout the development process.



6. **Environmental & Societal Well-Being:** The environment and society as a whole should be considered a "stakeholder" throughout the AI-system's life cycle. This requirement includes the encouragement of sustainability and ecological responsibility. It involves both research into AI solutions addressing climate change or other societal concerns, as well as being mindful about the ecological footprint of training and deploying an AI system. It also involves understanding and mitigating the larger societal, democratic or systemic effects AI can bring.



7. **Accountability:** This requirement necessitates that mechanisms be put in place to ensure responsibility and accountability for AI systems and their outcomes, both before and after their development, deployment and use. Students should be taught about auditing and record keeping, as well as legal frameworks for liability and being able to demonstrate minimization of negative effects.

These requirements must be addressed and evaluated continuously throughout the AI-system's entire life cycle - from the design and development phase to the end of usage - considering both technical and non-technical methods to ensure that they are met. By doing this, ethical and robust AI can be fostered and secured.

## Why teach Trustworthy AI?

Many students will likely be involved in AI development, deployment, procurement or use in their future occupations. Therefore, it is important that they acquire the necessary skills to develop and implement AI in a responsible and trustworthy manner. Thus, we have transformed the 7 requirements for Trustworthy AI (the "7 Requirements") into specific exercises and educational resources that help teach specific skills including the ability to:

1. Identify the applicability of the 7 Requirements in different contexts and its different dimensions for different stakeholders
2. Deliberate about possible implementations of the 7 requirements
3. Select and implement a course of action in response to ethical analysis regarding the requirements



By acquiring these skills and by developing the ability to understand and act according to ethical and social values, students can ensure a "human-in-command" approach to AI, where it remains up to them to decide if, when and how AI systems should be developed, deployed and used. Since the responsibility for Trustworthy AI does not only fall upon the developers, but on different stakeholders and experts in other fields, teaching students from different domains is of great importance.

## Trustworthy AI for STEM students



It is easy to imagine how students studying STEM subjects (mathematics, biology, physics, chemistry etc.) can easily end up being involved in any stage of AI development. Often discussions arise around AI's unintended consequences, algorithmic bias, data collection and protection etc. Many of these problems arise at the very first stages of the AI life cycle, its development. It is important that STEM students, the 'future AI developers', know how to appreciate ethical values early on and keep apply these values in different stages of AI development. This will enable them to identify unforeseen consequences or correlations or causal relationships between certain development choices and ethical problems, at an early stage. For example, they can ensure that data collection is done properly and responsibly, minimizing the risk of biased or harmful outcomes.

Every AI developer or AI researcher will study one or more STEM subjects at some point in their lives. Thus, as with any scientific researcher, it is important that ethics is well embedded in their educational career. Not only to prevent unforeseen consequences of a system's development, but to acquire reliable knowledge for making informed decisions.

## Trustworthy AI for Business Administration students



Teaching Trustworthy AI in the field of Business Administration is important given the exponential growth of AI in business settings or for commercial use and the fast growth of tech companies. Consumers are easily exposed to AI through e.g. customer service chatbots, predictive/recommending systems make decisions about them or IoT or IoB (Internet of Bodies) devices that use AI for their functioning. Within organizations, recruiting, hiring and worker assessment is done or supported by AI more often. Furthermore, businesses can benefit from the data analytics opportunities that AI can offer in any industry or use AI to manage business functions. For example, AI can be of great benefit for logistics as it can help determine the best organizational system for flights. Likewise, it can help in the healthcare sector by reviewing medical records or treatment approaches and supplement the knowledge of medical practitioners. The increased use of AI in an individual's daily life also makes it attractive for businesses to add AI to their product lineup. Meaning, by integrating and enhancing their product or services with AI features or technology.

Due to AI's wide applicability, it is important that businesses make sure that AI systems they develop, deploy or use is trustworthy. Although AI clearly brings many opportunities, companies should be cautious of risks, ranging from unsafe or unsecure AI to biased or unfair AI outcomes, but also to mass job displacement, or the potential to negatively impact society or democracy. One should not overlook these risks just because of the economic possibilities that AI can offer.

## Trustworthy AI for Political Science students



As we have discussed so far, it is not enough that AI is developed in a trustworthy manner, but it is also crucial that it is deployed responsibly. Unfortunately, we cannot put all of our trust and faith into businesses and developers for this to happen. Policies and legislative frameworks provide the boundaries within which AI can be acceptably developed and used. This is why it is important that political sciences students are educated about Trustworthy AI as well. Assuming that many of these students will become policy makers, or will have a role in governance, it is essential that they understand the true capabilities and limitations of AI and how policy and legislation can set the right boundaries for AI.

Section 2

# Using the Card Deck

## Trustworthy AI Card Deck

The purpose of using this card deck is to create meaningful ethical discussion in class, and to understand the diversity in which AI can be applied to different domains. Using this card deck, students will understand the complexity of some ethical dilemmas, and its relation to different stakeholders' points of view. The cards provide a general explanation of AI techniques that can be mixed and matched onto different domains to create a use case. A lack of familiarity with AI techniques can lead to certain difficulties with some games, which is why we have also created a deck of pre-selected use case cards for ease of use.

Before using the Card Deck, we advise to have students watch the short knowledge clips to gain familiarity with the concept of Trustworthy AI and the 7 Requirements for Trustworthy AI. Nevertheless, any extra knowledge about AI will be useful to come up with use cases required for some of the games.

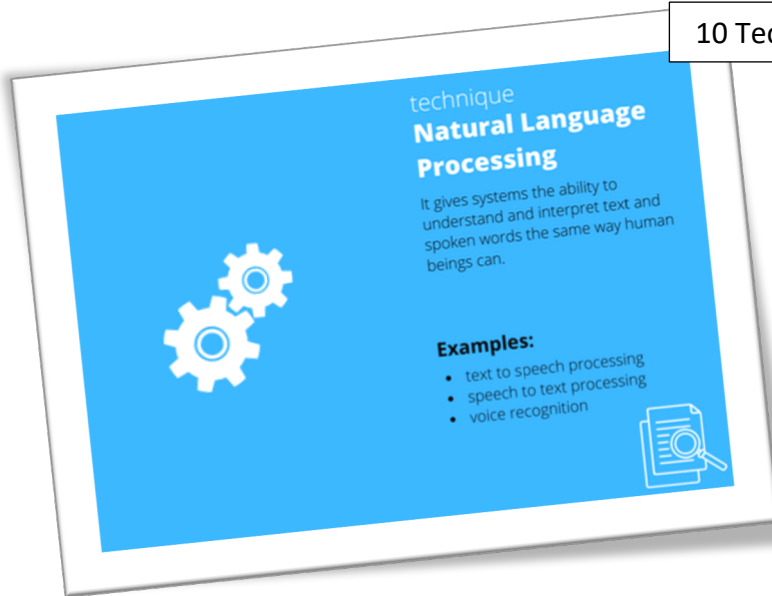
### *Composition of the Card Deck*

#### The deck consists of 5 different sets of cards:

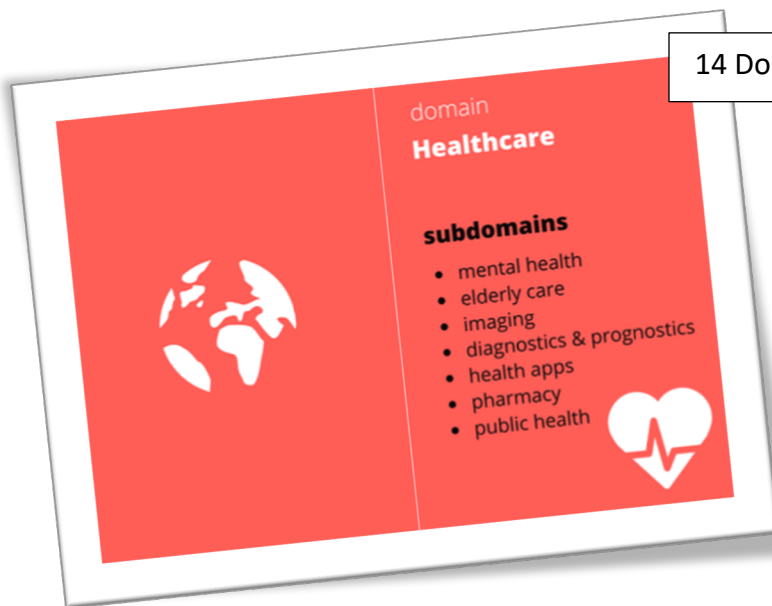
With these different sets of cards students can get inspired by use case examples to think of how to use different AI techniques in different domains. By using the requirement cards, they can critically analyze techniques and use cases from the viewpoint of different stakeholders. This can all be done through a series of class exercises/games as explained below.

- ◇ **AI technique:** A generic AI technique that can be used in many domains in various ways
- ◇ **Domain:** A sector in which an AI technique can be applied containing multiple subdomains falling under the same sector
- ◇ **Requirement:** An ethical requirement that evaluates the trustworthiness of an AI technique applied in a specific domain. These cards can also be used to evaluate Use Case cards.
- ◇ **Stakeholder:** A person's role in the development or deployment of an AI system having their own "competing interest" e.g., money, efficiency, safety, fairness, privacy, autonomy, etc.
- ◇ **Use case:** An example of the workings and goals of an AI technique applied to a domain where several stakeholders would be involved.

10 Technique Cards each with examples



14 Domain Cards, each with subdomains



8 Requirements Cards (7 Requirements for Trustworthy AI + Fundamental Rights)



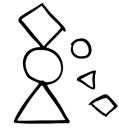


7 Stakeholders Cards with examples of the represented stakeholder



13 Use Case Cards

## Discussion starters



### Exercise 1 - Defend your use case

Materials: domain cards, technique cards, stakeholder cards, requirement cards

6 players: 1 judge, 5 developers

1. A card from the domain deck is drawn. This will be the domain in that round.
2. Each player draws and shows a stakeholder card.
3. Each player draws one technique card without showing.
4. Each player designs a use case for AI (on paper) within the domain from the point of view of their stakeholder role.
5. **NB: for an easier version: let each player picks up a use case card**
6. Once all players designed their use case, the judge will draw one requirement card per student.
7. The judge asks the first developer whether their use case complies with the requirement and if so, how.
8. If the answer is acceptable, the developer receives 2 points.
9. If it is not acceptable, or if the student is unable to answer, any of the other developers can argue why their solution complies with the requirement.
10. If this answer is acceptable, the other developer receives 2 points.
11. The judge repeats steps 7-10 for each of the developers.
12. The design with the most points is discussed by all developers to evaluate the results.

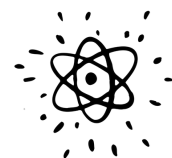
### Exercise 2 – Group design and discussion

Materials: Technique card, Domain card, Requirement card

5 players



1. As a group pick a domain card and a technique card.
2. Together, come up with a use case for that domain and technique.
3. Pick up one requirement card at a time and start discussing that requirement. Think of reasonably foreseeable uses/risks in which the AI application would disrespect such requirements, along with possible solutions to prevent this.
4. Repeat step 4 for each requirement.
5. Add on: As a group, rank the requirement cards in order of priority taking into account what was discussed in step 3.



## Exercise 3 - Discuss a use case card

Materials: Use case card, Stakeholder card, Requirement card

5 players

1. In small groups pick up a use case card.
2. Each player picks up a stakeholder card.
3. Pick up one requirement card at a time (5 in total) and start discussing that requirement from the perspective of your stakeholder.
4. Add on: Rank the requirement cards in order of priority from you own stakeholder's point of view.

## Stakeholder's views

A Stakeholder refers to a specific person in the development or deployment of an AI system having their own “(competing) interest” e.g., money, efficiency, safety, fairness, privacy, autonomy, *etc.* We understand that role-playing some of these stakeholders may be difficult for some, so we provide a few examples for inspiration!

**Governance:** Legal officers aim to ensure that AI is in line with rules and regulations, corporate sustainability officers aim to ensure that AI is ecologically and socially sustainable. Competing interest: legal compliance, sustainability.

**Authority/supervisor:** Supervisory authorities are independent public authorities that ensure uniform and consistent application and enforcement of the rules and laws (within their specific field of expertise). Think of Data Protection Supervisors, Market Authorities, Medical Safety Authorities, Labor Authorities *etc.* Competing interest: safety, privacy, competition, consumer protection, worker protection.

**Deployer:** The deployer of an AI system aims to ensure that the system operates correctly and at a profit and meets its goals. Competing interest: efficiency, money, safety.

**Affectee:** The affectee is the person or group of persons affected by the use of the system, such as consumers, citizens, workers, patients, children, people with disabilities. Competing interest: fairness, privacy, autonomy, agency, safety, equality, inclusiveness.

**Domain expert:** Domain experts aim to inform developers, deployers, and any other stakeholders with correct and reliable information about their field of expertise to ensure that a system works correctly both at a technical and social level. Competing interest: quality, safety, reliability.

**Developer:** Developers aim to develop systems that work correctly with the goals of the deployer in mind. Competing interest: efficiency, safety.



## The Card Deck

The next pages hold the full Card Deck in two versions:

1. Foldable version (for one sided printing)
2. Two-sided version (for two-sided printing)



domain

## Healthcare

### subdomains

- mental health
- elderly care
- imaging
- diagnostics & prognostics
- health apps
- pharmacy
- public health

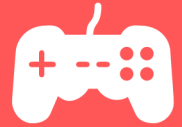


domain

## Entertainment

### subdomains

- music
- movies
- games
- social media
- art



domain

## Lifestyle

### subdomains

- smart homes
- smart speakers
- fitness apps
- online shopping
- travel



domain

## Production & Logistics

### subdomains

- manufacturing
- automated warehouse
- sales & marketing
- planning
- back office



domain

## Accessibility

### subdomains

- impaired sight
- impaired hearing
- dyslexia
- impaired mobility



domain

## Public sector

### subdomains

- law enforcement
- judiciary
- national security
- military
- public administration
- social benefits & services



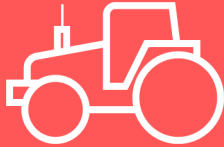


domain

## Agriculture

### subdomains

- animals
- plants
- large scale farming
- precision farming



domain

## Environment

### subdomains

- climate change
- data centres
- environmental protection



domain

## Transport

### subdomains

- cars, trucks
- trains
- planes
- boats
- bikes, scooters



domain

## Workforce

### subdomains

- recruitment
- worker management
- AI-worker interaction



domain

## Media

### subdomains

- online media
- social media
- political advertising



domain

## Research

### subdomains

- applied research
- fundamental research
- in-company research





domain

## Education

### subdomains

- access to education
- grading
- evaluating
- monitoring
- teaching & learning



domain

## Finance

### subdomains

- crypto
- loans, mortgages
- trading
- insurance







requirement

## Fundamental Rights

**Dignity:** Life, liberty & security,  
Privacy, Physical & mental integrity,  
Fair trial

**Freedoms:** Expression, Information,  
Assembly, Association

**Equality:** Non-discrimination, Gender  
equality, Equal treatment

**Social rights:** Education,  
Work, Social security



requirement

## Technical robustness and safety

**Resilience to attack and security:**  
Exploitations by adversaries, Dual use

**Fallback plan and general safety:**  
Unexpected situations,  
Harm to users/society

**Accuracy:** Correct judgements,  
Performance metrics

**Reliability and reproducibility:**  
Goals, Purposes,  
Intended applications



requirement

## Transparency

**Traceability:** Tracing of methods,  
development, validation and outcomes

**Explainability:** Opening the Black  
Box

**Communication:** Transparency  
about AI use (chatbots, etc.)



requirement

## Human agency and oversight

**Human agency:** Human autonomy,  
Self-determination

**Human oversight:** Human-in-the-  
Loop, Human-on-the-Loop,  
Human-in-Command



requirement

## Privacy and data governance

**Respect for privacy and data  
Protection:** (Sensitive) personal data,  
Encryption, anonymisation,  
aggregation

**Quality and integrity of data:**  
Correct labelling, Representativeness,  
DPIA

**Access to data:** Data governance  
protocols, processes and procedures



requirement

## Diversity, non-discrimination and fairness

**Unfair bias avoidance:** Data,  
Algorithm, Model, Design &  
development process

**Accessibility and universal design:**  
Disability, Age, Intelligence, Economic  
status

**Stakeholder participation:**  
Consumers, Workers,  
Citizens, Patients





requirement

## Societal and environmental well-being

**Sustainable and environmentally friendly AI:** Ecological footprint

**Societal impact:** Behavioral, psychological effects, Dangerous precedent

**Society and democracy:** Political discourse, Segregation, Systemic failure



requirement

## Accountability

**Auditability:** Independently auditable

**Minimising and reporting negative Impact:** Risk/impact assessment, Reporting vulnerabilities, biases, risks

**Documenting trade-offs:** Interests, Values

**Ability to seek redress:** Claims, Compensation of damage





stakeholders

## developer

### Examples:

- AI Expert
- Data scientists
- Statistician



stakeholder

## Affectee

### Examples:

- Citizen
- Patient
- Worker
- Student
- Child/Parent
- Consumer
- Person with disability



stakeholder

## Domain Expert

### Examples:

- Physician/doctor
- Judge/lawyer/DA
- Caseworker
- Teacher
- Financial expert
- Police inspector



stakeholder

## Deployer

### Examples:

- CEO
- Manager
- Head of Government Agency
- Minister





stakeholder

## **Authority / Supervisor**

### **Examples:**

- Privacy Authority
- Market Authority
- Financial Authority
- Consumer Authority



stakeholder

## **Governance**

### **Examples:**

- Legal Compliance Officer
- CSR Officer
- Sustainable Development Officer



stakeholder

## **Other**

### **Examples:**

- Society at large
- Environment
- Innovation
- Democracy







use case

## Translating Old Texts

To research the role of women in ancient Greece, a team of researchers must translate old parchment rolls before they can analyse the texts. They train a system using already translated rolls as input to easily translate the newfound texts. The system uses **optical character recognition** to convert the rolls into machine-encoded text/digital text.



use case

## The Perfect Teacher

The EU wants to fund a project to create the perfect **robotic tutor** for children on the autism spectrum. It is assumed that these children find it easier to communicate and open up to a robot tutor compared to a human teacher. Using a learning model, the robot helps them study, motivates them and simultaneously tracks their progress so that it can adapt its focus based on the student's needs.



use case

## Therapist at your Fingertips

Due to high demand for mental help, an app is developed to support those waiting to go to a psychologist. The main feature of the app is a **chatbot** that functions as a psychologist. It uses **natural language processing techniques** and transcribed conversations of therapy sessions to generate its answers. The goal is to make the user feel heard and possibly provide advice or even give exercises that can help the mental state of the user.



use case

## Hiring Machines

The HR team of a big company uses an AI classification system to help with their hiring process. The system classifies resumes as "yes" or "no" indicating whether an applicant is suited for the job. By doing this the HR team only needs to read the resumes classified as "yes" when manually reviewing them. The system is trained using **machine learning techniques**, taking judgements of previously reviewed resumes as its input and predefined features extracted using various **natural language processing tools**.



use case

## Smart Surveillance

At a large festival, a combination of **facial recognition** and smart **motion detection** are used for crowd control. The goal is to prevent outbreaks of crowd crushes or fights. Facial recognition is used to detect the emotions of the partygoers to be able to act quickly when things seem to get heated. Motion detection will help spot fights that are already happening or the beginnings of crowd crushes.



use case

## Paro: Your favourite Animal Companion

Paro is a **therapeutic robot** baby seal that is used primarily in care facilities. They are mostly used to calm and give a positive emotional response to people with dementia. They are able to analyse the specifics of the disorder and then intervene by creating controlled environments to reduce anxiety, as well as reduce the work-load of caregivers on a daily basis. It is intended to have the same effect as an animal-assisted therapy.





## use case Social Scoring

Due to the bad economy, the company you work for must make some changes regarding personnel. To not let personal relationships influence decisions and be as fair as possible, the company has used **machine learning** techniques to **evaluate and score** all employees based on their contributions to the firm.



## use case Crime Prevention

As an attempt to reduce crime in a city, its Mayor agrees to deploy an AI system that, through machine learning techniques, predicts the likelihood of someone being a victim of crime. The **crime forecasting** algorithm is trained on data such as searchable public records, court filings, licenses, addresses, phone numbers, social media data, demographics, age, gender, ethnicity, and socio-economic status. In response to the **predictive outcomes**, the city would mobilise its social support systems to protect them, without the use of law-enforcement.



## use case Increasing Senses

AI-synthetics generated media, also known as deep fakes, has begun to improve people's lives through deep learning technology. Advancements in this area has made accessibility tools smarter and more affordable for everyone, allowing blind people to see through AI-based recognition and synthetic voice to narrate objects, people, and the world. The technology is being used to empower people with disabilities in different ways.



## use case Disease Containment

In attempts to fight the Corona virus, countries have decided to deploy **thermal cameras** in public places such as malls, school, airports, and workplaces. These thermal cameras capture surface body temperature (**biometric data**) of individuals in crowds. People flagged with elevated body temperature are separated from the crowd to be subjected to further tests.



## use case Smart Money Advisors

In today's high frequency trading world, traders are starting to program or buy advanced software trading models that gives them a narrowed analysis of trading decisions. The models often use **fuzzy logic** to analyze thousands of factors in real time and present the investor with the best available opportunity. The technique has become attractive to use due to its accurate results using inaccurate data, and its lower hardware requirements than classical boolean logic.



## use case Increasing Senses

AI-synthetics generated media, also known as deep fakes, has begun to improve people's lives through deep learning technology. Advancements in this area has made accessibility tools smarter and more affordable for everyone, allowing blind people to see through AI-based recognition and synthetic voice to narrate objects, people, and the world. The technology is being used to empower people with disabilities in different ways.





use case

## Get Fit

Marie, a 16-year-old girl, has grown interest in joining the fitness community. She has started to do some research on the internet and social media about diets, exercise, and products that she can use to improve her health. Using machine learning techniques that learn from Marie's search history, Marie has been receiving personalized advertisements about certain products within the topic of fitness. She also noticed that the price of products she was interested seems to have gone up when she visited the websites a second or third time.





technique

## Natural Language Processing

It gives systems the ability to understand and interpret text and spoken words the same way human beings can.

### Examples:

- text to speech processing
- speech to text processing
- voice recognition



technique

## Image Processing

A method to manipulate or extract information from an image. This can be done in real time.

### Examples:

- object/face recognition
- temperature screening
- image quality improvement



technique

## Machine Learning

The capability of a computer to imitate human intelligence by learning from data to make informed decisions without explicit instructions.

### Examples:

- medical diagnosis
- classification
- prediction
- information retrieval



technique

## Deep Learning

A branch of machine learning that uses artificial neural networks to learn from data and make informed decisions. The networks are formed to mimic the human brain at a neuronal level.

### Examples:

- Virtual assistants
- translation
- chatbots
- image processing
- decision-making



technique

## Robotics

A branch of AI focusing on the design, construction and application of robots considering human-robot interactions.

### Examples:

- Social robots
- manufacturing robots
- delivery robots
- assistant robots
- humanoid robots
- monitoring robots



technique

## Predictive Systems

Predict future events based on historical data that the model has been fed. Through computational methods, it recognizes patterns in data to predict outcomes based on new data.

### Examples:

- Route-optimization
- Flight advisors
- Financial modelling
- Insurance/Risk Assessment







technique

## Recommending Systems

Recommending systems aim to predict the rating or preference a user would give to an item in order to give relevant item suggestions. It does this based on previous rating or data from the user collected in the past.

### Examples:

- Netflix
- YouTube
- Amazon
- Advertisement



technique

## Classification Systems

It gives systems the ability to understand and interpret text and spoken words the same way human beings can.

### Examples:

- Information retrieval
- Scoring



technique

## Content Generation

The ability to create content (written, visual, or spoken content) based on data of interest that allows the machine to learn the building blocks to create certain contents.

### Examples:

- Deepfakes
- Marketing,



technique

## Fuzzy Logic

A method of reasoning that imitates human reasoning. It recognizes variables containing more values other than true or false values. It aims to solve problems based on an imprecise spectrum of data

### Examples:

- Finance
- Facial pattern recognition
- Home appliances





domain

## Healthcare

### subdomains

- mental health
- elderly care
- imaging
- diagnostics & prognostics
- health apps
- pharmacy
- public health



domain

## Entertainment

### subdomains

- music
- movies
- games
- social media
- art



domain

## Lifestyle

### subdomains

- smart homes
- smart speakers
- fitness apps
- online shopping
- travel





domain

## **Production & Logistics**

### **subdomains**

- manufacturing
- automated warehouse
- sales & marketing
- planning
- back office



domain

## **Accessibility**

### **subdomains**

- impaired sight
- impaired hearing
- dyslexia
- impaired mobility



domain

## **Public sector**

### **subdomains**

- law enforcement
- judiciary
- national security
- military
- public administration
- social benefits & services



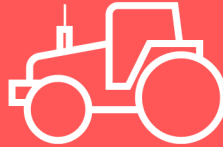
domain

## Agriculture



### subdomains

- animals
- plants
- large scale farming
- precision farming



domain

## Workforce



### subdomains

- recruitment
- worker management
- AI-worker interaction



domain

## Environment



### subdomains

- climate change
- data centres
- environmental protection



domain

## Media



### subdomains

- online media
- social media
- political advertising



domain

## Transport



### subdomains

- cars, trucks
- trains
- planes
- boats
- bikes, scooters



domain

## Research



### subdomains

- applied research
- fundamental research
- in-company research



domain

## Education



### subdomains

- access to education
- grading
- evaluating
- monitoring
- teaching & learning



domain

## Finance



### subdomains

- crypto
- loans, mortgages
- trading
- insurance



requirement

## Fundamental Rights



**Dignity:** Life, liberty & security,  
Privacy, Physical & mental integrity,  
Fair trial

**Freedoms:** Expression, Information,  
Assembly, Association

**Equality:** Non-discrimination, Gender  
equality, Equal treatment

**Social rights:** Education,  
Work, Social security



requirement

## Human agency and oversight



**Human agency:** Human autonomy,  
Self-determination

**Human oversight:** Human-in-the-  
Loop, Human-on-the-Loop,  
Human-in-Command



requirement

## Technical robustness and safety



**Resilience to attack and security:**  
Exploitations by adversaries, Dual use

**Fallback plan and general safety:**  
Unexpected situations,  
Harm to users/society

**Accuracy:** Correct judgements,  
Performance metrics

**Reliability and reproducibility:**  
Goals, Purposes,  
Intended applications





requirement

## Privacy and data governance

### **Respect for privacy and data**

**Protection:** (Sensitive) personal data, Encryption, anonymisation, aggregation

### **Quality and integrity of data:**

Correct labelling, Representativeness, DPIA

**Access to data:** Data governance protocols, processes and procedures



requirement

## Transparency

**Traceability:** Tracing of methods, development, validation and outcomes

**Explainability:** Opening the Black Box

**Communication:** Transparency about AI use (chatbots, etc.)



requirement

## Diversity, non-discrimination and fairness

**Unfair bias avoidance:** Data, Algorithm, Model, Design & development process

### **Accessibility and universal design:**

Disability, Age, Intelligence, Economic status

### **Stakeholder participation:**

Consumers, Workers, Citizens, Patients







requirement

## Societal and environmental well-being

**Sustainable and environmentally friendly AI:** Ecological footprint

**Societal impact:** Behavioral, psychological effects, Dangerous precedent

**Society and democracy:** Political discourse, Segregation, Systemic failure



requirement

## Accountability

**Auditability:** Independently auditable

**Minimising and reporting negative Impact:** Risk/impact assessment, Reporting vulnerabilities, biases, risks

**Documenting trade-offs:** Interests, Values

**Ability to seek redress:** Claims, Compensation of damage



stakeholders

## developer

### Examples:

- AI Expert
- Data scientists
- Statistician



stakeholder

## Domain Expert



### Examples:

- Physician/doctor
- Judge/lawyer/DA
- Caseworker
- Teacher
- Financial expert
- Police inspector



stakeholder

## Affectee



### Examples:

- Citizen
- Patient
- Worker
- Student
- Child/Parent
- Consumer
- Person with disability



stakeholder

## Deployer



### Examples:

- CEO
- Manager
- Head of Government Agency
- Minister



stakeholder

## **Authority / Supervisor**



### **Examples:**

- Privacy Authority
- Market Authority
- Financial Authority
- Consumer Authority



stakeholder

## **Governance**



### **Examples:**

- Legal Compliance Officer
- CSR Officer
- Sustainable Development Officer



stakeholder

## **Other**



### **Examples:**

- Society at large
- Environment
- Innovation
- Democracy





## use case **Translating Old Texts**

To research the role of women in ancient Greece, a team of researchers must translate old parchment rolls before they can analyse the texts. They train a system using already translated rolls as input to easily translate the newfound texts. The system uses **optical character recognition** to convert the rolls into machine-encoded text/digital text.



## use case **The Perfect Teacher**

The EU wants to fund a project to create the perfect **robotic tutor** for children on the autism spectrum. It is assumed that these children find it easier to communicate and open up to a robot tutor compared to a human teacher. Using a learning model, the robot helps them study, motivates them and simultaneously tracks their progress so that it can adapt its focus based on the student's needs.



## use case **Therapist at your Fingertips**

Due to high demand for mental help, an app is developed to support those waiting to go to a psychologist. The main feature of the app is a **chatbot** that functions as a psychologist. It uses **natural language processing techniques** and transcribed conversations of therapy sessions to generate its answers. The goal is to make the user feel heard and possibly provide advice or even give exercises that can help the mental state of the user.





## use case **Hiring Machines**

The HR team of a big company uses an AI classification system to help with their hiring process. The system classifies resumes as “yes” or “no” indicating whether an applicant is suited for the job. By doing this the HR team only needs to read the resumes classified as “yes” when manually reviewing them. The system is trained using **machine learning techniques**, taking judgements of previously reviewed resumes as its input and predefined features extracted using various **natural language processing tools**.



## use case **Smart Surveillance**

At a large festival, a combination of **facial recognition** and smart **motion detection** are used for crowd control. The goal is to prevent outbreaks of crowd crushes or fights. Facial recognition is used to detect the emotions of the partygoers to be able to act quickly when things seem to get heated. Motion detection will help spot fights that are already happening or the beginnings of crowd crushes.



## use case **Paro: Your favourite Animal Companion**

Paro is a **therapeutic robot** baby seal that is used primarily in care facilities. They are mostly used to calm and give a positive emotional response to people with dementia. They are able to analyse the specifics of the disorder and then intervene by creating controlled environments to reduce anxiety, as well as reduce the work-load of caregivers on a daily basis. It is intended to have the same effect as an animal-assisted therapy.





## use case **Social Scoring**

Due to the bad economy, the company you work for must make some changes regarding personnel. To not let personal relationships influence decisions and be as fair as possible, the company has used **machine learning** techniques to **evaluate and score** all employees based on their contributions to the firm.



## use case **Disease Containment**

In attempts to fight the Corona virus, countries have decided to deploy **thermal cameras** in public places such as malls, school, airports, and workplaces. These thermal cameras capture surface body temperature (**biometric data**) of individuals in crowds. People flagged with elevated body temperature are separated from the crowd to be subjected to further tests.



## use case **Crime Prevention**

As an attempt to reduce crime in a city, its Mayor agrees to deploy an AI system that, through machine learning techniques, predicts the likelihood of someone being a victim of crime. The **crime forecasting** algorithm is trained on data such as searchable public records, court filings, licenses, addresses, phone numbers, social media data, demographics, age, gender, ethnicity, and socio-economic status. In response to the **predictive outcomes**, the city would mobilise its social support systems to protect them, without the use of law-enforcement.





## use case

### Smart Money Advisors

In today's high frequency trading world, traders are starting to program or buy advanced software trading models that gives them a narrowed analysis of trading decisions. The models often use **fuzzy logic** to analyze thousands of factors in real time and present the investor with the best available opportunity. The technique has become attractive to use due to its accurate results using inaccurate data, and its lower hardware requirements than classical boolean logic.



## use case

### Increasing Senses

AI-synthetics generated media, also known as deep fakes, has begun to improve people's lives through deep learning technology. Advancements in this area has made accessibility tools smarter and more affordable for everyone, allowing blind people to see through AI-based recognition and synthetic voice to narrate objects, people, and the world. The technology is being used to empower people with disabilities in different ways.



## use case

### A Farmer's Best Friend

The agricultural sector has the enormous challenge of significantly reducing the ecological impact of their food production, while remaining able to feed the growing global population. Mike, a farmer, has hired an AI company to help him with crop forecasting, increasing output, and reducing costs. The company uses a **smart information system** that analyses data collected from farm equipment sensors. Mike pays 15\$ per acre and obtains around 100\$ increase in profits using the recommendations of the system. He also has access to the algorithms that show historical trends of soil moisture, and crop level weather patterns, and plug in different seeds and receive its predictive likely yields in the upcoming season.



## use case

### Get Fit



Marie, a 16-year-old girl, has grown interest in joining the fitness community. She has started to do some research on the internet and social media about diets, exercise, and products that she can use to improve her health. Using machine learning techniques that learn from Marie's search history, Marie has been receiving personalized advertisements about certain products within the topic of fitness. She also noticed that the price of products she was interested seems to have gone up when she visited the websites a second or third time.







technique

## Natural Language Processing

It gives systems the ability to understand and interpret text and spoken words the same way human beings can.

### Examples:

- text to speech processing
- speech to text processing
- voice recognition



technique

## Image Processing

A method to manipulate or extract information from an image. This can be done in real time.

### Examples:

- object/face recognition
- temperature screening
- image quality improvement



technique

## Machine Learning

The capability of a computer to imitate human intelligence by learning from data to make informed decisions without explicit instructions.

### Examples:

- medical diagnosis
- classification
- prediction
- information retrieval



technique

## Deep Learning

A branch of machine learning that uses artificial neural networks to learn from data and make informed decisions. The networks are formed to mimic the human brain at a neuronal level.



### Examples:

- Virtual assistants
- translation
- chatbots
- image processing
- decision-making



technique

## Robotics

A branch of AI focusing on the design, construction and application of robots considering human-robot interactions.



### Examples:

- Social robots
- manufacturing robots
- delivery robots
- assistant robots
- humanoid robots
- monitoring robots



technique

## Predictive Systems

Predict future events based on historical data that the model has been fed. Through computational methods, it recognizes patterns in data to predict outcomes based on dew data.



### Examples:

- Route-optimization
- Flight advisors
- Financial modelling
- Insurance/Risk Assessment





## technique **Recommending Systems**

Recommending systems aim to predict the rating or preference a user would give to an item in order to give relevant item suggestions. It does this based on previous rating or data from the user collected in the past.

### **Examples:**

- Netflix
- YouTube
- Amazon
- Advertisement



## technique **Classification Systems**

It gives systems the ability to understand and interpret text and spoken words the same way human beings can.

### **Examples:**

- Information retrieval
- Scoring



## technique **Content Generation**

The ability to create content (written, visual, or spoken content) based on data of interest that allows the machine to learn the building blocks to create certain contents.

### **Examples:**

- Deepfakes
- Marketing,



## technique

# Fuzzy Logic

A method of reasoning that imitates human reasoning. It recognizes variables containing more values other than true or false values. It aims to solve problems based on an imprecise spectrum of data



### Examples:

- Finance
- Facial pattern recognition
- Home appliances

