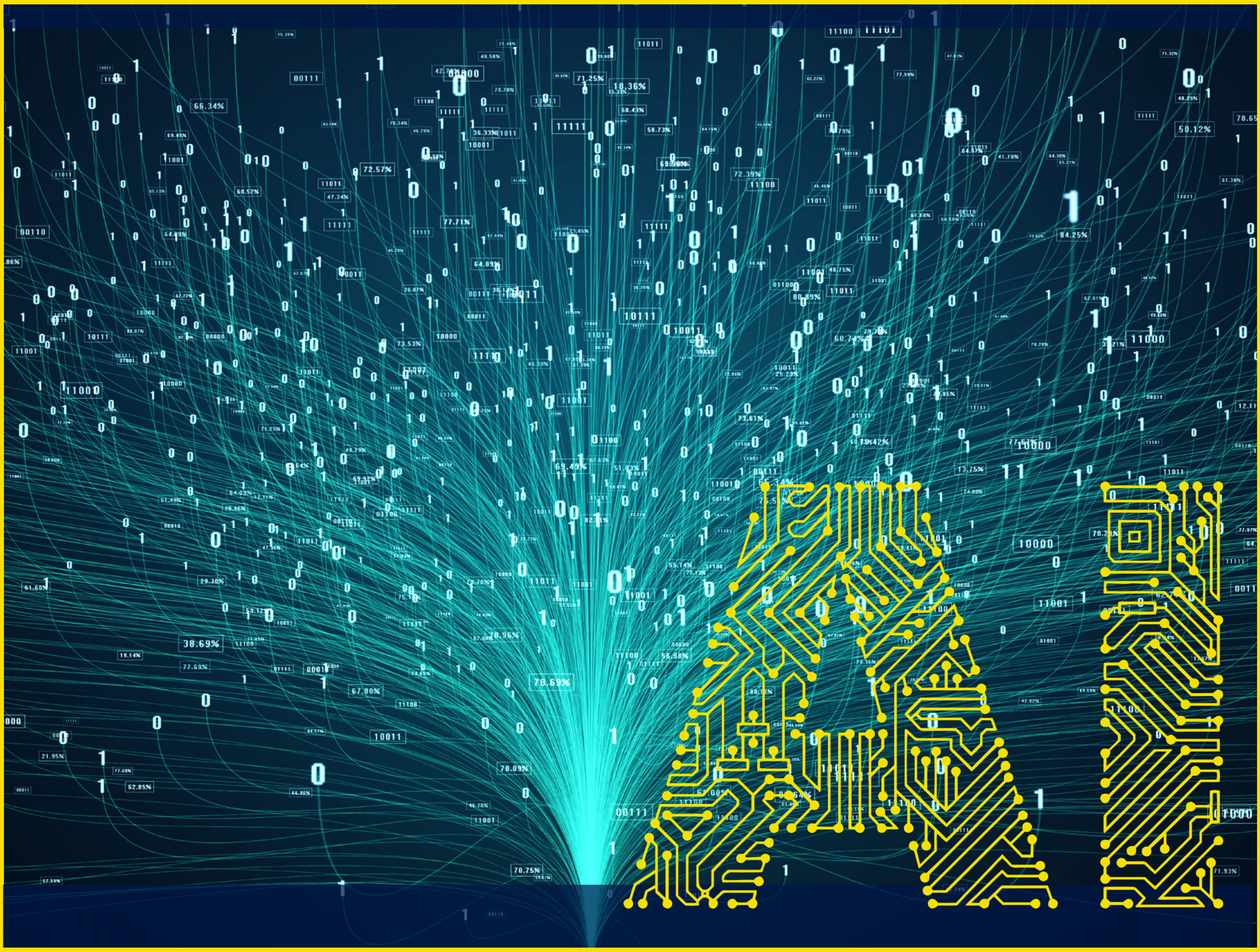
Funded by the European Union

CREATION OF A TAXONOMY FOR THE EUROPEAN AI ECOSYSTEM

A report of the Cross-KIC Activity "Innovation Impact Artificial Intelligence"



CONTENTS

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

The European Artificial Intelligence (AI) landscape is constantly evolving. Having put AI on the list of priorities, the European Commission fueled a multitude of activities. Additionally, the global race to AI is further accelerating. While each activity, report, landscape or any other output is valuable individually, there is a risk of losing oversight and efficiency.

To address this issue and allow for a comparison and integration of AI related studies, reports, companies, etc., in this report 35 existing frameworks have been scanned and analyzed. The developed AI taxonomy is compatible with the existing ones as long as they don't have used incorrect or inconsistent clusters or categories.

On top of that framework, a database structure has been built and integrated into the existing KIC tools. Moreover, an initial identification of the existing AI ecosystem in Europe has been conducted and added to the database. Finally, the central documents of the European Commission and other European bodies have been analyzed and compared. to ensure compatibility.

This work has been performed as part of an activity in which several innovation communities (KICs) of the EIT (the European Institute of Innovation and Technology) joined forces to address the challenges of AI in Europe. EIT ClimateKIC lead the workpackage and collaborated with UnternehmertTUM to create this report.

EIT Climate-KIC, EIT Digital, EIT InnoEnergy,
EIT Health and EIT Manufacturing, are co-funded
by the EIT, a body of the European Union.

1. CREATING TRANSPARENCY OF THE EUROPEAN AI LANDSCAPE

The prevalence and importance of AI has greatly risen during the last years spurred by advances in research and increasing dissemination and adoption of AI in business. The rapid development in the field, spurred by ever-new products, ideas and applications evolving out of research, startup activities and corporate innovation agendas, creates high intransparency of the current use case landscape, players in the field of AI as well as relationships between those. This intransparency has become a problem in coordinating and steering AI activities and developing an aligned, goal-directed and efficient approach to AI-related policy-making. Creating transparency of the European AI landscape must be the first step and the basis for future aligned actions on EU level to maximize impact.

Therefore, the appliedAI initiative has developed a comprehensive framework to map relevant actors and technologies in the AI space, that should serve as a basis to collect state-of-the-art information about organizations, use cases, financing opportunities and technologies in the AI space. Over time, the goal is to complete this basic framework by filling it with relevant content and thereby create a rich database that informs practitioners as well as policy-makers alike in steering their AI agendas. In particular, the goal of this project was to develop a taxonomy for mapping the European AI ecosystem that is:

- based on existing structures where reasonable
- extensive in covering the whole field of AI technologies
- able to map adjacent technologies like IoT, sensor devices etc.

and that serves the purpose of:

- Creating a consistent view and visibility to help steer decision-making
- allowing purpose-driven identification of relevant players and technologies.

This report will guide you through 1) the groundwork that was conducted to establish a sound basis for the development of the proposed framework (Chapter 2) as well as 2) introduce the resulting framework structure that has been derived based on the goals mentioned above and the results of step 1 (Chapter 3).

2. REVIEW OF EXISTING APPROACHES

2.1 OVERVIEW OF EXISTING FRAMEWORKS

Mapping the AI landscape has been attempted by many actors in the past already with very different goals in mind. As no standardized approach exists yet, the resulting frameworks vary significantly with regards to categorized objects, clustering dimensions and selection of subcategories.

However, the good work that has already been done in this field should not be ignored in this project. Quite to the contrary, it is an explicit goal of this project to also investigate existing landscapes to:

- make sure that no important aspects or dimensions are omitted in the development of the final framework and
- ensure integrability and alignment with existing structured where possible and reasonable

Therefore, 35 existing frameworks have been analyzed and compared in detail as a basis for the development of a comprehensive European AI taxonomy. 32 of these 35 frameworks have been developed and published by commercial actors, the remaining 3 are research-based taxonomies.

To identify these 35 frameworks, systematic research has been applied. At first, a list of keywords was defined and these keywords then were used for initial desk research. The results were screened and assembled in a comprehensive overview sheet. Next, based on the first screened frameworks, the set of keywords has been modified and extended. With the new set of adjusted keywords, the search and screening step was repeated until no new frameworks or no framework offering new insights were found. The assembled overview was then analyzed in depth as explained in the following sections.

The most basic differentiating characteristic is the object to be classified. Nearly all of the analyzed landscapes focus mainly on one of two objects - either companies (18 frameworks) or technologies (15). Only 2 of 35 do not and focus on showing an overview of AI use cases instead. Frameworks with companies as core elements can be further divided into two groups. The first group, the majority (14), maps startups, the second (4) focuses on companies in general with no specific restrictions.

To assess all existing frameworks in more detail and make them easier to compare and derive deviations, a detailed analysis of the dimensions used to structure the mappings has been carried out. This approach showed that depending on what is the core focus of the mapping, the dimensions that are being used as well as the respective categories per dimension vary widely. Especially company-focused frameworks are associated with a different set of clustering dimensions than technology-focused frameworks.

In total, eight different dimensions for clustering objects are being used within the 35 frameworks that have been investigated for this project:

- Industry
- Enterprise Function
- AI Capabilities (or "Subcategories of AI applications or structures to structure the wide range of possible AI technologies and approaches")
- Technical Infrastructure
- Customer Focus (B2C/B2B)
- Geographic Location
- Funding Stage, and
- Ecosystem Roles.

All frameworks that have been investigated as well as the dimensions used by each framework are shown below in table 1. The most prevalent dimensions of company-focused landscapes are industry, enterprise functions and AI capabilities and to a lesser extent technical infrastructure. Technology-focused frameworks include in all cases the dimension AI capabilities and additionally in one case technical infrastructure.

A key observation during the detailed assessment of frameworks it became obvious, that despite the general dimensions in which objects are being clustered are often shared between frameworks, the particular subcategories of the individual dimensions differ strongly between the publishers and even with the same publisher when comparing landscapes from consecutive years (ref. CB Insights 2017 - 2020). Due to the strong variation, analyzing changes over time and comparing

Name	Categorized Object	Industry	Capabilities	Enterprise Function	Infrastructure
Daxue Consulting - China AI Landscape	Companies	x	x	x	x
Element AI - Canadian AI Startups	Companies				
Firstmark Data & AI	Companies	x	x	x	x
Linux Foundation - AI Tech Ecosystem Landscape	Companies		x		x
appliedAI Startup Landscape	Startups	x	x	x	x
Bloomberg Beta	Startups	x	x	x	
CB Insights 2017	Startups	x	x	x	
CB Insights 2018	Startups	x	x	x	
CB Insights 2019	Startups	x		x	
CB Insights 2020	Startups	x	x	x	x
CB Insights Agriculture	Startups		x		
CB Insights Healthcare	Startups		x		
CB Insights Retail	Startups		x		
Cognite Ventures DL	Startups	x	x	x	x
MMC Ventures UK AI	Startups	x	x	x	x
StartHub Computer Vision	Startups	x	x	x	
StartHub Israel	Startups	x	x	x	
StartHub NLP	Startups	x	x	x	
3XN - AI Taxonomy	Technology		x		x
AI & Intelligent Automation Network - AI Classification	Technology		x		
Algorithmia - AI & ML	Technology		x		
appliedAI Use Case Cards	Technology		x		
appliedai.com Search platform	Technology	x	x	x	
Classification of AI - Ajit Nazre & Rahul Garg (Investors)	Technology		x		
Deloitte - AI Definition	Technology		x		
Forbes - AI Knowledge Map	Technology		x		
McKinsey - AI Capabilities	Technology		x		
McKinsey Mapping	Technology		x		
PWC - AI layers	Technology		x		
Small Data Group - Taxonomy of AI & ML	Technology		x		
Taxonomy of Machine Learning Based Anomaly Detection and its suitability	Technology		x		
Thomson Reuters - AI	Technology		x		
Papers with Code - Tech mapping	Technology		x		
AI and Climate Paper	Use Cases		x		
Plattform Lernende Systeme	Use Cases	x	x	x	

Table 1: Overview of AI frameworks - clustering dimensions 1 - 4.

the different landscapes from a content perspective is difficult. Additionally, the dimensions are often incomplete and include subcategories with heterogeneous granularity. For example, the dimension AI capabilities is especially problematic as it is often technically incorrect in addition to being incoherently structured. This leads to the conclusion that no currently existing framework fits the criteria for a comprehensive framework.

2.2 ASSESSMENT OF SUITABILITY OF EXISTING APPROACHES FOR A COMPREHENSIVE EUROPEAN AI TAXONOMY

Based on the analysis of thirty-six existing frameworks and the goals described in section 1, the requirements for a comprehensive European AI taxonomy can be derived. The analysis of existing frameworks has shown that there is no aligned approach regarding the mapping of AI-related organizations or technologies yet. The use of dimensions as well as the specific categories per dimension differ widely. To derive a comprehensive mapping based on the existing frameworks, each dimension has been assessed in more detail to derive a unified set of dimensions and categories by dimension, that fulfils the following requirements:

- Based on the investigated frameworks, all relevant dimensions should be covered.
- The dimensions as well as the categories in each dimension should be mutually exclusive as well as collectively exhaustive, to make sure each organization or technology can be mapped to the framework and the category to map it to is as unambiguous as possible.
- Technology-related dimensions should clearly focus on core AI technologies, other related technologies like those from the field of IoT or Augmented Reality should be mappable, but the framework should make clear that these are not core AI technologies.
- The structuring of the wide field of AI use cases and technologies should be as technologically sound as possible.
- The final structure should be integrable with existing frameworks and structured, especially those used in policy-making processes on EU-level.

As explained in section 1, the taxonomy shall be used to support and steer decision-making processes and therefore information must be mapped in a technically correct, coherent and consistent way. Conclusions derived from the analysis of existing frameworks highlights this as well. One of the most important problems currently is the low consistency and coherence of the mapping criteria. This relates to the selection of high-level mapping dimensions as well as the composition of individual dimensions. Furthermore, the taxonomy needs to cater to a broad range of users with different purposes. Consequently, this requires comprehensiveness and a high degree of flexibility. The taxonomy on the one hand must cover the whole technology field of AI and the adjacent technologies extensively. On the other hand, include flexible elements which can extend and adapt the existing static structure.

To ensure that these criteria are met, the following steps have been followed:

1. Discussion of dimensions used in existing frameworks to decide on whether those should be part of the final framework
2. Expert input on potentially missing dimensions, that have not yet been addressed by existing frameworks
3. For each dimension: In-depth analysis of categories and discussion of:
   a Review of existing "standard approaches" to cluster a dimension (e.g., for distinguishing different industries), apart from those used in the frameworks mentioned above.
   b. In-depth discussion of existing approaches for each dimension to decide whether to go with established approach or self-develop a new structure.

For the dimensions used in the frameworks above, the results of this process will be described briefly in the following section.

2.2.1 Core findings per dimension

- **Industries**

The dimension "industries" is used in all but two company focused frameworks and once for technologies and use cases. The selected subcategories differ between publishers and with the same publisher when comparing landscapes from consecutive years. The following table 2 depicts the selection of industry subcategories by CB Insights from 2017 to 2020 and shows how strongly the selection changes from year to year. To fulfil the requirements for a comprehensive taxonomy, the industry dimension would need to be consistent.

- **Enterprise functions**

The dimension enterprise functions mainly consists of different sets of customer-facing functions such as Marketing, Customer Support or Sales and horizontal support functions, e.g. IT and HR. Functions directly related to the product (e.g. procurement, R&D, production) are not covered as

Name	Categorized Object	Customer Focus (B2B/C)	Geography	Funding Stage	Ecosystem Roles
Daxue Consulting - China AI Landscape	Companies				
Element AI - Canadian AI Startups	Companies		x		x
Firstmark Data & AI	Companies				
Linux Foundation - AI Tech Ecosystem Landscape	Companies				x
appliedAI Startup Landscape	Startups				
Bloomberg Beta	Startups				
CB Insights 2017	Startups				
CB Insights 2018	Startups				
CB Insights 2019	Startups				
CB Insights 2020	Startups				
CB Insights Agriculture	Startups				
CB Insights Healthcare	Startups				
CB Insights Retail	Startups				
Cognite Ventures DL	Startups		x		
MMC Ventures UK AI	Startups	x		x	
StartHub Computer Vision	Startups				
StartHub Israel	Startups				
StartHub NLP	Startups				
3XN - AI Taxonomy	Technology				
AI & Intelligent Automation Network - AI Classification	Technology				
Algorithmia - AI & ML	Technology				
appliedAI Use Case Cards	Technology				
appliedai.com					
Search platform	Technology				
Classification of AI - Ajit Nazre & Rahul Garg (Investors)	Technology				
Deloitte - AI Definition	Technology				
Forbes - AI Knowledge Map	Technology				
McKinsey - AI Capabilities	Technology				
McKinsey Mapping	Technology				
PWC - AI layers	Technology				
Small Data Group - Taxonomy of AI + ML	Technology				
Taxonomy of Machine Learning Based Anomaly Detection and its suitability	Technology				
Thomson Reuters - AI	Technology				
Papers with Code - Tech mapping	Technology				
AI and Climate Paper	Use Cases				
Plattform Lernende Systeme	Use Cases				

Table 2: Overview of AI frameworks - clustering dimensions 5 - 8..

CB Insights 2020	CB Insights 2019	CB Insights 2018	CB Insights 2017
Agriculture & Food	Agriculture	Agriculture	Autonomous Driving
Construction	Autonomous Driving	Autonomous Driving	Commerce
Education	Finance & Insurance	Commerce	Fintech & Insurance
Energy	Government	E-Sports	Healthcare
Finance & Insurance	Healthcare	Education	
Government & City Planning	Industrials	Fintech & Insurance	
Healthcare	Legal, Compliance & HR	Healthcare	
Legal	Media	Life Science	
Manufacturing	Real Estates	News & Media	
Media & Entertainment	Retail	Physical Security	
Mining	Semiconductor	Sports	
Real Estate	Telecom	Travel	
Retail & Warehousing			
Telecom			
Transportation			

Table 3: Overview of Industry Categorizations.

exhaustive and are lacking completely in all analyzed frameworks except for MMC Ventures, Plattform Lernende Systeme.

- **AI Capabilities & Technical Infrastructure**

AI capabilities and technical infrastructure are not clearly differentiated by most frameworks and often grouped together under the label "Tech Stack" or similar high-level descriptions. Another problem is the high heterogeneity of the selected subcategories and the lack of exhaustiveness. Technology-focused frameworks cover in general more subcategories of AI capabilities but are not necessarily more technically correct and exhibit the same problems of heterogeneity and lacking mutual exclusiveness.

The following four categories are not used systematically for classifying AI related objects and appear only in one or two frameworks.

- **Customer Focus (B2C/B2B)**

The dimension customer focus appears only once in all 35 frameworks (MMC Ventures UK AI) and consist of the properties B2C and B2B.

- **Geographic Location**

The dimension geography is used twice. Element AI map companies of different regions within Canada. In contrast, Cognite Ventures focus on global regions and map startups onto three global regions (Asia, Americas and Europe, Middle East, Africa).

- **Funding Stage**

The dimension funding stage is included only in the MMC Ventures UK AI landscape. Four funding stages are differentiated (angel, seed, early stage and growth).

- **Ecosystem Roles**

Ecosystem roles include different types of companies related to the provision of technical infrastructure for AI and non-technical support such as financing or education. Two frameworks, the Canadian Element AI landscape and the Linux Foundation Tech landscape, apply this categorization.

3. A CONSOLIDATED FRAMEWORK FOR MAPPING THE EUROPEAN AI LANDSCAPE

3.1 INTRODUCTION

Based on the observations during the investigation of existing frameworks and the respective findings discussed in the previous chapter, we will now suggest a new framework for mapping the AI landscape that addresses the shortcomings mentioned above and, as already noted, tries to cover the goals of:

- Covering all relevant dimensions;
- The categories per dimensions should be mutually exclusive as well as collectively exhaustive;
- Technology-related dimensions should clearly focus on core AI technologies;
- Ability to map related infrastructure and enabling technologies;
- Being integrable with existing frameworks on EU-level.

The proposed framework can be described by answering the following two questions:

1. **How can elements be mapped?** This refers to which dimensions are being used to structure the actual landscape (e.g., organizations, technologies and use cases) and which categories exist per dimension. This structure is generally static and does not change over time. Please not that to make sure further structuring opportunities can be added over time we do not limit the dimensions to those described, but allow more dimensions to be introduced, however, once integrated the dimension as well as the categories it entails are static. We will refer to this part of the suggested mapping approach as the Structural Framework.

2. **What elements are actually mapped?** Once it has been decided which dimensions and categories per dimensions structure the space of AI-related entities, the next question is which content elements are of relevance and should

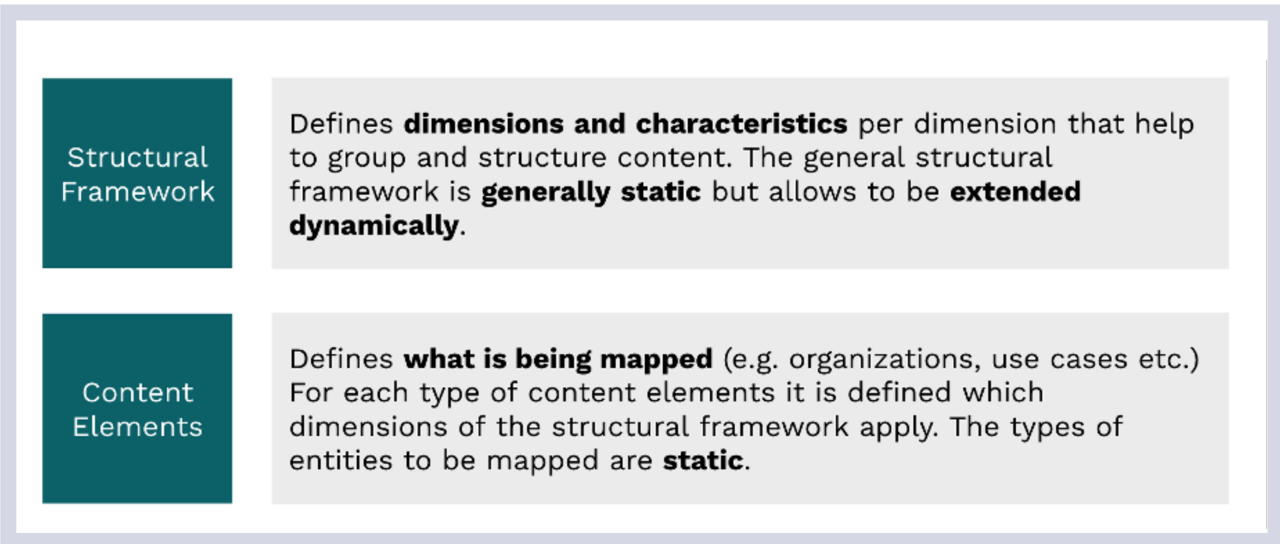


Figure 1: Structure of the taxonomy.

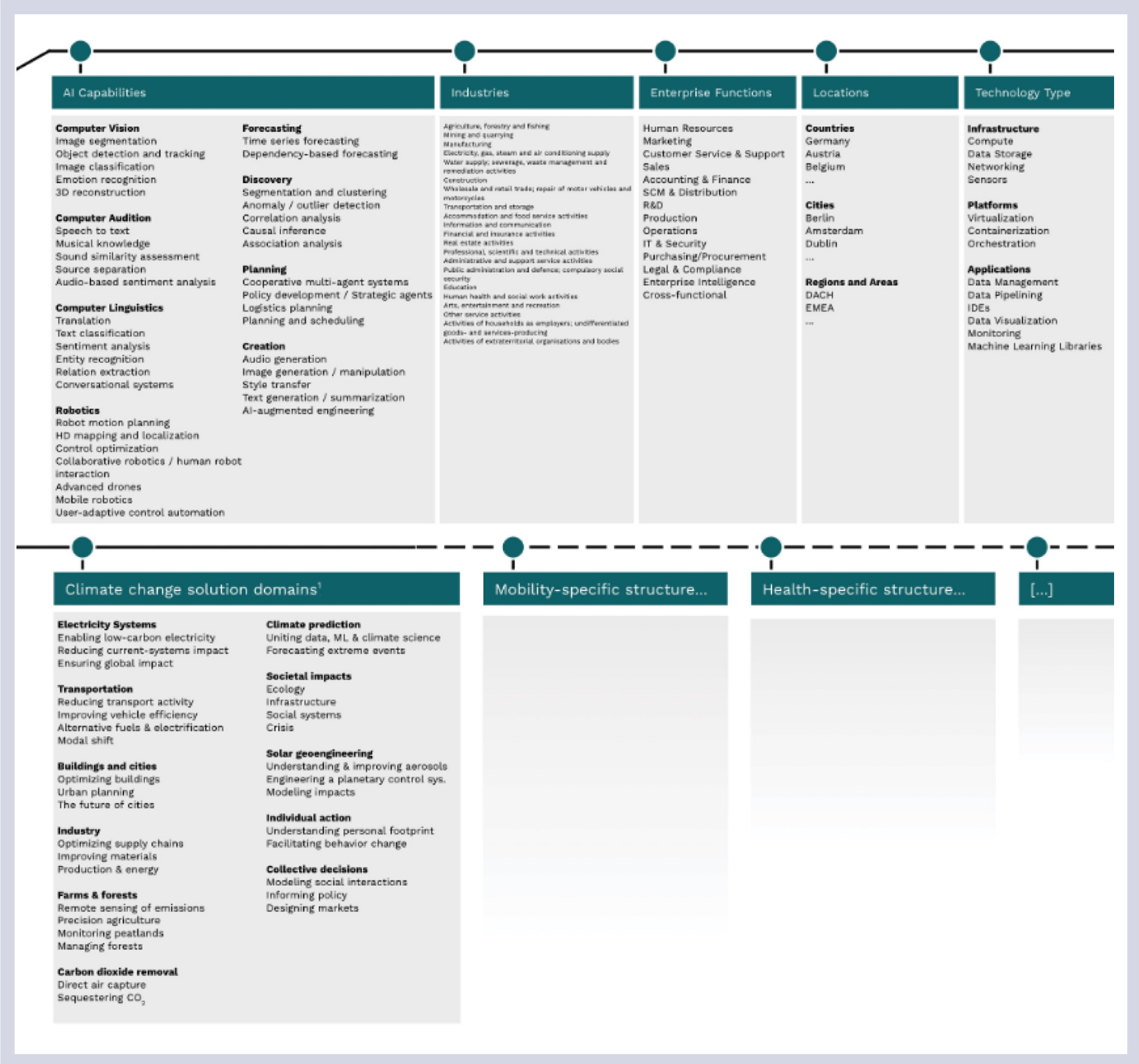


Figure 2: Full structure of the taxonomy including all subcategories.

therefore be mapped to the structural framework. We have already touched upon a few in section 1, like organizations or use cases, but we will further describe which entities will be mapped to this framework and which aspects of the structural framework apply to each in more detail.

These entities are what is ultimately being "filled into" the structural framework, meaning it should reflect the current state of the real AI landscape as good as possible and will therefore be dynamic in its nature. Entities can be added and removed anytime if necessary. We will refer this aspect of the mapping as Content Elements

The following chapter will first introduce the structural framework (3.2) and then the content elements and how they are linked to the structural framework in section 3.3.

3.2 Structural framework

The structural framework answers the question "How can elements be mapped?" by defining dimensions and characteristics per dimension that help to group and structure content. The structural framework is generally static but allows it to be extended dynamically. The following dimensions are static:

- Industries
- Enterprise Functions including Definitions
- Locations
- AI Capabilities
- Enabling Technology Types

and can be extended dynamically by domain-specific categories.

Domain-specific categories are custom mapping dimensions which are added for individual domains and enable the integration of additional characteristics apart from industry, enterprise function, location, AI capability and enabling technology type. The following visualizations show an overview of all dimensions and categories covered by the structural framework.

3.2.1 Industries

The choice for the final list of industries used in the structural framework was mainly driven by the strong necessity of applying standards and ensuring that the framework can be integrated into existing structures at EU-level. We therefore decided to base the list of industries on the NACE code list. NACE stands for "nomenclature statistique des activités économiques dans la Communauté européenne" and refers to Statistical Classification of Economic Activities in the European Community. By using an established structure usability, comparability and transferability can be maximized.

With regards to companies, industry can be considered in two ways. When classifying content elements of the type company, it is important to keep in mind the difference between the industry served by a company's offering and the industry classification of the company itself. A startup offering a software for analyzing medical images serves the healthcare sector but the company itself would be classified as "Information and communication"-company as it is programming and selling software. For the attached database architecture, both aspects are taken into consideration.

The following list depicts the top-level NACE code structure:

3. Agriculture, forestry and fishing
4. Mining and quarrying
5. Manufacturing
6. Electricity, gas, steam and air conditioning supply
7. Water supply; sewerage, waste management and remediation activities
8. Construction
9. Wholesale and retail trade; repair of motor vehicles and motorcycles
10. Transportation and storage
11. Accommodation and food service activities
12. Information and communication
13. Financial and insurance activities
14. Real estate activities
15. Professional, scientific and technical activities
16. Administrative and support service activities
17. Public administration and defence; compulsory social security
18. Education
19. Human health and social work activities
20. Arts, entertainment and recreation
21. Other service activities
22. Activities of households as employers; undifferentiated goods- and services-producing
23. Activities of extraterritorial organisations and bodies

3.2.2 Enterprise Functions

The dimension enterprise functions covers all business activities related customer-facing functions, horizontal support functions and functions directly related to the product creation process. The here presented structure is largely based on the existing set of enterprise functions developed by Plattform Lernende System with small adjustments regarding subcategory names and composition. In contrast to the original set developed by Plattform Lernende Systeme, Sales and Legal & Compliance constitute independent subcategories.

The subcategory planning has been renamed to the more general operations. Additionally, two new subcategories have been added: Enterprise Intelligence and cross-functional. Enterprise Intelligence comprises all data and analytics activities that are based on aggregating and descriptively analyzing corporate information. The results of these kinds of analyses are usually presented or consumed as some kind of report or dashboard showing the results-related aspects and the subcategory cross-functional can be used for mapping elements related to more than one enterprise function:

1. Human Resources
2. Marketing
3. Customer Service & Support
4. Sales
5. Accounting & Finance
6. SCM & Distribution

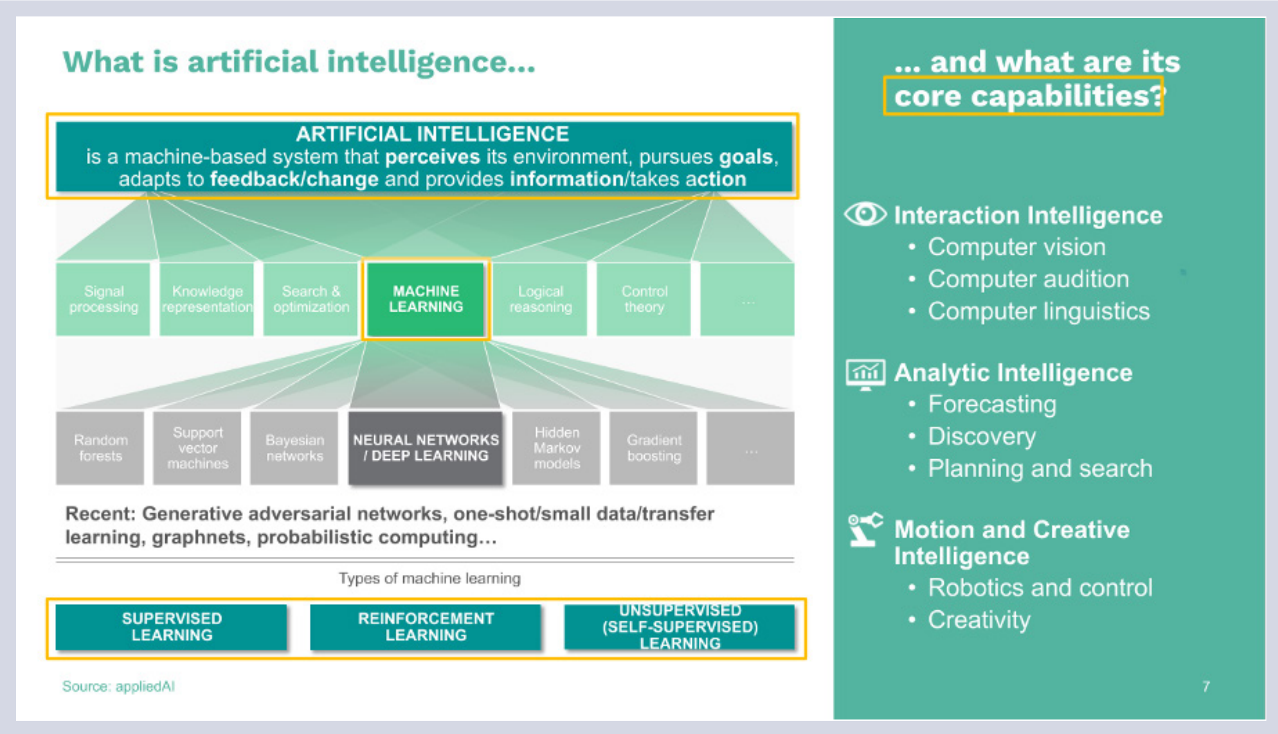


Figure 3: Categorization of AI technologies into capabilities.

7. Research & Development
8. Production
9. Operations
10. IT & Security
11. Purchasing and Procurement
12. Legal & Compliance
13. Enterprise Intelligence
14. Cross-functional

3.2.3 Geographic Locations

The dimension locations describes where a specific element is active or executed and consists of the three subcategories:

- Countries
- Cities
- Regions, e.g. DACH, Benelux. Regions are thereby conceived as a “set of countries”, so one region is always made up of multiple countries.

3.2.4 AI Capabilities

AI capabilities describe basic problem types AI is able to cope with and that can be applied in use cases. They have been derived as an alternative to the diverse and mostly technically incorrect clusterings used in existing frameworks. In general, the AI capabilities used as a structure in this approach can be sorted into eight broader categories, each with certain subcategories. The following table provides an overview and a few examples of what these eight capabilities refer to and provides some examples.

The first three capabilities comprise skills related to the processing of images, audio data and natural language - Computer Vision, Computer Audition and Computer Linguistics. The fourth, Robotics, as already implied by the name, describes capabilities connected to the steering and control of physical systems, robots in particular. Forecasting refers to the analysis and forward-thinking of time series data. Discovery involves all analysis based on clusters and finding structures in datasets. Planning describes ML-approaches that allow to develop long-term strategies and multi-step plans to solve problems.

The last capability, Creation, describes the creation of new data with ML systems, meaning the creation of images or audio data or applying machine learning to augment disciplines like engineering in coming up with new structures or approaches that humans would most likely not think of. More information about the eight capabilities and examples are given below.

Capability	Computer Vision	Computer Audition	Computer Linguistics	Advanced robotics and control
Description	Enables machines to make sense of visual data by recognizing objects or persons and understanding the semantics of images or video sequences.	Enables machines to process and make sense of audio signals by recognizing, classifying and understanding audio signals.	Enables machines to process, interpret and potentially "understand" the content of language (e.g. human) as well as interact in such a language with a desired semantic content.	Enable machines to analyze, interpret and learn from data representing physical systems and use such data to change the systems' behaviour (e.g. control of machines).
Exemplary Use Case	* Object detection, identification, and description * Pose detection and recognition * Image captioning segmentation and generation * Text recognition in complex scenes	* Speech to text and text to speech * Music recognition or ambient sound removal * Machine monitoring and anomaly detection * Autonomous navigation	* Semantic text understanding, translation, and summarization * Bot interaction and conversation * Sentence and text generation	* System, machine and plant control and optimization * Motor control and trajectory planning for robots * Autonomous robots
Input Data	Typically image and video (sometimes streams).	Typically audio streams, e.g. environmental noises, spoken language.	Typically digital (often structured) text or language representations.	Typically structured sensor, actuator and environment data as well as feedback / result values of actions.

Capability	Forecasting	Discovery	Planning	Creation
Description	Machines are enabled to find complex patterns in data and use them to make predictions about the future course of a time series or the likelihood of certain events that may happen (e.g., defaulting on a credit).	Machines can process huge amounts of data and find, for example, patterns, logical relationships, similarities, groups or dependencies in them.	Machines can look for optimal solutions to problems requiring long sequences of actions in complex environments, where the search space is too vast for human or even exact computer solutions.	Machines can generate images, speech, or music based on examples they have been shown beforehand.
Exemplary Use Case	* Demand prediction * Price prediction * Prediction of energy consumption * Default prediction * Customer lifetime value estimation	* Customer segmentation * Recommender systems (e.g. through collaborative filtering) * Anomaly (e.g. fraud) detection	* Optimization of project plans * Optimal routing * Molecule search / material design	* Image style transfer * Creation of fake images and video * Music generation * Making realistic photos out of rough descriptions of what should be on them
Input Data	Historical data on the trend and patterns in the development of certain variables and / or information on external factors that have some relation to and therefore help to explain or predict the target variable.	Can work with various types of input data. May comprise customer and sales data, machine sensor data or audio streams.	Problems that have been formalized in a way that can be solved by search algorithms. The result of such a "search" can be the optimal strategy to play a game, molecule structures with certain properties or optimal plans.	Examples of the type of data to recreate, for example music composed by Bach or images of celebrity faces, landscapes etc.

Figure 4: Description and examples of AI Capabilities.

For each capability, respective subcategories have been derived based on multiple rounds of in-depth discussions with a range of AI experts. The subcategories allow more detailed structuring of elements, but also help to provide an impression of which type of content falls into which category. It thereby helps to focus on core AI technologies and leave out related or enabling technologies:

3.2.5 Enabling Technology Types

As described above, the focus of the mapping is to structure core AI technologies. However, AI as a technological field is deeply integrated with other related technologies that fuel developments in the field of AI, these should not go unconsidered. We therefore introduce a special part of the structural framework to account for this. We refer to this category as Enabling Technology Types. This category is intended to cover four broader categories of technologies: Infrastructure, platforms, frameworks and applications.

- **Infrastructure** refers to technologies that provide basic compute, data storage or capturing as well as networking technologies, that are commonly used for building AI use cases. Examples are cloud infrastructure providers or specialized hardware (e.g., GPUs) for running modern machine learning models.

- **Platforms** refer to software platforms that are used to manage AI model training or operations, maintain data assets or pipelines or steer the development process of machine learning applications. Examples are Amazons Sagemaker, IBM Watson Studio, Domino Data Lab, Kubernetes or Data Warehousing and Management Solutions.

AI Capabilities subcategories allow more detailed structuring of elements.	
Computer Vision Image segmentation Object detection and tracking Image classification Emotion recognition 3D reconstruction	**Forecasting** Time series forecasting Dependency-based forecasting
Computer Audition Speech to text Musical knowledge Sound similarity assessment Source separation Audio-based sentiment analysis	**Discovery** Segmentation and clustering Anomaly / outlier detection Correlation analysis Causal inference Association analysis
Computer Linguistics Translation Text classification Sentiment analysis Entity recognition Relation extraction Conversational systems	**Planning** Cooperative multi-agent systems Policy development / Strategic agents Logistics planning Planning and scheduling
Robotics Robot motion planning HD mapping and localization Control optimization Collaborative robotics / human robot interaction Advanced drones Mobile robotics User-adaptive control automation	**Creation** Audio generation Image generation / manipulation Style transfer Text generation / summarization AI-augmented engineering

- **Frameworks** refer to libraries, code repositories or technical resources that can be built upon in AI or ML development and that provide certain necessary functionality, of which the goal is often to facilitate AI development and reduce effort of redundant implementation of the same functionality.

- **Applications** refer to software products that are commonly used during AI development or operations, for example to facilitate the setup of infrastructure, IDEs, data Visualization or monitoring applications.

All subcategories are given in the following. Detailed short definition of each category are listed in the table below:

A. Infrastructure
   a. Compute
   b. Data Storage
   c. Networking
   d. Sensors

B. Platforms
   a. Virtualization
   b. Containerization
   c. Orchestration

C. Frameworks
   a. Machine learning libraries
   b. Data visualization
   c. Data preparation and transformation
   d. Model management
   e. Others

D. Applications
   a. Data Management
   b. Data Pipelining
   c. Labeling
   d. IDEs
   e. Data Visualization
   f. Monitoring
   g. Others

Infrastructure	Infrastructure describes all hardware that is being used to run applications and software upon.
Compute	All hardware that is being used to run program code, meaning servers that house, for example, CPU or GPU-based compute hardware
Data Storage	Hardware that is providing storage for data. This category comprises various different storage technologies like hard disk drives, flash drives or others.
Networking	Networking appliances are used to interconnect other infrastructure elements. Examples are infiniband or ethernet networking via switches etc.
Sensors	Sensors refer to physical devices that allow capturing information about the real world. Examples are radar appliances, cameras for image capturing
Platforms	A platform describes some standardized layer of software, that provides unified interfaces and functionalities to build application software on top of it.
Virtualization	Virtualization solutions allow to abstract from physical hardware infrastructure, by simulating hardware assets via software solutions and therefore increasing flexibility.
Containerization	Containerization is defined as a form of operating system virtualization, through which applications are run in isolated user spaces called containers, all using the same shared operating system. A container is essentially a fully packaged and portable computing environment.
Orchestration	Orchestration describes the automatic assignment of hardware resources to software solutions. These solutions allow to abstract from actual hardware and resource requirements, which are handled automatically.
Frameworks	A framework is a collection of software modules that are not a usable end-user-software themselves, but allow programmers to build upon them by using them as apart of their final software solutions.
Machine learning libraries	Machine learning libraries provide functionalities for training machine learning models. They provide, for examples, languages to describe and handle statistical models and implement optimization approaches to train such statistical models based on some data.
Data visualization	Data visualization frameworks provide functionality to plot graphs and other means of data visualization. and other means of data visualization.
Data preparation and transformation	Data preparation and transformation libraries allow handling of large data resources, transforming them and running simple analyses to derive insights about a dataset that are relevant in the later machine learning phase.
Model management	Model management frameworks allow to track training of models and handle models in production (i.e., storing and loading them, deploying vs. undeploying them etc.)
Others	This category covers all other frameworks that focus on more general tasks, but in the end are also being used in the development of machine learning applications (e.g., to develop user interfaces)
Applications	Applications are software artifacts that are designed and being used by end-users, that are not programmers themselves.
Data Management	Data management applications support users in handling and working with large datasets. They provide a graphical user interface to run basic data transformation and transferring.
Data Pipelining	Data pipelining applications allow to define the flow of data between different applications and software solutions across the organization.
Labeling	Labeling applications support users during data labeling, meaning they provide an easy and GUI-based approach to annotate data and might also support data annotation by pre-labeling functionalities.
IDEs	IDEs are software solutions that are being used to program software. They provide text editor functionalities including support for software testing etc.
Data Visualization	Data visualization applications allow users to create graphs and plots in a GUI-based fashion.
Monitoring	Monitoring applications help users to track the outputs of their machine learning models and keep track of model quality.

Table 4: Overview of industry categorization.

3.2.6 Domain-specific extensibility

Depending on the individual requirements, domain-specific dimensions can be integrated into the framework. These custom mapping categories are sets of characteristics and criteria, which can be added to the framework and extend the pre-defined static dimensions onto which then the content elements can be mapped. This enables the analysis of AI in the context of individual domains or jointly with other problem fields.

For example, it is possible to analyze the potential opportunities of AI in the context of climate change by extending the set of dimensions with a domain-specific category. In this case, we can use the climate change solution domains defined in “Tackling Climate Change with Machine Learning”, a paper published by Rolnick, David; Donti, Priya; Kaack, Lynn; Kochanski, Kelly; Lacoste, Alexandre; Sankaran, Kris et al. (2019) as the basis for a new dimension. The individual climate change solution domains constitute the new subcategories of this dimension.

Climate change solution domains with subcategories as an example for a domain-specific dimension:	
Electricity Systems Enabling low-carbon electricity Reducing current-systems impact Ensuring global impact	**Carbon dioxide removal** Direct air capture Sequestering CO_2
Transportation Reducing transport activity Improving vehicle efficiency Alternative fuels & electrification Modal shift	**Climate prediction** Uniting data, ML & climate science Forecasting extreme events
Buildings and cities Optimizing buildings Urban planning The future of cities	**Societal impacts** Ecology Infrastructure Social systems Crisis
Industry Optimizing supply chains Improving materials Production & energy	**Individual action** Understanding personal footprint Facilitating behaviour change
Farms & forests Remote sensing of emissions Precision agriculture Monitoring peatlands Managing forests	**Collective decisions** Modelling social interactions Informing policy Designing markets

Content elements, here for example use cases, can be mapped onto these subcategories indicating that a certain use case supports a certain climate change solution domain comparable to an industry or enterprise function.

3.3 CONTENT ELEMENTS

Content Elements define what is being mapped (e.g. organizations, use cases etc.). For each type of content elements, it is defined which dimensions of the structural framework apply. The types of entities to be mapped are static.

The following types of Content Elements can be mapped and will be explained in detail in the following sections:

- Use Cases
- Organizations
- Financing Opportunities
- Educational Offerings
- Technology Solutions
- Domain-specific Scenarios
- Regulations

3.3.1 Use Cases

AI use cases are a clearly defined set of activities designed to solve a specific problem from a business or customer perspective, in which one or more AI technologies are involved in solving the respective problem.

For example, agricultural companies face the problem of a very volatile and complex environment. They are impacted by the local weather conditions, global trade developments and demand volatility. Better prediction quality or a longer prediction time frame can help with managing these factors. A possible use case could therefore be "Prediction of crop yield & demand" powered by the AI capabilities computer vision and forecasting.

A use case can be involved in a scenario, receive financing, use a technology solution and can be supported or applied by organizations.

3.3.2 Organizations

Organizations are organized groups of people with a particular common purpose. They can offer products or services related to AI on a for- or non-profit basis or aim at advancing the field of AI in general.

We have defined and explained the five following subcategories in detail:

- **Startups**: new companies offering an innovative product or service.
- **Corporations:** organizations offering a service or product on non- or for-profit basis.
- **Associations:** group of individuals or other organizations with a common interest and goal, e.g. EurAi.
- **Research Institutions**: independent or academic organizations with focus on research in AI-related fields, e.g. German Research Center for Artificial Intelligence.
- **Standardization and regulation institution**: organization working on developing regulations and setting standards for AI, e.g. IEEE.

Organizations have direct relations to all other content elements except domain-specific scenarios.

3.3.3 Financing Opportunities

Financing opportunities describe systematic investment activities into AI. Examples are among others funds for research projects, startup grants, public or corporate investments or tax reductions for particular innovation activities. The financing vehicles to be mapped here should specifically focus on driving innovation in the AI space or related to enabling technologies for AI. Each financing opportunity is related to organizations in two ways: They are provided by an organization - typically a governmental institution or industry players - and they are received by other players - typically industry or ecosystem players.

We distinguish between to different general financing approaches as mentioned above:

- **Taxation-based financing** - referring to tax reduction measured to create incentives for AI innovation projects.
- **Funding-based financing** - referring to special funding that is being offered to support AI innovation projects.

Financing opportunities can either be provided or received by organizations and can fund use cases and/ or domain-specific scenarios.

3.3.4 Educational Offerings

Educational offerings are products or services aimed at educating the audience about AI.

- Online Training, e.g. Elements of AI
- Offline Training, e.g. Mittelstand Digital - AI Trainings
- Whitepapers and Reports
- Books

Educational offerings are offered by organizations.

3.3.5 Technology Solutions

Technology solutions are all kinds of technological assets, tools or systems that are being used to develop and run AI solutions. They fall in one of the categories described under enabling technology types (see chapter 3.2.5). These classifications already provide a good overview of what types of technologies can fall under the term technology solution. I.e. those might be hardware assets, software platform, individual application or frameworks. Each technology solution can have a provider (usually a company that is developing and offering the solution and can be used in AI use cases (e.g., a particular image analysis system to detect draughts might run on NVIDIA graphics cards.

Technology solutions can either be provided or used by organizations and can be used by use cases.

3.3.6 Domain-specific Scenarios

Domain-specific scenarios allow for the inclusion of more complex multi-dimensional problems specific to the respective domain. Scenarios consist of different individual problems which can be solved by AI and the description of a solution approach based on concrete use cases. Mapping domain-specific scenarios offers the opportunity to depict the complexity of real-world problems adequately For example, when mapping organizations realizing climate-focused use cases the lens of domain-specific scenarios can help make visible the multiplicity of possible solutions for a problem and the interplay between seemingly distinct use cases.

Scenarios involve different use cases and can receive financing opportunities.

3.3.7 Regulations

Regulations describe legal or regulatory requirements, frameworks or guidelines, that affect the application or use of AI technologies in particular areas. Examples are The EU Product Liability Directive or GDPR. Such regulations can be related to particular use cases (e.g., related to the banning of facial recognition technologies in the EU) or application scenarios. In the latter case, regulations might not only be related to

AI in particular, but can also be related to the actual scenario (e.g. climate-related) but still have impact on AI technologies applied in that area. Draft documents or discussion papers are not counted as regulations.

Regulations can influence use cases and/or scenarios as well as single AI capabilities.

3.4 INFORMATION ARCHITECTURE

3.4.1 Overview of Database Structure

The structure described above has been implemented in a relational database schema. The schema is a technically implemented reflection of the structure, content and relations that has already been described. This section therefore focuses on the technical aspects of how the schema has been realized in a technical manner. The following figure gives a high-level overview of the database schema.

The resulting database schema contains four general types of tables:

- Tables that store the different elements of the structural framework. These are marked with the prefix "STRUCT" in the implemented database.
- Tables that store content elements. These are marked with the prefix "CONT".
- Tables that store many-to-many relations (relations) between content elements. These are marked with the prefix "REL"
- Tables that carry additional background information and are necessary to ensure integrity of the database schema without being at the core of it. These tables are marked with the prefix "UTIL"

The list on table 5 gives an overview of all tables in the database schema and their purpose.

Due to the large number of database tables we will not list all columns here individually. Instead, we would like to highlight two aspects:

1. The structure of the tables storing the contents of the framework (those marked with the prefix "CONT" as they reflect what can be stored in the database scheme about each type of content element, and
2. How content elements are linked to the structural framework.

The other tables follow the basic rules of relational database design and therefore can be interpreted based on the entity relationship diagram given in figure 6.

The tables storing content elements have the following columns (which reflect the attributes that are stored for each content element. Please note that apart from these attributes there is also the information stored in the "MAP"-tables, meaning the mapping of content elements to the structural framework as well as relationships between the content elements in the "REL"-tables.

To link these content elements to the structural framework, the tables carrying the "MAP" prefix are being used. As there

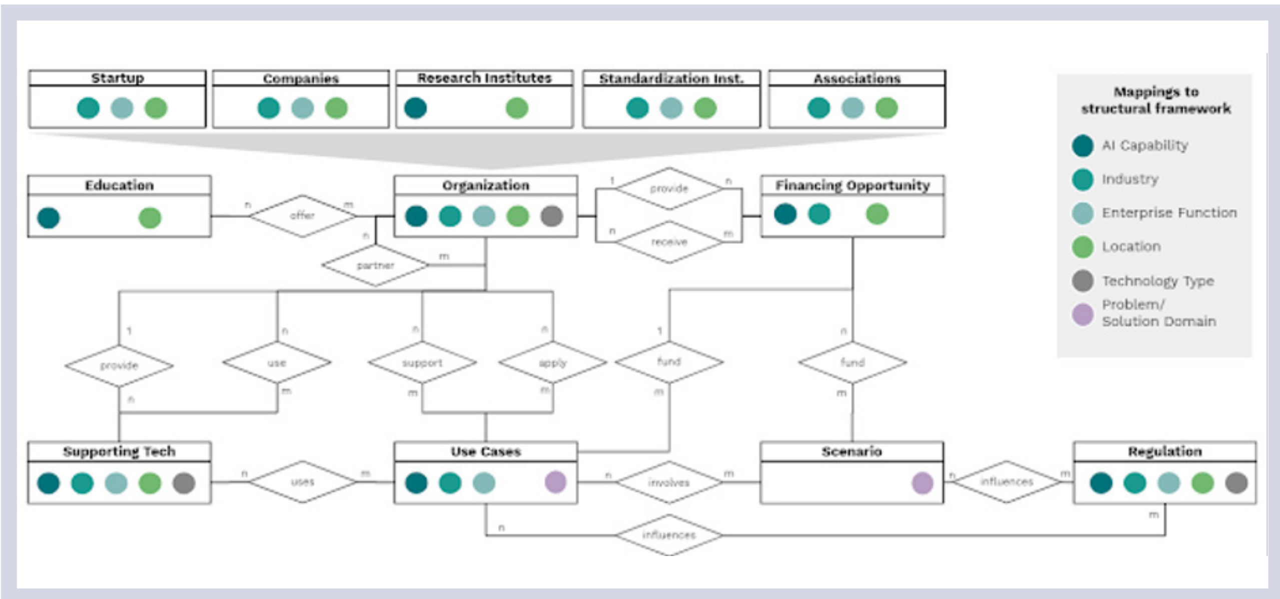


Figure 5: Overview of database structure.

Table	Description
CONT-Education	Stores information about educational offerings
CONT-Financing	Stores information about financing vehicles
CONT-Organizations	Stores information about organizations
CONT-Regulation	Stores information about regulations
CONT-Scenario	Stores information about scenarios
CONT-Supporting_Tech	Stores information about supporting technologies
CONT-UseCases	Stores information about AI use cases
MAP-Capabilities	Stores information about the assignment of content elements to different AI capabilities
MAP-Climate_domains	Stores information about the assignment of content elements to different climate change solution domains
MAP-Enterprise_Functions	Stores information about the assignment of content elements to different enterprise functions
MAP-Industries	Stores information about the assignment of content elements to different industries
MAP-Locations	Stores information about the assignment of content elements to locations (locations can be Areas, Cities or Countries)
MAP-Technology_Type	Stores information about the assignment of content elements to different technology types
REL-Financing-funds-Scenario	Stores information about which scenarios are funded by which financing vehicles
REL-Organizations-apply-UseCase	Stores information about which Use Cases are applied by which organizations
REL-Organizations-offer-Education	Stores information about which organizations offer which educational content
REL-Organizations-partner-Organizations	Stores information about which organizations partner with each other
REL-Organizations-receive-Financing	Stores information about which organizations receive which financing vehicles
REL-Organizations-support-UseCase	Stores information about which organizations support the development of which use cases or offer products that help implement this use case
REL-Organizations-use-Tech	Stores information about which organizations use which technologies
REL-References	Stores information about which references provide more background information about a particular scenario
REL-Regulation-influence-Scenario	Stores information about which regulations have influence on which scenarios
REL-Regulation-influence-UseCase	Stores information about which regulations have influence on which use cases
REL-Scenario-involves-UseCase	Stores information about which use cases are involved in which scenarios
REL-UseCase-uses-Tech	Stores information about which technologies are used to implement different AI use cases
STRUCT-EXT-Climate_Domains	Stores information about climate change solution domains
STRUCT-Capabilities	Stores information about AI capabilities
STRUCT-Capability_Clusters	Stores information about clusters of AI capabilities (e.g, Computer Vision or Computer Audition). Each entry in the table STRUCT-Capabilities is linked to one entry in this table
STRUCT-Enterprise-Functions	Stores information about different enterprise functions
STRUCT-Industries	Stores information about different industries
STRUCT-Location_Areas	Stores information about areas
STRUCT-Location_Cities	Stores information about cities. Each entry links to one entry in the table STRUCT-Location_Countries indicating which country a certain city belongs to
STRUCT-Location_Countries	Stores information about countries
STRUCT-Location_Countries_in_Areas	Stores information about which countries are part of which areas
STRUCT-Technology_Clusters	Stores information about technology clusters. These refer to broader categories of technologies like hardware, platforms, frameworks or applications
STRUCT-Technology_Types	Stores information about types of technologies used for AI development. Each entry in this table is related to one technology cluster (see table "STRUCT-Technology_Clusters)
UTIL-References	Stores information about references that are referred to

Table 5: Database schema.

Scenarios	Use Cases	Organizations	Education	Financing	Supporting Technology	Regulation
id	id	Startups Companies Research Institutes Standardization Inst. Associations id	Online Offline Training Whitepaper and Reports Books id	Academic financing programs Industry financing programs id	id	id
scenario_name	usecase_name	organization_name	ecuation_name	financing_name	tech_name	regulation_name
problem_description	description	organization_type (see above)	description	type (see above)	description	description
solution_approach	image_link	website	target_group	description	software (yes/no)	link
	high_risk (yes/no)	logo_link	publicly_available (yes/no)	prerequisites	hardware (yes/no)	
	funded _by (->organization)	description	free_of_charge (yes/no)	submission_deadline	provider (->organization)	
		b2b (yes/no)	author	funding_period_start		
		b2c (yes/no)	education_type (see above)	funding_period_end		
		non-profit (yes/no)	duration	link		
		hardware (yes/no)	link	provided_by (->organization)		
		software (yes/no)		financing_period_end		
		hq_city (->city9		link		
		industry (->industry)				

Table 6: List of attributes per content element.

are quite many relationships between structural framework and content elements to consider, a special approach was being used for the mapping: For each element of the structural framework there is exactly one table to map this structural element to all possible content elements. This means each "MAP"-table is structured as follows: Each table starts with one column that refers to the actual characteristic of the dimension of the structural framework. For example, in the MAP-Capabilities table the first column might refer to "Object detection". Apart from that there is one column for each type of content element that an AI capability could be mapped to. This means only one table is required to be able to map in this case AI capabilities to all the different content elements that this dimension of the structural framework applies to.

3.4.2 Using the structure to derive insights

The suggested database structure can now be used to explore the European AI landscape. This is supported by the relations reflected in the schema described above. In the following, we would like to give an example of two of many conceivable questions that could be addressed based on the structure. Technically, these requests to the database can be implemented with simple SQL-Queries towards the database.

Example 1: Fanning out use cases and organizations related to a scenario

To identify all organizations that are related to a scenario one would traverse the database schema as follows: First, you start with selecting the scenario of interest from the table "CONT-Scenarios" and read the column "id" from this table. Via the relation table "REL-Scenario-involves-UseCase" it is possible to find all related Use Cases. This table has two columns: "scenario" and "usecase" which contain id's of entries in the tables "CONT-Scenarios" and "CONT-UseCases". This means by selecting all rows from this table, where the scenario-column equals the id of the Scenario of interest and getting the respective id's for Use Cases, one is able to select all these cases from the use case table. At this point all Use

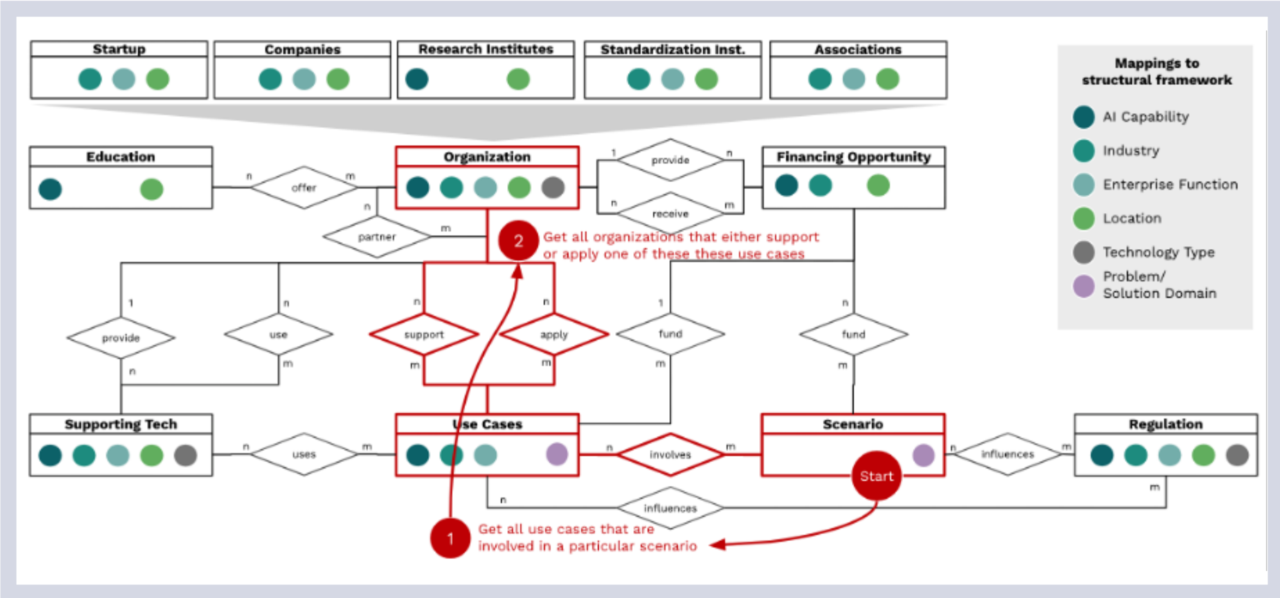


Figure 6: Example 1 for applying the outlined structure.

Cases related to the scenario have been identified.

To now also read all Organizations related to a Scenario one would now take all the id's of the use cases identified in the first step and use these to search the tables "REL-Organizations-support-UseCase" and "REL-Organizations-apply-UseCase". These are the two relations that Use Cases and Companies can have, meaning that to identify all Organizations that are somehow related to a use case one can search for all identified Use Case ids in these tables and fetch the related id's of Organizations. These then allow to pull, e.g., names of organizations from the "CONT-Organizations" table. This approach is visualized on Figure 6.

Example 2: Technology leadership for a particular climate solution domain

Another exemplary use case could be identifying the organizations that provide the key technologies to impact a certain scenario in the climate domain.

Again, one would start to select the respective scenario to investigate from the "CONT-Scenarios" table and reading its id from the id column. This allows - as in the aforementioned example to get all related Use Cases by using the "REL-Scenario-involves-UseCase" table. The list of id's that results is then used to find all technologies used in these use cases

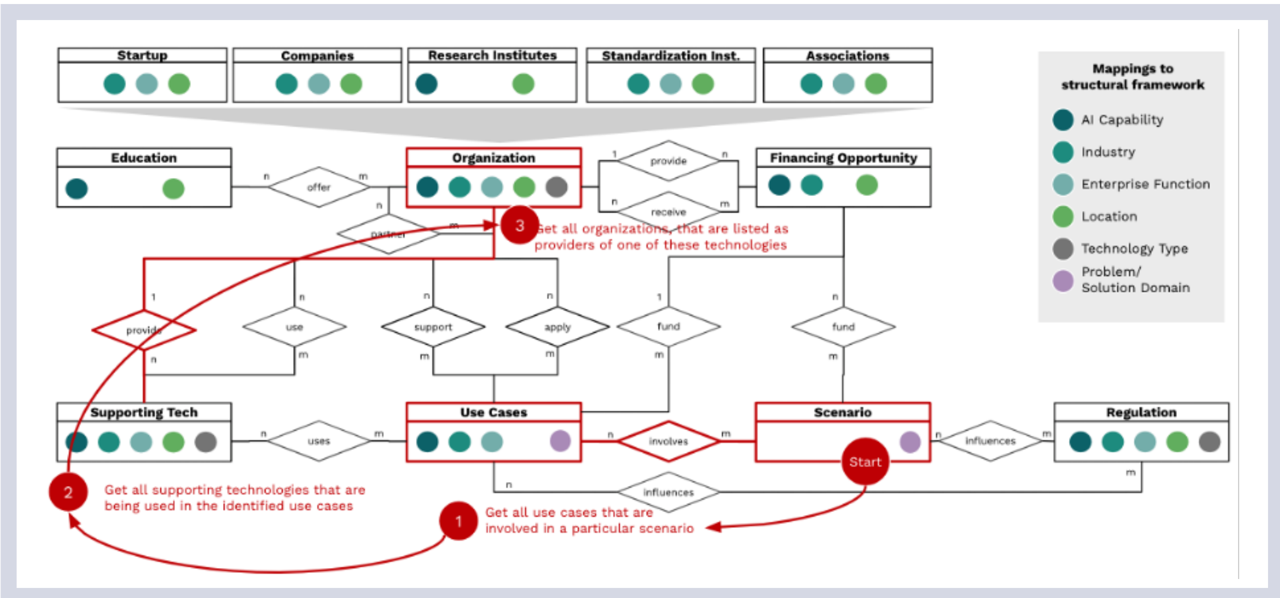


Figure 7: Example 2 for applying the outlined structure.

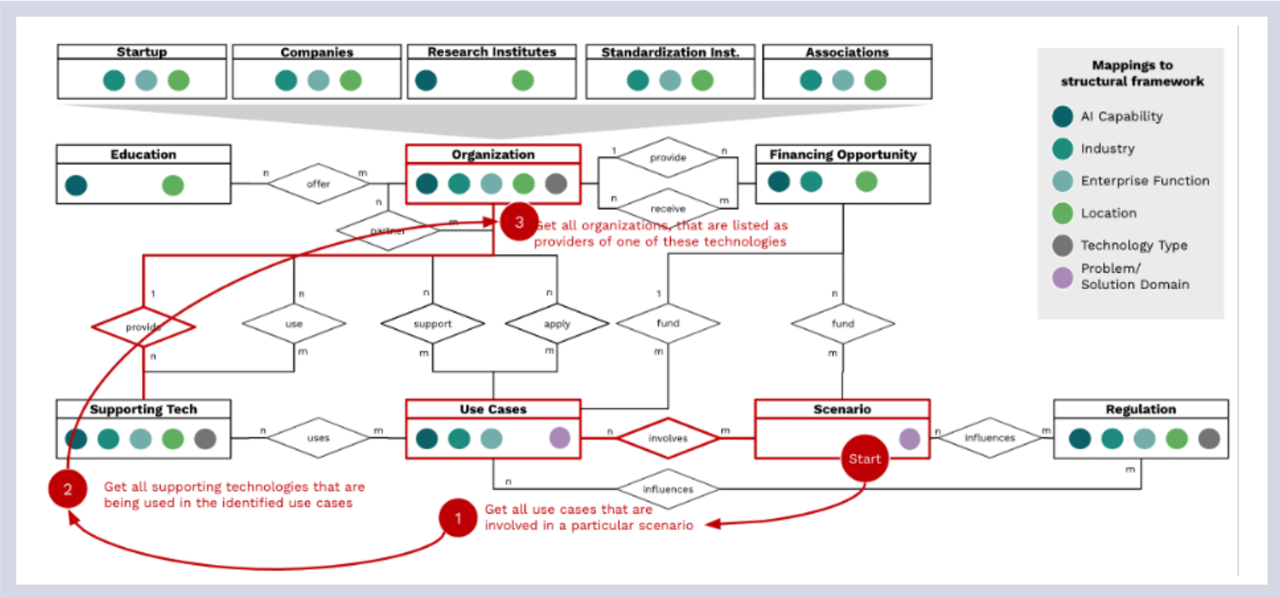


Figure 8: Example 3 for applying the outlined structure

by finding all id's of entries in the table "CONT-Supporting_Tech" that are listed next to one of the identified use case id's in the "REL-UseCase-uses-Tech" table. This would result in a comprehensive analysis of the use cases for a selected scenario and its underlying organizations and technologies. Filtering for those gives us all providers of technologies that are used in a certain scenario to fight, e.g., climate change.

Example 3: Assess impact of Regulation on Scenarios

A third exemplary use case could be to analyze existing or potential regulation to a specific scenario and the underlying use cases, technologies and organizations as outlined in the EIT Digital document A Policy Perspective.

Again, one would start to select the respective scenario to investigate from the "CONT-Scenarios" table and reading its id from the id column. This allows - as in the aforementioned example to get all related Use Cases by using the "REL-Scenario-involves-UseCase" table. The list of id's that results is then used to find all technologies used in these use cases by finding all id's of entries in the table "CONT-Supporting_Tech" that are listed next to one of the identified use case id's in the "REL-UseCase-uses-Tech" table. In addition, one would read the tables "REL-Organizations-support-UseCase" and "REL-Organizations-apply-UseCase".

In a final step, one would use the "REL-Regulation-influence-Scenario" table to identify relevant regulation. Based on the received information, a comprehensive analysis of the impact of existing regulation on new use cases or adjusted scenarios or alternatively an assessment of potential new regulation on existing use cases or scenarios can be made.

4. RETROSPECTIVE ON THE INTEGRABILITY OF OTHER LANDSCAPES AND STRUCTURES WITH THE NEW FRAMEWORK

A final aspect of the new architecture is it's compatibility with existing frameworks.

4.1 Exemplary selection & description of existing frameworks

CB Insights

CB Insights provides one of the most well known startups rankings. On a yearly basis, it assembles a list of the most promising 100 AI startups worldwide. The applied structural framework varies from year to year with regards to the domain and category selection with industries being covered most thoroughly and consistently. Enterprise functions and AI capabilities play a lesser role and subcategories within these fields are changing more often.

Bloomberg Beta

Bloomberg Beta is an early-stage venture capital firm backed by Bloomberg L.P issuing on an irregular basis a landscape covering the broader Machine Learning ecosystem. In comparison to most other landscapes, Bloomberg Beta includes a separate section for enterprise intelligence, which is divided into the different types of data the startups provide e.g. visual data.

StartHub Israel

StartHub.ai is the leading source for data regarding the Israelian AI ecosystem. The 2018 published landscape provides a very detailed overview of the AI startup ecosystem and nearly all relevant players in Israel. The landscape consists of subcategories from industries, enterprise functions, AI capabilities and supporting technologies. Specific industry verticals are represented more granular and in depth.

Plattform Lernende Systeme

Plattform Lernende System is a German consortium consisting of different working groups with the goal of shaping and promoting "learning systems" in Germany. They offer an interactive web-based landscape of the Germany AI ecosystem. The landscape can be filtered by application field, industry, enterprise function and AI capability and in contrast to most other publications depicts not only startups but also AI development projects, different types commercial providers of AI solutions and relevant users of AI use cases.

4.2 Compatibility analysis of selected dimensions

Compatibility analysis is conducted for the four most frequently applied dimensions in existing frameworks.

AI capabilities & Infrastructure

A direct mapping of the proposed AI capabilities with the respective capabilities of the selected existing frameworks is not possible. One reason is that the properties of subcategories vary too strongly to allow for a transformation of one framework into another. This is mostly due to heterogeneous levels of granularity within the subcategories. For example, CB Insights combines vision and speech processing in one joint category "NLP, NLG & Computer Vision" and Bloomberg Beta lists "Machine Learning" separately on the same level as "NLP" and "Autonomous Systems".

Another problem is that not all landscapes limit their technological capabilities to pure AI technologies. Nearly all frameworks do not distinguish between dedicated AI capabilities, adjacent technologies such as VR and enabling technologies

Capabilities	CB Insights	Bloomberg Beta	StartHub Israel	Plattform Lernende Systeme
Computer Vision			Computer Vision	Computer Vision
Computer Audition			Speech & Text	Speech & Text
Computer Linguistics		Natural Language	Speech & Text	Speech & Text
Discovery				
Forecasting				
Planning				
Robotics		Autonomous systems		Robotics & Autonomous Systems
Creation				
Enabling Technology Types	**CB Insights**	**Bloomberg Beta**	**StartHub Israel**	**Plattform Lernende Systeme**
Infrastructure				
Compute	AI Processors	Hardware	Hardware	
Data Storage		Hardware	Hardware	
Networking		Hardware	Hardware	
Sensors				Sensoric & Communication
Platforms				
Virtualization				
Containerization				
Orchestration				
Frameworks				
Machine Learning Libraries		Open Source Libraries		
Data Visualization				
Data Preparation and Transformation				
Model Management				
Others/ General				
Applications	AI Model Development	Agent Enablers; Research; Development	AI Enablers	
Data Management				Data management & Analysis
Data Pipelining		Data Science		Data management & Analysis
IDEs		Data Science		Data management & Analysis
Data Visualization		Data Science		Data management & Analysis
Monitoring				
REST	NLP, NLG & Computer Vision, Other R&D	Machine Learning Data Capture	AR/VR/ MR Bots & Biometrics Misc	Human-MachineInteraction and Assistive Systems VR & AR

Table 7: Comparison to existing frameworks.

which constitute the necessary infrastructure. As can be seen at StartHub Israel, their Technology cluster includes only two pure AI capabilities (Computer Vision and Speech) and mixes it with adjacent technologies such as Bots & Biometrics or VR/AR/MR and an unspecified category AI enablers. Therefore, AI capabilities and infrastructure elements are analyzed jointly. Additionally, criterias for good capabilities are defined and the subcategories of the selected frameworks analyzed to what extent they fulfill the requirements.

	CB Insights	Bloomberg Beta	StartHub Israel	Plattform Lernende Systeme
Separating core AI technologies and enabling technologies	no	no	no	no
Mutually exclusive categories to structure field of "AI"	yes, but very high level	no (e.g., NLP and autonomous systems part of data science)	yes	no (Sensoric overlaps with robotics and NLP overlaps with communications)
Categories on the same level (i.e. there are on categories that are part of others that are listed on the same level	no	no (e.g., Natural Language is part of Machine Learning)	no (hardware layer vs. type of AI application)	Data management more of basic requirement for other categories
Exhaustive (covers whole field of AI)	no	no	no	no

Table 8: Criterias for AI Capabilities.

Mapping of industries

The proposed set of industries is the most comprehensive one in comparison with the analyzed existing frameworks. It covers all of the industry-subcategories used by existing taxonomies. CB Insights' and "Plattform Lernende Systeme"'s selection of industries is largely similar to the standardized NACE code list and comparatively exhaustive. The other analyzed frameworks cover a significantly smaller selection of industries. Only three industries (Agriculture, Healthcare and Education) are covered fully and consistently by all five sample frameworks and additional two industry subcategories (Manufacturing and Finance) are covered by all five but not consistent with regards to abstraction level.

With regards to the subcategory Manufacturing, Bloomberg Beta covers only AI startups supporting the manufacturing of materials and StartHub Israel selects three individual areas (Industrial, Automotive and Drones). Bloomberg Beta differs also in their representation of the finance sector as they split "Investment" and "Retail Finance" into two separate industry subcategories. Additionally, Plattform Lernende System extends the finance related industry-subcategory by including Real Estate.

Mapping of enterprise functions

The here proposed structure for the enterprise function class comprises significantly more subclasses than the respective

appliedAI Framework	CB Insights	Bloomberg Beta	StartHub Israel	Plattform Lernende Systeme
Human Resources		HR	HR	HR
Marketing		Marketing	Marketing	Marketing, Sales
Customer Service & Support		Customer Support	Customer & Support	Customer & Support
Sales	Sales & CRM	Sales	Sales	
Accounting & Finance				Finance & Accounting, Legal
SCM & Distribution				Supply Chain
R&D				R&D
Production				Production
Operations				Planning
IT & Security	Cybersecurity	Security	Development & IT;	
Security	Infrastructure			
Purchasing/Procurement				Purchasing/ Procurement
Legal & Compliance			Legal	
Enterprise Intelligence	BI & Ops Intel	Enterprise Intelligence	Internal Data & Intel	
Cross-functional				

Table 9: Comparison of enterprise functions.

appliedAI Framework	CB Insights	Bloomberg Beta	StartHub Israel	Plattform Lernende Systeme
Agriculture, forestry and fishing	Food & Agriculture	Agriculture	Agritech	Agriculture
Mining and quarrying	Mining			
Manufacturing	Manufacturing	Materials	Industrial;	
Automotive; Drones	Manufacturing			
Electricity, gas, steam and air conditioning supply	Energy		Cleantech	Energy & Environment
Water supply; sewerage, waste management and remediation activities				
Construction	Construction			Construction
Wholesale and retail trade; repair of motor vehicles and motorcycles	Retail & Warehousing		Retail	Retail
Transportation and storage	Transportation	Logistics		Mobility & Logistics
Accommodation and food service activities			Travel	
Information and communication	Telecom; Media & Entertainment		New Media	ICT
Financial and insurance activities	Finance & Insurance	Investment;		
Retail Finance	Fintech	Finance, Insurance & Real Estate		
Real estate activities	Real Estate			
Professional, scientific and technical activities	Legal	Legal		
Administrative and support service activities				Other services
Public administration and defence; compulsory social security	Government & City Planning		Home, City & Misc	Government & Security
Education	Education	Education	Education	Education
Human health and social work activities	Healthcare	Healthcare	Healthcare	Healthcare
Arts, entertainment and recreation			Sports & Fitness	
Other service activities			Parenting	
Activities of households as employers;				
Activities of extraterritorial organisations and bodies				

Table 10: Comparison of industry categorization.

dimension of the analyzed existing landscapes. This difference is mainly due to the lack of representation of production related functions. In contrast, horizontal support functions such as HR, Marketing or IT are included in nearly all existing frameworks. One exception are the fields Finance & Accounting and Legal & Compliance. Plattform Lernende Systeme is the only one incorporating these fields in their framework as well. In general the in this report proposed list of enterprise functions is most similar to the dimension of Plattform Lernende Systeme.

5. STRATEGIC RECOMMENDATIONS WITH REGARDS TO EU DOCUMENTS

While the proposed structure serves the purposes of mapping stakeholders as well as use cases and fits to existing frameworks and landscapes, it also needs to be integrable in the EU documents - first and foremost the EU White Paper on AI and the recommendations of the High Level Expert Group on AI (HLEG).

5.1. INTENDED VALUE OF THE TAXONOMY

The taxonomy allows to map the AI ecosystem in Europe. AI is a basic technology that can be applied in every industry and a multitude of use cases. It will affect all our lives to a similar extent as electricity or the internet did. As a potentially disruptive technology, AI will become a key enabler for scenarios that have been hard to reach without using AI. Therefore, the single view at AI is misleading when it comes to regulation and innovation. AI needs to be considered in the context of use cases and scenarios. It affects ecosystems as well as our society as a whole. AI specific measures always require the context.

"The business plans of the next 10,000 startups are easy to forecast: Take X and add AI. This is a big deal, and now it's here."

Kevin Kelly, Founding Executive Director, Wired Magazine, 2014

Following the goal of the EC that the EU wants to become leaders in the application of AI, we emphasize that AI as a technology is only the enabler for a variety of applications. Ultimately, we want to lead in specific applications and scenarios with AI being a central technology. Consequently, the taxonomy is constructed to reflect use cases and scenarios as starting points for any analysis with regards to AI. We consider strong visions like the European Green Deal (EGD) with all the mentioned fields or the UN SDGs as the north star for a purposeful application of AI. If we consider AI as a disruptive technology that helps us reach our visions, we can create scenarios and identify the most relevant use cases within each scenario. Connecting the use cases with technology, regulatory or skills requirements helps us to monitor and track our ability to reach the vision.

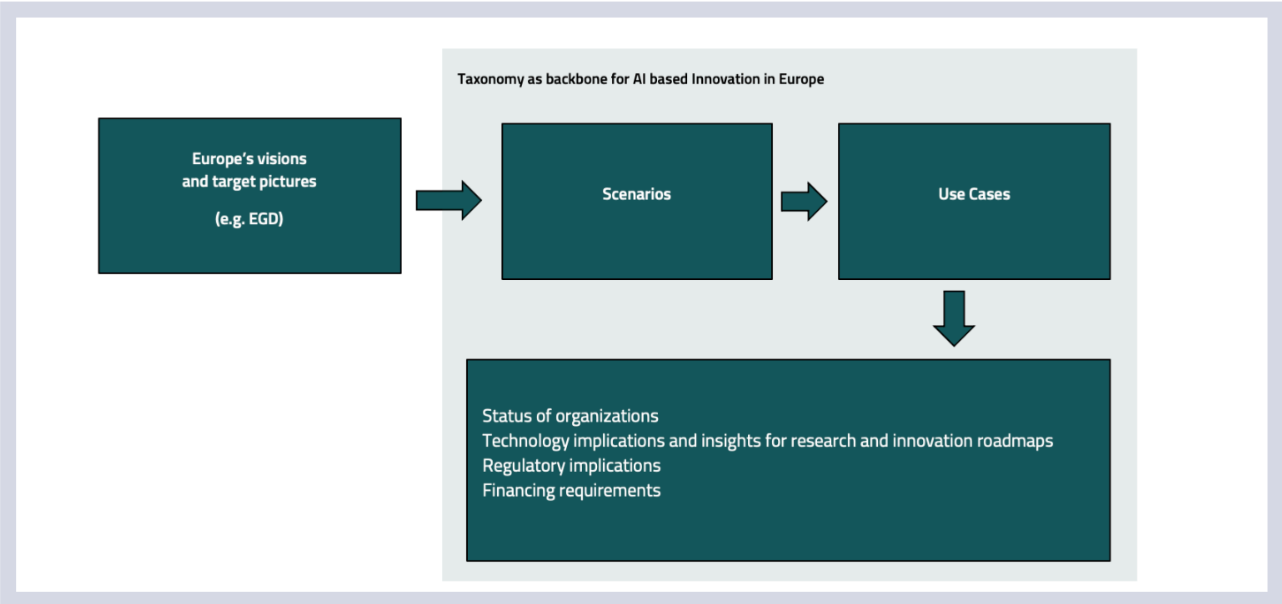


Figure 9: Intended value of the taxonomy,

Thus, used correctly, the taxonomy can serve as a backbone for a vision driven, accelerated approach to reaching Europe's goals to become a leader in AI. It serves several purposes

Common language: While there is no common definition of AI and its technologies, this taxonomy starts at the capability level. For example, independent of the exact algorithmic approach being used, Computer Vision is and will be a field of application. Starting at this abstraction layer, allows for a unifying framework that can be applied and used across Europe. It enables us to create consistent reports, studies, recommendations, or actions across several domains and units within Europe.

Dynamic: AI develops at an incredible speed. Therefore, the taxonomy needs to reflect constant changes while maintaining a stable structure. This is solved by separating the more stable structural parts from the more dynamic application specific components.

Acceleration: If Europe aims to become a global leader in AI, we need to accelerate our activities and focus on impactful measures. The scenario-based approach with the information structure of the taxonomy reflects this goal and allows for constant tracking and targeted activities.

Networked: AI application happens in ecosystems with new horizontal and vertical business models being developed. The interconnectedness of single actions (e.g. single regulatory proposals like the storage of training data) with a multitude of use cases and scenarios is reflected in the information architecture of the data base model.

All aspects are needed for the EGD to become successful. The Commission rightfully points out that the EGD requires a circular economy and ecosystem perspective while many effects are interconnected. As of November 2020, the EC focuses on nine policy areas[1] that could be used as a basis for scenarios in this structure. By analyzing the published action plans[2] regulatory, financial, educational, and technological measures can be delineated and integrated in the taxonomy. Ultimately, the most relevant AI based use cases can be identified to help the goals of the scenarios.

5.2 COMPATIBILITY WITH EU WHITE PAPER ON AI

The EU Whitepaper on AI is the central document of the EC and describes an Ecosystem of Excellence and an ecosystem of trust. Other documents are based on central concepts outlined in the whitepaper. Consequently, the taxonomy is designed to be fully integratable with the EU Whitepaper.

5.2.1. ECOSYSTEM OF EXCELLENCE

"The Commission is committed to enabling scientific breakthrough, to preserving the EU's technological leadership and to ensuring that new technologies are at the service of all Europeans—improving their lives while respecting their rights" (p.1, EC Whitepaper on AI).

While the EC outlines very broadly the ambition of Europe and working with member states, the goals need to be described in scenarios as they are reflected in the taxonomy. This also supports the mentioned focus of the efforts of the research and innovation community as well as the access to data and computing infrastructures in the whitepaper. By using the proposed taxonomy, one can identify the use cases and the underlying technological domains that are most impactful for reaching the scenarios. Lighthouse centers, moonshot activities and research clusters can be added in the proposed structure.

Another strong emphasis of the ecosystem of excellence is on skills which is directly reflected as the content element "Education" in the taxonomy. The focus on SMEs, the partnerships with the private sector as well as the promotion of the adoption of AI by the public sector is reflected throughout the structure in financing, regulation, organization as well as the education categories while being reflected in the use cases. Detailed analyses on the effectiveness of proposed measures can be made in the future using the proposed taxonomy.

The Whitepaper as well as the taxonomy is focused on Europe. However, we emphasize that in a globalized world players and activities outside of Europe need to be considered when developing strategic recommendations or action plans. Thus, the taxonomy might be extended to also monitor and track international activities outside of Europe.

5.2.2. ECOSYSTEM OF TRUST

Trustworthiness is a property of an AI system that is integrated in the taxonomy. AI capabilities and technology types each define use case properties containing individual aspects of trustworthy AI. For example, a use case using the AI capability of planning by using production data defines a different set of aspects regarding trustworthy AI than a use case of human surveillance using computer vision. Regulatory measures are then applied to use cases e.g. through the risk-based approach outlined in the Commission's White Paper. Having separated use cases from AI capabilities and AI infrastructure emphasizes the differentiation of technological aspects and technology specific regulatory measures from use case specific regulation. More importantly, regulation could also be scenario specific, applying the Ecosystem of Trust to a higher level.

The risk-based approach in its current proposal distinguishes two risk categories: Sector and use-case. Both aspects are reflected in the taxonomy which allows for the reflection of future regulatory measures if they follow the current line of thinking within the commission. The outlined requirements for potential high-risk cases regarding e.g. storing data, record-keeping, robustness, or human oversight would be integrated as part of a specific regulation.

While the ecosystem of trust highlights the need for regulation, it also emphasizes the importance of standardization, certification and voluntary labelling. Accordingly, the taxonomy includes associations and standardization institutions in addition to regulation as categories in its structure.

5.3. COMPATIBILITY WITH HLEG DOCUMENTS

5.3.1. Ethics Guidelines to Trustworthy AI

In its first report, the HLEG outlined seven dimensions for trustworthy AI. The Whitepaper builds on these dimensions in its ecosystem of trust. As outlined in section 5.2, the taxonomy is able to reflect the ecosystem of trust. Therefore, the Ethics guidelines are equally embedded in the taxonomy.

Dimensions:

1. Human agency and oversight (Including fundamental rights, human agency and human oversight).

2. Technical robustness and safety (Including resilience to attack and security, fall back plan and general safety, accuracy, reliability and reproducibility).

3. Privacy and data governance (Including respect for privacy, quality and integrity of data, and access to data).

4. Transparency (Including traceability, explainability and communication).

5. Diversity, non-discrimination and fairness (Including the avoidance of unfair bias, accessibility and universal design, and stakeholder participation).

6. Societal and environmental wellbeing (Including sustainability and environmental friendliness, social impact, society and democracy).

7. Accountability (Including auditability, minimisation and reporting of negative impact, trade-offs and redress).

These dimensions are all properties of use cases and the applied AI capabilities. Therefore, they are reflected within the use case and capability mappings and descriptions.

5.3.2. Policy and Investment Recommendations for Trustworthy AI

In its policy recommendations, the HLEG translates the structure that it described in its first publication to several application areas.

In the first chapter, the HLEG describes how trustworthy AI can be used to create positive impact in Europe. The HLEG emphasizes four different stakeholder groups (society, private sector, public sector and research). While public sector and society are described as indirect stakeholders (users, customers) and therefore also reflected indirectly in the taxonomy, the private sector and research institutions are active stakeholders in a European AI landscape as those who develop use cases or provide the necessary technologies. In line with the HLEG's structure, the taxonomy embeds private sector (through startups and companies) and research Institutions.

In the second chapter of the Policy and Investment Recommendations, the HLEG describes enablers for a European AI ecosystem. Consequently, all aspects highlighted by the HLEG are integrated into the taxonomy. The HLEG outlines the supporting technology infrastructure (data and infrastructure), skills and education, governance and regulatory framework, and funding and investment. All four components are part of the proposed structure and connected to scenarios, use cases, and organizations reflect the embedding in the European AI landscape.

5.4. COMPATIBILITY WITH EIT DIGITAL - A EUROPEAN APPROACH TO AI (POLICY)

The EIT Document analyzes the European policy framework and elaborates a variety of potential levers for establishing trustworthy AI in Europe while continuing to the Ecosystem of Excellence. It rightly emphasizes the need for context-specific analyses of regulation with regards to AI. In particular five concern/opportunities are displayed with Data, Organisation, Human capital, Trust and Markets and applied to the sectors Manufacturing, Mobility, Climate and Health. All aspects would be reflected in the regulation category of the taxonomy with connections to scenarios and use cases. Following Example 3 in Section 3.4.2, the taxonomy could be used for a deeper assessment of regulation and policy recommendations. By systematically analyzing use cases that are tied to a scenario and its underlying technologies and organization, the impact of various existing or potential regulations on the selected scenario can be assessed.

5.5. FURTHER DOCUMENTS

The HLEG documents and the EC's white paper are considered as the most relevant documents. Additionally, Members of the European Parliament have written three additional papers that are considered in the taxonomy.

- Civil liability regime[3]
- Framework of ethical aspects [4]
- IP Rights[5]

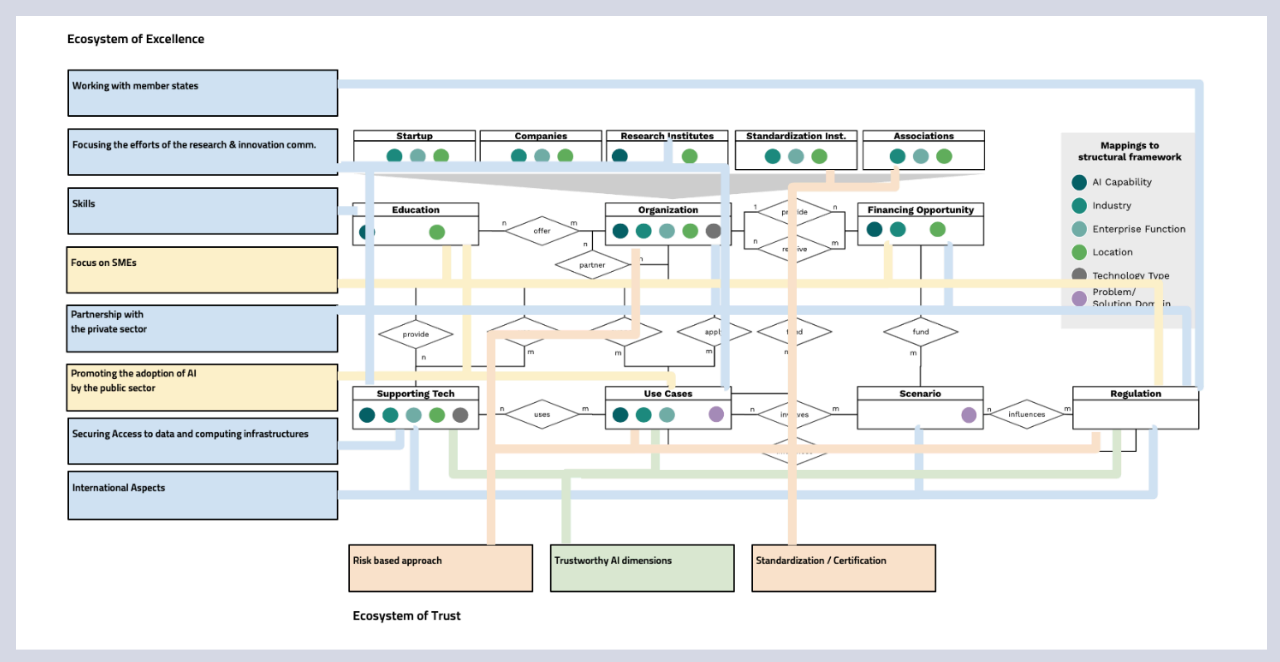


Figure 10: Mapping of the EU white paper on AI to the outlined structure.

The civil liability regime will need to and can be implemented in the regulation category when it is translated into binding rules. The ethical aspects and the IP rights impact the taxonomy only on a meta-level through use case design and scenarios.

5.6. OVERALL MAPPING OF STRUCTURE WITH THE COMPONENTS OF EU DOCUMENTS

Overall, the taxonomy is suited to reflect central European documents like the White Paper on AI of the European Commission including the ecosystem of excellence and the ecosystem of trust. Figure 10 displays the connection of the categories in the whitepaper's ecosystems with the categories in the taxonomy. The colors vary for readability purposes only.

Consequently, the proposed taxonomy in its current structure is suited for becoming the information architecture and model for the existing AI landscape in Europe. In addition, due to the dynamic nature and the flexibility in the application, the taxonomy is adaptable to future changes in the AI landscape like new structural elements, new unforeseen characteristics of AI, or additional stakeholder groups

Section Footnotes:

[1]Biodiversity, From Farm to Fork, Sustainable Agriculture, Clean Energy, Sustainable industry, Building and Renovating, Sustainable Mobility, Eliminating Pollution, Climate Action

[2]E.g. energy system integration, biodiversity, circular economy, European Industrial Strategy

[3]https://www.europarl.europa.eu/doceo/document/TA-9-2020-0276_EN.html

[4]https://www.europarl.europa.eu/doceo/document/TA-9-2020-0275_EN.html

[5]https://www.europarl.europa.eu/doceo/document/TA-9-2020-0277_EN.html

Funded by the
European Union

Visit: ai.eitcommunity.eu