AI industry – is there such a thing and how does it look?

Henten, Anders; Tadayoni, Reza

Publication date:
2024

Document Version
Accepted author manuscript, peer reviewed version

Link to publication from Aalborg University

Citation for published version (APA):
Henten, A., & Tadayoni, R. (2024). AI industry – is there such a thing and how does it look?. 1-17. Paper presented at 24th Biennial ITS conference , Seoul, Korea, Republic of. https://www.econstor.eu/handle/10419/302459/

Anders Henten and Reza Tadayoni

CMI, Electronic Systems, AAU, Denmark

AI industry – is there such a thing and how does it look?

24th Biennial ITS conference

23-26 June 2024

Seoul, Korea

Anders Henten, Electronic Systems, AAU, Denmark

Reza Tadayoni, Electronic Systems, AAU, Denmark

AI industry – is there such a thing and how does it look?

1. Introduction

The aim of this paper is to present a descriptive analysis of Artificial Intelligence (AI) from an industry supply side point of view. Almost all non-technical research in AI is concerned with the user side, i.e. what AI is used for and where, and what the likely implications are, etc. Indeed, there is no doubt that this is where most emphasis should be. However, the supply side is grossly under-researched, with few exceptions (Simon, 2019; Jacobides et al., 2021), and the purpose of this paper is to shed light on what the AI industry looks like – presupposing that such an industry exists in the sense that it can, to some extent at least, be set apart from IT industries in general. The research question of the paper is, therefore, concerned with the extent to which an AI industry can be set apart from the IT industry as such and what the structure of the industry looks like.

An analogy to the telecoms industry could be made. Most people would understand the telecoms industry as consisting basically of companies producing telecoms hardware and software and companies delivering telecoms services. And then, on top of that, there are all the companies developing and delivering applications and content, etc., that are not commonly included in the conceptions of the telecoms industry. But can such an analogy be used for understanding AI from an industry structure point of view?

The first step is to define what AI is, namely, to specify the boundaries of the AI concept and, therefore, also what it is an AI industry is producing. This is not an easy task, as there are variations in the definitions and conceptualizations of AI (Tobin, 2023). Thereafter, it is important not only to define the boundaries but also to explain what's inside these boundaries, i.e. which are the different kinds of technologies that can be characterized as part of the AI concept. In order to do so, it is helpful to include a historical perspective, as AI and the AI concept has developed over time, and new technologies have become part of the overall AI concept.

Following this, the next step is to pinpoint the companies and organizations that can be seen as producing goods and services within the AI area. This comprises hardware, to the extent that the hardware is specifically developed for AI use. It also includes AI software, data specifically assembled and used for training of AI systems, and AI-based services and applications. From this description of the different parts of an AI industry, it can already be seen that the boundaries of an AI industry are blurry to a large extent, the reason being that it can be difficult to differentiate between hardware used for AI and hardware used for general IT purposes just as it can be difficult to differentiate AI software from software in general. Also, it can be difficult to differentiate between data assembled specifically for AI training purposes and data in general. In spite of these difficulties in clearly defining the boundaries of the AI concept and an AI industry, the following steps are to estimate the size of the AI industry and to find out where AI industries are located.

In terms of policy initiatives around the globe, there is no doubt that in the minds of policy makers there is an AI industry, and that policies are being developed to establish and/or further develop the industry (Simon, forthcoming 2024). An important question is the extent to which an AI industry is becoming a separate industry or the extent to which it is an indifferentiable part of the broader IT area.

The paper is a descriptive analysis and does not use a theory framework for its analysis. The methodological approach used is first to delimitate the AI concept and its constituting elements, followed by an overview of the size and industrial structure of the area, the major companies within the AI field and their location. The paper is concluded by an illustration of the structure of the AI area and a discussion as to the extent to which the AI area is becoming a separate sector of the IT industry or whether it is an indifferentiable part of the IT industry as such. In this context, a revue of the deliberations and developments of the United Nations' International Standard Industrial Classification (ISIC) system is included (Advisory Expert Group on National Accounts, 2021).

2. Motivation

When introducing the idea of writing a paper presenting a descriptive analysis of AI hardware, software, dataset and services from a supply side point of view, the reaction has now and then been slightly skeptical or at least hesitant. Why is this interesting, and what can we learn from that?

Searching for academic literature, there is very little on the topic. However, on web pages, consultancies and industry observers publish overviews and trends. This would indicate that there is an interest 'out there' in knowing more about the development of AI from an industrial production point of view. What does 'the industry' look like? Who are the producers? What is the size of AI production? Etc.

This has been the approach from the 'statistical community'. Once something new comes up, statisticians will be interested in knowing what the contours of the new phenomenon are and how it can be measured. At the moment, the System of National Accounts (SNA, which used to be called the United Nations Systems of National Accounts) does not include a category entitled Artificial Intelligence. One reason is that a new update of the SNA is still under negotiation and that Artificial Intelligence was not considered a separate category in the latest update in 2008. However, it has also turned out to be an issue that is highly debated. The question is whether it is possible to separate AI from other kinds of IT hardware, software, datasets and services. This probably needs to be settled soon, as the new update of the SNA is due by 2025.

Questions relating to AI have been how AI can be defined and how it can be set apart from IT in general. In a presentation made by a staff member of IMF, Andrew Baer, in 2022, there is reference to a survey made among statistical experts from different parts of the world. The first question was: ‘Is this topic of relevance for your country?’ – ‘this topic’ meaning improving the visibility of AI in national accounts. Actually, only 33 out of 48 respondents considered this as being of medium or high relevance – and of these only 14 of high relevance. Among the experts from the EU, only 1 out of 10 respondents considered it as highly relevant, 4 indicated medium relevance, 4 low relevance, and 1 no relevance (Baer, 2022).

When asked: ‘Do you agree that Artificial Intelligence should be explicitly mentioned in the asset classification in a new class called “Computer Software and Artificial Intelligence” with AI appearing with an “of which” category?’, 38 out of 48 answered yes. In the comments to the answers, negative responses anticipated difficulties in differentiating AI from other software (Baer, 2022).

There were also questions regarding the definition of AI. The brief version of what has been agreed upon in the negations of a new SNA is: ‘AI is a computer program operating a system capable of recognition and reasoning consistent with human recognition and reasoning’. 40 out of 48 agreed with this definition though some, for instance, suggested ‘simulating’ instead of ‘consistent with’ (Baer, 2022).

Though this definition could be subject to change in the final version of a new SNA, it would seem that AI is considered as a sub-category of computer software and that hardware specifically designed for AI applications is not included. Also, there is a question as to whether training datasets will be considered part of AI. One question in the survey was: ‘Do you agree that the value of the cost of producing training datasets should be excluded from the value of own-account AI and included instead in the value of data assets?’. 38 respondents out of 48 agreed with this statement though there were comments anticipating difficulties in separating software and databases (Baer, 2022). It is interesting (confusing) to note that the Advisory Expert Group on National Accounts recommends that ‘the value of the cost of producing training datasets be included in the value of own-account AI’ (AEG, 2021). There is apparently confusion regarding defining the contours and delimitations of AI.

How to define and delimitate the concept of AI, will be discussed in a separate section. However, it seems clear that it is not an easy task to define and delimitate AI as a separate statistical category, let alone an ‘industry’. On the other hand, it *will* be done, and AI will appear as a separate category in the upcoming SNA.

Another line of motivation is the increasing amount of policies and regulations regarding AI. For instance, in March 2024, the EU parliament passed a much-debated AI Act (EU, 2024) Also, industrial policy regarding AI is on the rise all over the world, as countries and their governments seek to position their countries in the AI area (Simon, forthcoming 2024). Production of AI-related hardware, software, datasets and services is highly unequally distributed around the world, and countries seek to find as favorable as possible positions in this international division of labor.

However, little research attention has been on the supply-side regarding AI in social sciences. Primary attention has been on the usage side. This becomes a problem when wanting to be more specific about possible industrial policy interventions and when wanting to regulate AI. An example of this was the 'surprise' that Generative AI was to the policy and regulatory community when developing the EU AI Act (EU, 2024). Had the supply side of AI been followed more closely, Generative AI would not have come, to the same degree, as a surprise. The same applies to academic studies of the use of AI applications. Increasing the knowledge on the supply-side will make it easier to follow the diffusion and use of AI. Otherwise, the risk is that studies of AI implementation and use will stay relatively abstract, as the specific technical solutions are not sufficiently part of the analyses.

3. Literature overview

There is a massive dearth of academic literature in the field of social science including economics analyses of the supply side of AI. In social science academic research on AI, practically all attention has been on the use side – not even demand, as this would entail some quantification. Attention has been on qualitative analyses of prospective and actual implementations of AI in different sectors and professions with emphases on how AI can support work functions (augmented intelligence) and/or on the potential displacement effects of AI on human labor.

This, one can ascertain by searching with keywords such as 'AI supply', 'AI industry', 'AI sector', 'AI companies' in academic databases. Very few items appear, while social science literature on the use of AI and how AI will affect different work functions and sectors is increasing steeply. While this is a good thing for better understanding business and work implications of AI implementation, it provides no understanding of how the industries providing AI solutions look. Citing Agrawal et al. (2019), Jacobides et al. (2021) also make this point, namely that supply/provision side analyses of AI are rare.

In a book collection of paper edited by Agrawal et al. (2019), entitled 'The Economics of Artificial Intelligence', different economics-oriented perspectives on AI developments are given, including AI as a General Purpose Technology, Growth, Jobs and Inequality, Machine Learning and Regulation, and Machine Learning and Economics. Though the collection of papers throw light on AI from different economic perspectives, it does not in reality provide any insights into how the production of AI solutions are structured and what the sizes of AI markets are. A very preliminary start on this is present in a paper by Varian (2019), but this is only a very early indication.

Jacobides et al. (2021) themselves present an analysis that focuses on the development and innovation of AI ecosystems in an evolutionary perspective. This paper is one of the very few social science academic papers performing an analysis of AI developments from a supply side point of view. There are, however, many papers that are concerned with developments of AI from a technical perspective. But they do not provide any insight into the markets for AI solutions.

Another paper written from a social science perspective is one by Jean Paul Simon that was published in 2019. This paper does not have a theory framework or theory ambition as the Jacobides et al. paper (2021). Simon's contribution is an empirically based paper aiming at showcasing the industrial development of AI and providing information on the geo-economics of AI with pointers as to the developments in the USA, China and the EU.

Simon's paper (2019) is based on reports and insights from large international consultancy companies and from governmental reports. This is where information on the production side of AI can be found including information on the companies involved and where estimations of market sizes are made public. The strongest aspects of such contributions are related to the structuring of the production of AI solutions. In these contributions, one will often find a structuring of the production of AI solutions comprising hardware, software, and services. The estimations of market sizes are highly fluctuating, and the quality of the estimations is hard to assess – as can be seen in the section of the present paper on market estimations.

Government reports from different countries also try to assess the development of AI production in their countries. The depth of these kinds of publications also differs, but there are in some countries a tradition for producing high quality government reports on many different topics including industrial developments. This applies, for instance, to the UK, where a report by Perspective Economics for the Department of Science, Innovation and Technology (DSIT), entitled 'Artificial Intelligence Sector Study' was published in 2023.

An additional source of information on AI from a supply side point of view is the deliberations from statistical institutions. In the present paper, we have a section reporting on such deliberations, where the main topic is the extent to which one can delimitate AI production and how one can elaborate statistical accounts of AI developments.

The question in these statistical deliberations is whether or the extent to which it is relevant to determine what an AI sector is and how the border lines of an AI sector can be set. This is a very relevant question, however, the ambition of the present paper is to attempt to define what AI production is and how the different parts of the production of AI solutions are related to one another. For the telecoms industry this has been done for a long time. A theoretical contribution to understanding the different elements of the telecoms industry and how they are related was published in the journal Telecommunications Policy by Martin Fransman in 2002. The title of the paper is 'Mapping the evolving telecoms industry: the uses and shortcomings of the layer model'.

The ambition of the present paper is similar to the Fransman paper: Can a layered model of an AI industry be made? There are many things that speak against such a model of the AI area understood as an industry. However, there are also things that speak in favor. First of all, that it might provide a better insight into the relationships between the different areas of AI solutions production. Secondly, that industrial policies are currently being developed in different countries, based on preliminary understandings of the structures of the AI production area and on the political ambition to have part in the international distribution of AI production. And thirdly, that it will be 'done'. Statistical institutions and departments will eventually develop models for structuring the production of AI solutions in order to assess the sizes of production, consumption exports and imports, etc.

4. AI in historical perspective

The concept of artificial intelligence was first introduced in the mid-1950s. It was the then assistant professor at Dartmouth College John McCarthy, who for the first time used the concept of artificial intelligence. John McCarthy has been since recognized as the founding parent of AI and his interest in AI "was sparked by the Hixon Symposium on Cerebral Mechanisms and behavior he attended at Caltech in 1948" (Sabanovic et al., 2012). McCarthy organized a scholarly event in 1955 where he invited mathematicians from different universities to discuss the question if machines can think. In this event he coined the term 'artificial intelligence' (AI) to refer to machines and processes that imitate human cognition and make decisions like humans (Tlili et al., 2023). Small developments were done in this period but for a long period of time AI was not seen as a societally crucial research area and it was difficult to get funding for research in the area.

Anders Søgaard in his popular-scientific book 'Artificial intelligence from behind' proposes an AI Winter period from 1970s for a period about 40 years where we don't see significant developments in AI. He mentions lack of data, computing, and necessary toolbox as the main arguments for the low attention to AI during this period (Søgaard, 2022). These limitations have been considerably minimized during recent years where access to data, computing power, storage, and advanced infrastructures have enabled development of very advanced AI algorithms that are used in several use cases such as health technologies, automotives, media technologies, surveillance etc. Furthermore, AI is becoming an important element in many optimization algorithms used in different sectors of the society.

Since 2010, the development of AI has been intensified and provided as specific applications through, for instance, Appel, Google, Microsoft and other larger technology provides. As examples we can mention voice assistance applications like Apple's Siri, use of Natural Language Processing in translation applications like Google Translate, and recently the launch of Generative AI applications such as ChatGPT by OpenAI.

Today there are different ways that AI is categorized in the literature. IBM divides the different types of AI based on capabilities in 1) Narrow, 2) General and 3) Super AI. The Narrow or weak AI can be trained to perform a single or narrow task, the General or Strong AI presupposes no human involvement in training or tagging the underlying data and the Super AI is a futuristic theoretical AI with cognitive abilities beyond the human mind. Machine learning, Computer Vison and robotics are important components of AI developments.

5. Relation between different layers in AI

Like the OSI 7-layer reference model (Day et al., 1983), a layered approach called the AI stack is proposed in an article named 'The AI stack: a blueprint for developing and deploying artificial intelligence' by (Moore et al., 2022), depicted in figure 1. The AI Stack 'establishes a model of an AI system wherein each layer uses results from the layers beneath it and passes its results up to the layers above it to build for other things needed to achieve AI' (Moore et al., 2022).

In the paper the different layers are defined as:

1. Computing Layer: Systems, networks, programming languages, operating systems and interactions between devices that make computing possible.
2. Device Layer: The device layer is all the sensors and components needed for machines to perceive the world around them.
3. Massive Data Management Layer: The data, which is used by ML, including tagging and training and management of data.
4. Machine Learning Layer: This layer belongs to ML algorithms.
5. Modeling Layer: The paper defines this layer as models used to construct and manipulate abstract representations of situations and natural phenomena in the world.
6. Decision Support Layer: This layer includes technologies that help humans make decisions.
7. Planning and Acting Layer: According to the paper, systems in this part of the stack rely on optimization, safety, the knowledge network, and strategic reasoning to make the best possible decision available and learn from the information researchers give them.

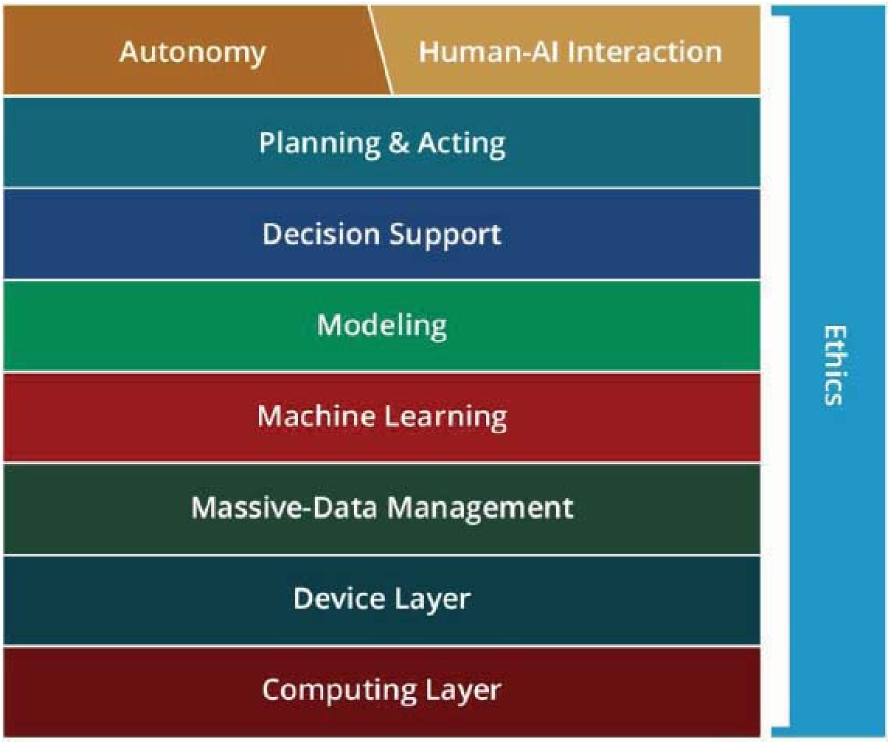


Figure 1 – AI stack (Andrew W. Moore et al., 2022)

8. Human-AI Interaction Layer: This layer is concerned with the situation where AI interfaces to human actors, as stated in the paper 'when we create artificial intelligence in this part of the stack, we're augmenting what humans can do'.
9. Autonomy Layer: AI technologies at this level concern with 'creating systems that make their own decisions without human intervention'.

As seen in the figure, ethics has been illustrated as a vertical layer that permeates the entire AI stack. The decisions people make as they build AI systems involve serious ethical questions that we can't ignore (Moore et al., 2022)

In the following we focus on a modified version of the AI stack model where we combine the layers such that the new reference model has closer relation to the topic of our analysis that studies to what extent we can identify a specific AI industry. The new layered model is depicted in figure 2. We remove the ethics layer, not because it is not important but because it is not directly relevant for the topic of our paper.

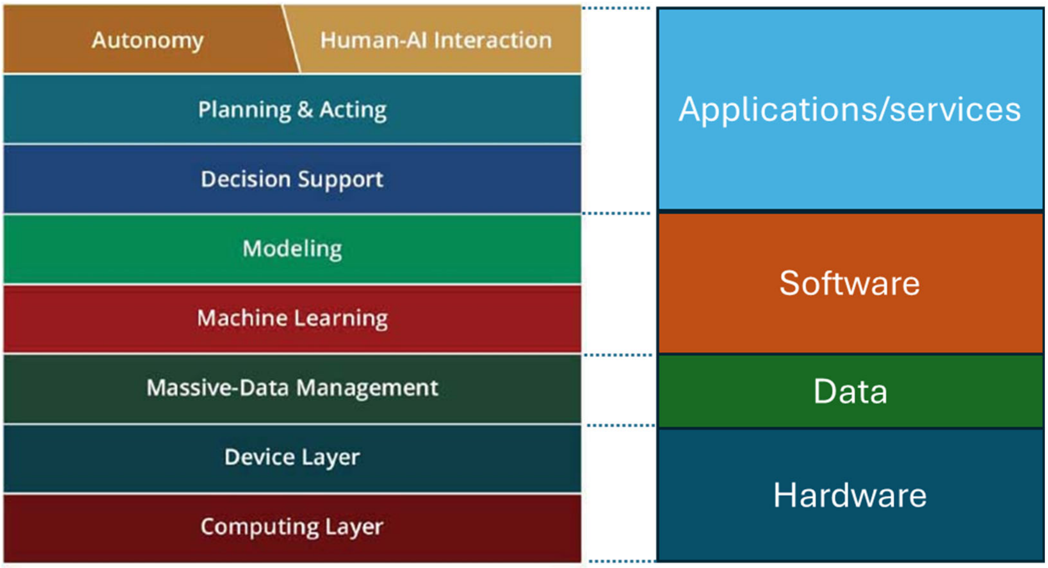


Figure 2 – AI industry Layered model

The new hardware layer is composed of the computing and device layer.

In the computing layer we can see tremendous developments in capacity of processors, storage, network interfaces, GPUs and the network infrastructure enabling optimized architectures using high bandwidth wired and wireless network technologies and Cloud and Edge computing and high-capacity data centers. As mentioned before, computing was a big challenge for developing AI in the 'AI winter' that with the recent developments can be considered as one of the major enablers for the development we are witnessing now.

The device layer has also taken a jump forward in the recent years by developments of sensors and miniaturization of embedded computing resulting in development of Internet of Things (IoT). The development of IoTs is also due to development of dedicated network infrastructures connecting and managing the massive number of devices that have specific requirements to scalability and to power consumption and general maintenance. In this regard, the development of 4G and 5G and specific IoT infrastructures like LoRa and Sigfox have been important drivers for these processes. The development in IoTs has enabled the possibility for gathering massive amounts of data concerning users and use cases and the context of use that have been essential for development of AI.

The data layer concerns the big data gathering and handling, including tagging and training of the data. The data is gathered amongst others through the IoT devices and stored in the Cloud and Edge devices.

The software layer concerns Machine learning (ML) and modeling, including a number of developments within Neural Networks (NN), Natural Language Processing (NLP), Large Language Models (LLM), speech recognition and Computer Vision, including pattern recognition. With regards to ML, there are different models being supervised or unsupervised. In supervised learning labeled /tagged data are used, while in unsupervised or self-supervised, there is no labeling of the data, and the algorithm looks at underlying structures and patterns in the data. Re-enforcement learning and deep learning are other algorithms that do not rely on tagged data.

The application and services layer concern the use-cases including decision support systems, planning and acting systems, as well as the interface to human and nonhuman actors. With regards to the human actor, it refers to applications and services AI provides for optimization and accuracy of the work being done, and the autonomy layer concerns autonomous systems without user interaction like the AI used in autonomous cars or other appliances.

6. AI markets

As official statistics have not (yet) been developed for the AI area – see discussion on UN work to establish a foundation for AI statistics – it is extremely difficult to assess the sizes of AI markets. This has obviously not stopped consultancy companies and other market observers from estimating a wide range of metrics for market sizes, investments, implications for job replacement, etc. As expected, the figures reported by consultancies differ to a very large extent. Also, the data presented are mostly intended towards making forecasts more than for reporting on the past and the present – part of the reason being that the forecasts are aimed at companies trying to make sense of future developments of AI to determine whether and when to invest and implement AI solutions. Nevertheless, the forecasts presented take their point of departure in estimations of current market sizes, and these are the figures that we will report on in this section.

There are websites that compile different sources for AI market developments, for instance www.semrush.com with references to Statista, Crunchbase, PwC, McKinsey, World Economic Forum, Gartner, etc. This is useful in the sense that it provides a quick overview of different sources for estimations of AI developments. However, it also shows the wide disparity in estimations and the fact that it is only seldomly clearly presented what the background sources are and how the figures have been estimated.

But if one takes just a few of the estimations provided, Statista has published figures stating the global AI market size was 208 billion USD in 2023. Fortune Business Insight estimates that the global AI market size in 2023 was valued at 515 billion USD. Grand View Research estimates the world market to be 197 billion USD in 2023, and the assessment by marketandmarkets is that the global market was at 145 billion USD in 2023. What the right size is can be difficult to assess as it is seldomly stated what the 'AI market' comprises and what the bases for the estimations are. However, there is agreement on vast growth. In 2030, the estimations are in the trillions – 2, 3, 4 or more. This is not said to ridicule such estimations, as they provide some kind of 'ballpark' assessments. But they do call for caution, as the bases for the estimations remain unclear.

Just as the estimations regarding total AI market sizes differ very much, it is even more difficult to assess the market sizes for the different segments of the AI market, i.e. hardware, software, datasets, and services. However, in order to indicate some orders of magnitude, we will report on estimations of the global markets for hardware, software, datasets and Generative AI. Generative AI has since the release of ChatGPT in late 2022 become the 'beast in the revelation' and is already subject to market size estimations.

The global AI hardware market is by SkyQuest estimated to be app. 17 billion USD in 2023. Precedence Research has an estimation saying app. 53.7 billion USD, and Intent Market Research reaches around the same figure: 54.3 billion USD. The global AI software market is estimated to be app. 150 billion USD in 2023 by Verified Market Research, and Precedence Research says app. 170 billion USD. Hardware and software are assessed to be the largest parts of the overall AI market (not taking services into account), while datasets constitute a much smaller segment. According to Fortune Business Insight, the global 2023 AI training dataset market size was valued at app. 2.4 billion USD. Peculiarly enough, the figures provided by the different market forecasts are mostly very exact with decimals after the comma – which does not add to the credibility of the forecasts.

Market size estimations for Generative AI have already begun to come forward. As with the other AI market segments, the estimations of the market for Generative AI also fluctuate. Zion Market Research, e.g., estimates that the market in 2023 was 13.2 billion USD, while the estimation by Statista is at almost 45 billion USD.

It goes without saying that the credibility of the above estimations is relatively low. However, if staying with the intention of providing a 'ballpark' assessment, it would (not unsurprisingly) seem that software is the biggest area, constituting app. 2/3 or 3/4 of the overall global AI market (excluding services). Hardware is the second largest segment with 1/4 to 1/3 of the market, while training data only represents a small segment of the overall market. How the market for Generative AI is to be exactly understood in this context is not clear. However, an assumption will be that Generative AI is part of the software segment and possibly the training data segment. In other accounts, e.g. by Precedence Research, the market segments are hardware, software and services - with services covering implementation and configuration. In this account, services represent the largest share with app. 40%, while software is app. 37% and hardware is app. 23%. This market segmentation seems reasonable, as the actual deployment and implementation of IT solutions in general often constitute a large portion of the total costs.

On a global scale, supply will (eventually) equal demand, however, at a national level, supply can either be smaller or bigger than demand. There will be countries where demand is higher than and different from what is nationally produced, and there will be countries where exports are larger than imports. In terms of demand, estimations are that the three largest markets are the US, China and Europe, where some estimations are that the US market is twice as large as both the Chinese and the European markets respectively. In other estimations, for instance, Precedence Research, the North American market (including Canada) constitutes a little more than one third of the global market, while Asia Pacific and Europe respectively represent one fourth. In terms of supply, the countries with the largest productions are the US and China with Europe trailing behind, where the UK is the most prominent country in Europe.

In an overview of the largest companies producing AI solutions, Exploding Topics list 22 companies of which 17 are based in the US. One is from China (Baidu), two are from Israel, one from Norway, and one from the UK. This may very well be an exaggeration of the US dominance – in other accounts, there are more companies from China. However, it does provide a picture of the stark US dominance in AI production, and it also indicates the position of Israel in the global AI market.

A prominent manner of illustrating the structure of the AI industry and its companies is by making a layered illustration of an 'AI ecosystem'. In such layered ecosystem illustrations, there is generally a hardware layer, a software layer, and a services layer. Marketsandmarkets has made such layered ecosystem illustration, comprising hardware (hardware and devices and network), software (including security), platforms, and then services (including cloud). In each of these layers, the companies that are included are, to very large extent, companies based in the US. The most prominent companies are the big tech companies from the US. These companies dominate the markets within each of the layers. These are, furthermore, the same companies that dominate the whole IT field, illustrating the fact that AI is becoming a more or less inseparable part of IT in general.

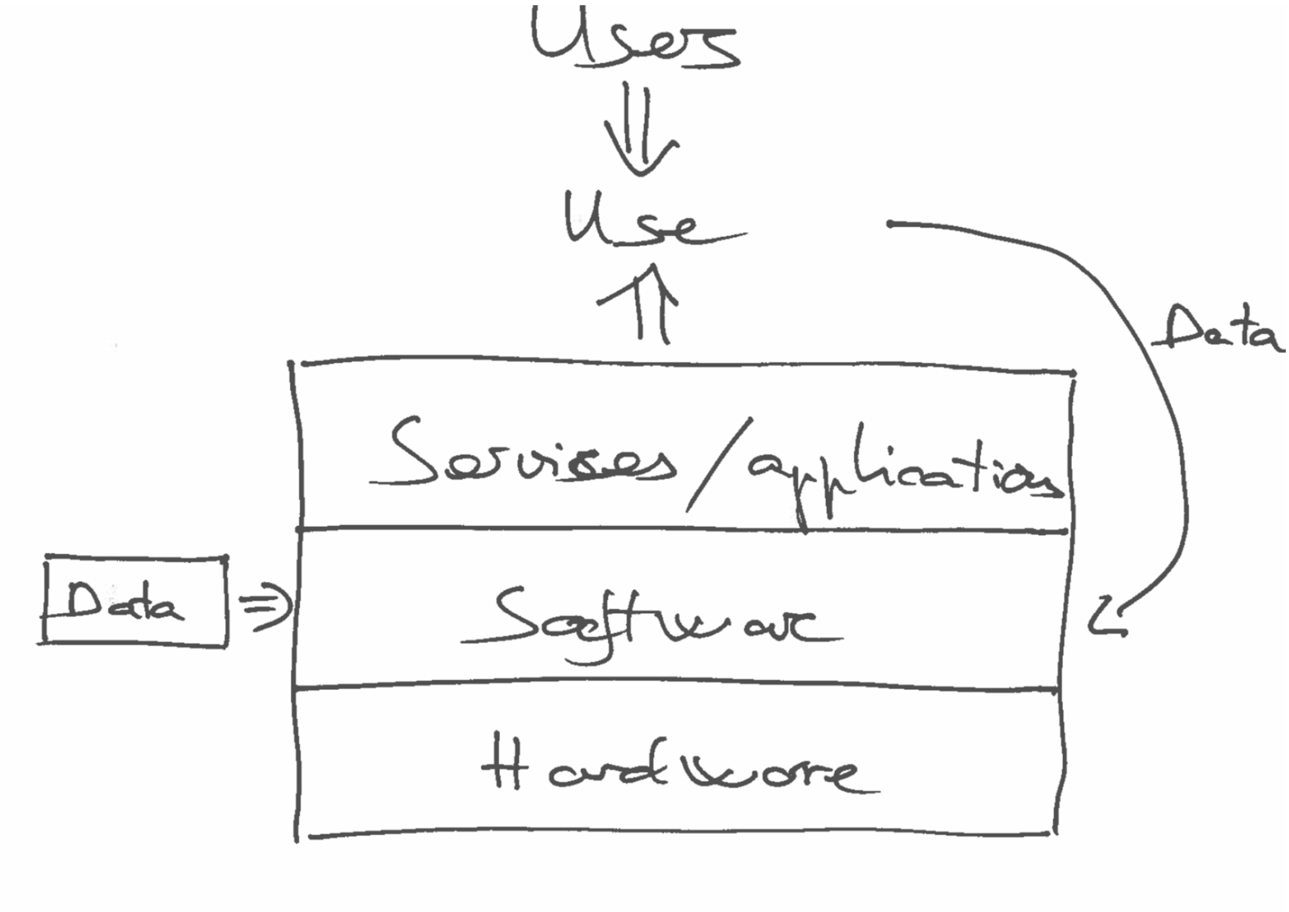
7. Analysis

In the motivation section of the paper, focus is on the 'soul searching' discussion as to whether one can define what an AI industry is. The approach is whether AI can be separated from other IT hardware, software, data and services and be defined as a separate industry area. Among statisticians this is not settled, though the tendency is that an AI area will be defined and recommended to be included in national statistics. Among consultancies and other observers, this discussion is long past, and statistical data is being produced and published in large varieties.

The aim of this paper is to visit the discussion regarding the definition of an AI industry and to discuss the structure of such an industry area. In spite of all difficulties in defining an AI industry, the direction seems to be that defining the boundaries of an AI industry area will be done, because of the great importance of AI technology and solutions. With respect to the structure of the AI area, the type of model that can be used to better understand the structure of the area is suggested to be a layered model – as was discussed by Fransman (2002) regarding the telecoms industry. The purpose of a layered model as proposed by Fransman is to define the different segments of the industry and to facilitate analyses of the relationships between these different segments including the developments of innovation. In fact, layered models have for long been used in technical as well as techno-economic analyses. The OSI-model (Open Systems Interconnection) was developed in the 1970s by the International Organization for Standardization (ISO). Later a similar model was developed for the Internet, the TCP/IP-model. Such layered models are also used by consultancy companies when illustrating what is often termed AI ecosystems.

A model for the AI industry area would take its point of departure in models similar to the telecoms industry with a hardware layer at the bottom, a software layer in the middle, and a services layer at the top. In addition, there would be a training dataset segment. The core of the hardware layer would be processing power necessary for supporting AI solutions, including processors specifically developed for AI solutions. At the software layer, there would the algorithms that are specifically designed for AI solutions, and there would be the training of the algorithms and the datasets that are used for training the algorithms. At the services layer, there would be implementation and configuration of AI solutions.

Figure 3: Layered model of AI supply



The figure illustrates the three main layers of any IT but also AI solutions. The special thing about AI is the vast importance of training data input and input of data from the use of the AI services and applications. This is shown with the training data input to the left and the data input from use to the right.

Conclusion

In recent years, there has been a steeply increasing interest in AI, which simply exploded with the launch og ChatGPT in late 2022. For many years, AI was not a topic that received much public attention and not either much interest from social science research. AI was developed 'behind the screens' in technical research and development in IT companies primarily and also in research in IT departments at universities. With the ‘break through’ of AI during the past five years, social science research has primarily focused on the usage side and very little attention has been given to the supply side.

The research question of the paper is concerned with the extent to which the AI area can be set apart from the IT industry as such and what the structure of the industry area looks like. The first part of the research question is in the paper dealt with by way of discussing the difficulties of separating AI hardware, software, data and services from IT hardware, software, data and services in general. In the UN context, where principles for defining the categories in the System for National Accounts are developed, there has been an ongoing deliberation and discussion of this, which still is not concluded. However, the likelihood is that decisions will be made no later than in 2025, and that these decisions will include an AI category. However, it seems that only AI software will be included. Among consultancies and other industry observers, statistics for the developments of the different segments of an AI industry have long been produced, including hardware, software, data and services. The problem with these statistics is that they differ to a large extent and that the bases for the estimations are seldomly clear.

One could argue that trying to define AI is a ‘fool’s errand’ – just as Lehr et al. (2018) wrote about defining Internet. It is much too complex and inseparable from IT in general. However, industrial policies regarding AI are currently being developed around the world, and so are regulations. If countries wish to support and promote AI production in their countries, there will be a requirement for understanding what an AI industry is. With respect to regulation, it could be argued that it is the use of AI that needs regulation, not the production. However, intervention in AI products may also be necessary and, therefore, intervention in AI production.

The structure of the AI industry area can be illustrated by a traditional layered model with hardware at the bottom, software in the middle, and services at the top. In addition to this traditional model, data for the initial and continuing training has to be added, as the core of understanding AI is based on algorithms being pre-trained and continuously trained by use data in order to deliver services that take new situations into account.

Information primarily provided by consultancy overviews of AI production clearly shows that, though there are increasing numbers of companies that specialize in AI equipment, software and services, the largest companies developing AI solutions are the existing big tech companies. These companies are primarily based in the US and in China, and the implications are that AI production is first and foremost based in the US and in China. From a European angle, this is a problem. In Europe, the UK is the primary site for AI development and production.

References

Advisory Expert Group in National Accounts (2021) Draft guidance note on 'Recording Artificial Intelligence in the national accounts', 17th Meeting of the Advisory Group on National Accounts, November 2021.

Agrawal, A., Gans, J. and Goldfarb, A. (2019) (eds.) The economics of artificial intelligence – An agenda, NBER.

Baer, A. (2022) DZ.7 Improving the visibility of AI in the national accounts – Global consultation outcomes, Statistics Department, IMF.

Day J. D. and Zimmermann H. (1983) The OSI reference model, in Proceedings of the IEEE, vol. 71, no. 12, pp. 1334-1340, Dec. 1983.

EU (2024) Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence.

Fransman, M. (2002) Mapping the evolving telecoms industry: the uses and shortcomings of the layer model, *Telecommunications Policy*, 26, pp. 473-483.

Jacobides, M.G., Brusoni, S. and Candelon, S. (2021) The evolutionary dynamics of the artificial intelligence ecosystem, *Science Strategy*, 6(4), pp. 412-435.

Lehr, W., Clark D., Bauer, S., Berger, A. and Richter, P. (2018) Whither the Public Internet?, TPRC 46: The 46th Research Conference on Communication, Information and Internet Policy 2018.

Moore, A. W., Martial, H., and Shane, S. (2018) The AI stack: a blueprint for developing and deploying artificial intelligence, Ground/Air Multi-sensor Interoperability, Integration, and Networking for Persistent ISR IX. Vol. 10635. SPIE.

Perspective Economics (2023) Artificial Intelligence sector study – Research report for the Department for Science, Innovation & Technology (DSTI).

Sabanovic S., Milojevic S. and Kaur J. (2012) John McCarthy [History], in IEEE Robotics & Automation Magazine, vol. 19, no. 4, pp. 99-106.

Simon, J.P. (2019) Artificial intelligence: Scope, players, markets and geography, *Digital Policy, Regulation and Governance,* 21(3), pp. 208-237.

Simon, J.P. (2024, forthcoming) Back to industrial policies with a human-centric flavour: the case of artificial intelligence, in Scupola, A, Sundbo, J, Fuglsang, L and Henten, A. (eds.) Handbook of artificial intelligence in services, Edward Elgar.

Søgaard, A. (2022) Kunstig Intelligens Bagfra, Aarhus: Aarhus Universitetsforlag.

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023) What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education, Smart Learning Environments, 10(1), 15.

Tobin, J. (2023) Artificial intelligence: Development, risk and regulation, UK Parliament, House of Lords Library.

Varian, H. (2019) Artificial Intelligence, economics, and industrial organization, in Agrawal, A., Gans, J. and Goldfarb, A. (eds.) The economics of artificial intelligence – An agenda, NBER, pp. 399-422.