AI-Powered Data Products

Jarkko Moilanen

Venkata Pradeep Tatiraju

15th Oct 2023

Contents

Preface

The authors are seasoned experts in the field of data economy, with extensive leadership experience in data productization, machine learning, data exchange, and monetization. We have developed data-driven services and have firsthand insights into the needs of customers.

This book was written with the goal of helping businesses capitalize on the potential of AI and data products. It offers practical guidance for business people who are struggling to navigate the new opportunities presented by these emerging technologies. Additionally, it provides valuable insights for legacy data management practitioners on how to take the next step and incorporate AI into their data-driven strategies.

The authors recognized the historical breakthrough of AI and were particularly impressed by the OpenAI-driven products and services. We saw the potential of combining AI with data products and decided to write a book on the subject. We believed it was time to move beyond the theoretical discussion of "what is a data product" and provide concrete strategies

for businesses to create value through the combination of AI and data products.

The authors come from diverse backgrounds but share a passion for data-driven value creation and innovation. We have worked together on a cutting-edge data platform that utilized decentralization, data exchange, and monetization with the help of machine learning and AI. One author initiated the idea for the book and invited the other to collaborate on this groundbreaking project.

The authors are recognized pioneers in their field, utilizing scientific tools and theories to create real-world value. We wrote the book while applying the principles and concepts discussed in their daily operations, and our passion for the subject kept us motivated through long hours of work.

We hope that our book will inspire businesses to embrace the potential of AI and data products and provide a roadmap for achieving success in the data-driven economy.

1 Introduction

This is an opportunity you cannot afford to miss. Right now, we are navigating the tidal wave of a transformative shift in the global landscape. If you are not currently leveraging the power of AI technologies like ChatGPT, the unsettling truth is that your organization is already a step behind. This isn't intended as a fear tactic, but rather a clarion call to acknowledge the monumental changes and groundbreaking innovations unfolding before our eyes—developments we could only dream of a few short years ago.

In 2022 Altman (CEO of Open AI inc) believes that the capabilities of artificial general intelligence (AGI) will align closely with the intellect of a *"median human that you could hire as a co-worker."*[1] This rather intriguing statement includes the possibility to have AI teams or or at least AI agents as team members. We are not that far from "AI team members" (even without AGI, since first full AI agents driven teams / companies have been tested already. Furthermore, cognitive

---

[1] https://futurism.com/sam-altman-replace-normal-people-ai

clones of humans as decision-makers have been successfully tested. This opportunity lurking in the future is discussed in great detail in the last chapter.

According to McKinsey, the use of AI in data-driven applications is projected to generate around $13 trillion in new global economic activity by 2030.[3] AI holds the substantial potential to revolutionize the data economy in numerous ways. The advent of ChatGPT has truly catapulted us into a new era. Whether your company is just beginning its journey of data productization or already has experience and success with data products, this book, particularly the part on AI, will prove beneficial.

The speed at which society and economies are adapting and growing, fueled by these remarkable opportunities, can indeed feel dizzying at times. Yet, remember that every so-called "revolution" opens up fresh avenues for unprecedented wins and new frontiers. Don't shrink back from change — instead, welcome it with open arms, and ride the crest of this wave. This is your chance to take the reins of this momentous shift—let's seize the day, not let it pass us by!

There are several indicators of the growing interest and investment in Artificial Intelligence (AI) across various sectors and industries. There has been a significant surge in funding for AI startups and projects, indicating that investors are increasingly interested in the potential of this technol-

ogy. There's been a significant increase in the number of AI-related research papers published and patents filed, signaling a rise in innovative activity in the field. There is an increasing demand for AI professionals, including data scientists, machine learning engineers, and AI specialists, indicating the growing importance of AI in business operations. Large technology companies like Google, Amazon, Facebook, Microsoft, and Apple are investing heavily in AI, acquiring AI startups, and integrating AI into their products and services. All the above indicate strong business-level interests and belief in AI to be a profitable bet.

On the technology level, Natural Language Processing (NLP) is a critical component of AI applications for various reasons such as it enables Human-Computer Interaction, Understanding of Unstructured Data, Sentiment Analysis, Information Extraction, and Machine Translations. When you use the ChatGPT, it is enabled by NLP. Thus measuring the use of NLP is a good KPI indicating the rise of AI. Natural Language Processing (NLP) is the top application for Python, the preferred programming language for data science. It constitutes nearly half (49%) of the daily usage of Python's data science libraries. In essence, every second usage of Python in data science is devoted to processing and understanding human language. This demonstrates the significant role NLP plays in current data science applications and the increasing use of AI.

Data is the lifeblood of AI. The performance of AI systems and the accuracy of the insights and predictions they generate are heavily dependent on the quality and quantity of data they are trained on. In the digital era we live in, businesses across various sectors are producing enormous quantities of data through their operations and customer interactions. In 2020, it was estimated that every person created 1.7 megabytes of data per second. When you consider this in the context of a global population of over 7.8 billion people, that's an astronomical amount of data being generated each second!

Customer data offers insights about demand, such as market trends and preferences. Platform data helps optimize operations to enhance customer experience, while supply data aims to understand sourcing constraints. By effectively leveraging this data, businesses can gain a competitive edge, augment their products and services, and bolster their bottom line.

To extract maximum value from this wealth of data, a network of organizations creates a digital ecosystem. In this ecosystem, data is shared, analyzed, and transformed into actionable insights that benefit all parties involved. This cooperative approach allows businesses to access a broader data pool, enhancing their decision-making capabilities and driving innovation. The power of a digital ecosystem lies in

its collective intelligence, turning raw data into strategic assets that fuel business growth and success.

The data economy has also sparked the emergence of innovative business models, including data brokers and analytics companies. These businesses specialize in gathering, analyzing, and selling data to other companies. This development has given birth to a burgeoning data market, wherein companies can capitalize on their data assets and cultivate new revenue streams. According to a study by MIT, more than half (53%) of business leaders reported that their participation in the data economy has driven the creation of new business models. [15].

The conservative estimate of the market size for the data economy in the global hospitality industry alone was US $43.2 billion in 2018, and it has been doubling its size every three years.[7] The value of the data economy of EU27 was almost €325 billion in 2019, representing 2.6 % of GDP. The same estimate predicts that it will increase to over €550 billion by 2025, representing 4 % of the overall EU GDP.[6]

Data Exchange started the Data Economy

A crucial element of the data economy is the exchange of data among diverse entities, including businesses, governments, and individuals. This data exchange is enabled by a variety of technologies, such as cloud computing, big data

analytics, and artificial intelligence. The data exchange acts as a critical catalyst for innovation and economic growth in the data economy.

By sharing data, businesses can innovate new products and services, enhance operations, and gain a better understanding of customer needs and preferences. This can result in heightened efficiency, productivity, and profitability, as well as improved customer satisfaction. In essence, participation in the data economy offers businesses a significant advantage in today's rapidly evolving digital landscape.

Data exchange is the practice of sharing data among distinct entities or groups. This process entails the secure transmission of data from one organization to another, fostering collaboration, insight generation, and value creation. The mechanisms of data exchange can vary and include partnerships, marketplaces, or formal agreements.

The primary goal of data exchange is to encourage data flow, thereby allowing organizations to tap into varied data sources, glean insights, and improve decision-making. Its emphasis is on the act of data sharing itself, irrespective of any associated direct financial transaction. While data exchange and data monetization are interconnected concepts, they diverge in their central focus and objectives.

Data Monetization growing fast

Data monetization, a crucial facet of the data economy, involves transforming data into revenue or extracting economic value from it. This process essentially capitalizes on data as an asset to generate financial returns. Data monetization can adopt various forms such as directly selling data to third parties, licensing data to other organizations, or utilizing data to develop and offer value-added products or services.

The primary aim of data monetization is to convert data into a source of revenue, whether through direct transactions or by creating new business models based on data-driven insights. Its focus is on the economic value that can be derived from data and the financial benefits it can yield. Notably, data monetization is a key driving force of economic growth in the data economy. It underscores the compelling monetary potential that data holds when harnessed effectively.

By monetizing their data assets, businesses can create new revenue streams, increase profitability, and gain a competitive advantage in their markets. Data monetization can also lead to new business models, such as subscription-based services and data marketplaces. The global data monetization market size stood at USD 2.60 billion in 2022. The market will expand from USD 2.99 billion in 2023 to USD 9.10 billion by 2030, exhibiting a CAGR of 17.2% over the

estimated period.[14]

Not all data are equally valuable

Data is the lifeblood of modern organizations. Yet, not all data holds the same value, and this is a distinction many data teams have often overlooked. Recently, we embarked on an ambitious journey to categorize the data of one customer into various tiers of criticality, ranging from 1 to 4. The exercise contained a workshop in which collection of over 300 datasets were analyzed, with aim to identify which ones would fall under tier 1. By the close of our discussions, we found that approximately 4% of the datasets were of this utmost criticality. The decision-making process was not a walk in the park. It was filled with engaging debates and illuminating discussions, making the journey just as valuable as the end result. But one might wonder, why undergo this rigorous exercise?

There are two central motivations. Firstly, efficiency and focus. Data teams, while skilled, have capacity limits. Assigning tiers to our data enables us to channel our energies where they're needed most. Secondly, clarity for end users. By making the criticality of data transparent, users can gauge the degree of trust they should place in a particular dataset.

To assess the criticality of a dataset, it's essential to evalu-

ate its impact on several key areas, such as core operations, user experience, legal obligations, rider safety, and financial reporting. But identifying a dataset's tier is only half the battle. The real challenge lies in how we manage and treat this data. For each tier, we've outlined distinct management expectations and requirements. For instance, a dataset categorized under tier 1 necessitates more stringent measures, such as faster incident response times and superior quality assurance, compared to those in tiers 3 or 4.

A few lessons we've gleaned from this experience. Firstly, the process, while challenging, is immensely gratifying. Secondly, seeking the "perfect" classification immediately is a folly; it's more practical to adopt an iterative mindset, refining criteria as we progress. Finally, data tiering isn't a one-and-done task; it requires continual reassessment as conditions evolve.

Do the above sounds familiar? It should since that is Master Data Management (MDM), which is a comprehensive method of enabling an enterprise to link all of its critical data to a common point of reference. When done properly, MDM streamlines data sharing among personnel and departments. In today's digital world where data drives decisions, having a robust MDM strategy in place is crucial for ensuring that decisions are based on accurate, up-to-date, and reliable data. As businesses grow and diversify, the need for a

central source of consistent data becomes even more critical, making MDM an essential practice for modern enterprises.

MDM has been in use for a long time already as a tool, but what has changed lately is the approach. Now the data product concept is seen as a key element. That approach emphasizes the commodity value of data instead of treating it as a pure technical asset. As you can see from this simple example, identifying the MVPs (Most Valuable Product) of your data is the key to efficiency and value creation. The above-mentioned 4% of the datasets are what we label as core data products.

Data has become heavy

At the same time, AI is becoming omnipotent, and the volume of data being generated worldwide has increased exponentially in recent years, driven by factors such as the growth of the internet, the proliferation of smart devices, and the rise of social media. This has resulted in a flood of data that organizations must collect, store, manage, and analyze in order to gain insights, make informed decisions, and remain competitive. Replicating data around the systems over the web is no longer an efficient solution. In other words, data has become heavy and "wants" to stay near the creating system.

In addition to the volume of data, the complexity of data is also increasing. Data is no longer limited to simple struc-

tured information such as numbers and text but now includes multimedia content, unstructured data such as social media posts and comments, and data generated by sensors and Internet of Things (IoT) devices. Managing and analyzing this diverse and complex data requires new tools, technologies, and expertise.

Overall, the term "heavy data" reflects the growing challenge of managing and analyzing the massive and complex data sets that organizations must deal with today. As a result, there is a growing demand for data management and analytics solutions that can handle this "heavy" data, including big data platforms, machine learning, and artificial intelligence.

In data management, the term commonly used for this phenomenon is **data gravitation**. Data gravity is the observed characteristic of large datasets that describes their tendency to attract smaller datasets, as well as relevant services and applications. It also speaks to the difficulty of moving a large, "heavy" dataset. Large datasets are attractive because of the diversity of data available. They are also attractive (i.e. have gravity) because of the technologies used to store such large datasets. Placing data in one central arena means that data gravity will not collect slowly over time, but rather increase significantly in a short time.

Data gravity presents data managers with two issues: latency and data nonportability. By its very nature, a large

dataset requires the applications that use it to be close, in its orbit, or suffer latency. This is because the closer the applications are to the data, the better the workload performance. Data gravity increases with the size of the dataset, and the larger the dataset, the more difficult the dataset is to move. After all, moving a planet would be quite a feat. Moving vast quantities of data is slow and ties up resources in the process.

So what to do? One plausible approach is to take the data processing near the source. **Edge computing** is an emerging computing paradigm that refers to a range of networks and devices at or near the user. Edge is about processing data closer to where it's being generated, enabling processing at greater speeds and volumes, leading to greater action-led results in real-time.

Data products drive Data Economy

At this point is needed to remember the concept of data fabric. Data mesh emphasizes data decentralization, autonomy, and productization. In contrast, the data fabric architecture advocates centralization and unified data access. Both are valuable approaches to democratizing data and insights, but they differ in their underlying philosophy and architecture. The more decentralized your organization and the more autonomy your teams have in their day-to-day operations, the

more likely it is that a data mesh will prove more effective than data fabric, which is better suited for centralized organizations with a more hierarchical structure. While the data fabric seeks to build a single, virtual management layer atop distributed data, the data mesh encourages distributed groups of teams to manage data as they see fit, albeit with some common governance provisions. Despite the differences, attempts to combine the pros of both concepts have been done and indicate that it's not that black and white question. It's more about what your business and objectives require. In this book, we chose to approach data from productizement and servitization point of view, which also combines the aforementioned 'rival' models.

As previously discussed, data exchange revolves around sharing data to stimulate collaboration and insights, while data monetization aims to draw financial value from data assets. Data exchange often serves as a precursor or enabler for data monetization, as it allows organizations to access and obtain data from various sources, which can then be used for monetization objectives. Both concepts are pivotal in the data-driven economy and are reliant on data products. Data exchange establishes the groundwork for collaboration and innovation, while data monetization empowers organizations to tap into the economic potential of their data assets.

The concepts of the data economy and data products are

intimately linked. The data economy encompasses economic activities founded on data collection, processing, analysis, and dissemination. It includes all businesses, organizations, and individuals involved in creating, managing, and utilizing data. Therefore, within this ecosystem, data products serve as valuable assets that facilitate data exchange and enable data monetization.

Product mindset is the key

Data products, on the other hand, are products that are created using data as the primary input. These physical or digital products may include software, applications, reports, analytics, and insights. The Economist wrote an article in 2017 with the title *Data is giving rise to a new economy* and declared that *"data are to this century what oil was to the last one: a driver of growth and change."*[9]. More recently, Harvard Business Review published an article *Approach Your Data with a Product Mindset* which emphasizes the need to *"abstracting these solutions into [...] products that can be used to address additional needs down the line, and packaging these products in a way that drives adoption by stakeholders across the organization."* In this approach, data is treated as a commodity rather than a pure technical asset.

The Product mindset establishes a foundation for effective data exchange and the ability to generate revenue. It seems

evident that data products in different forms have become a standard block in the data economy. This has also changed the skills requirements in companies and given space for new positions like Data Product Managers.[21] We'll discuss this keyrole in more detail in Chapter *Build winning teams.*

In this book, we emphasize that Data Products are created with a specific purpose in mind through a cyclic process. The starting point is typically a use-case, a customer problem, or feedback concerning existing data products. The model applied in this book, depicted in Figure 1.1, can be used across all data product development, irrespective of whether it's intended for internal or external use. Moreover, this model can be utilized to iterate new versions of the product to customers, accommodating their evolving needs and prolonging the lifecycle of a data product.

To preempt unforeseen issues down the line, it's crucial to include legal representation as part of the team. This ensures all legal aspects are adequately addressed, circumventing potential disputes that could escalate to court proceedings.

The use-case is analyzed as part of the business design with the help of mockups. The aim is to find product-market fit by creating mockups until the customer says *take my money!* You stay in this cycle until you get the green light or decide to ditch the idea for other business reasons.
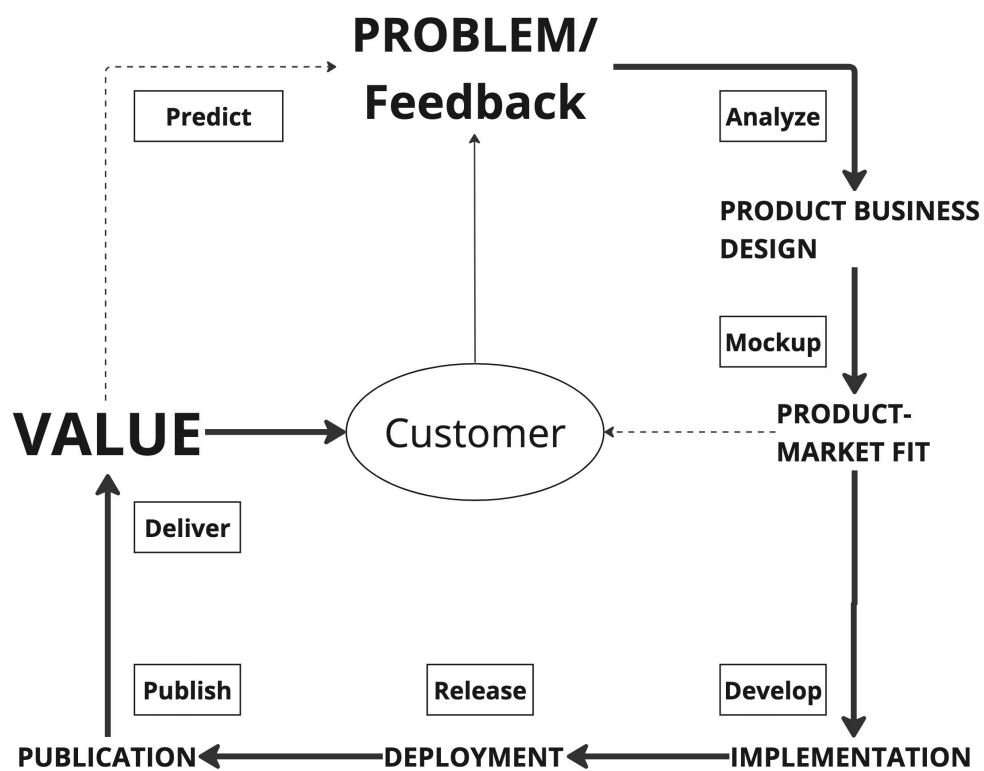
Figure 1.1: Data Products are developed and eventually delivered as value to customers in phases.

Once the design of the data product, including the business model, pricing, and initial data model, is finalized, and the data sources identified, it's time for implementation, provided that the expected revenue justifies the investment. The implementation phase is primarily technical work.

After implementation, it's time for a release. At this juncture, it's crucial to ensure all security and privacy issues have been addressed (ideally, if proper data governance is in place, this should already be the case). Additionally, contemplate and implement strategies for the customer's adoption of the product. Consider the customer's perspective, testing the ease or difficulty of onboarding and value creation with your data product. Develop guides, optimize the customer journey, refine documentation, and possibly provide code snippets. These efforts accelerate value creation and help reduce the 'Time to First Success' timespan.

When you're confident that the data product package is well-rounded, all automated tests are successful, and the customer experience (CX) is satisfactory, you can publish the release to the market (marketplace, catalog). However, don't forget to test the customer paths in the platforms you use after publishing. It would indeed be a shame to have a great data product in the market that, for some reason, customers cannot see, find, use, or pay for!

Many organizations halt their efforts once the data product

is published in the marketplace or catalog. However, it's essential to remember that *your goal isn't merely to publish, but to deliver value! Completing the last mile in value delivery is crucial.* If you stop at publishing, you're focusing on the product and likely missing the broader business perspective. Ensure that you communicate to customers that the solution is available and encourage them to use it. You've achieved your goal when customers use the data product and derive value from it. Thus, when setting KPIs for data products, it's important to consider the business value.

Even after achieving this, your job isn't done. An additional task is to anticipate possible customer challenges and future needs. Doing so can speed up value delivery and enhance the customer experience. Analytics and AI can assist in achieving this.

Throughout this book, we will revisit the aforementioned model and delve into detail about how AI can be applied in each phase.

One might wonder why the data economy hasn't seen more rapid growth yet. Currently, the leading data marketplaces offer limited data products, and it's often a challenge for data consumers to find the specific dataset that addresses their particular problems on these platforms. It's worth noting that we are still in the early phases of the data economy. Although comparing data to oil isn't entirely fitting, it's in-

sightful from a business perspective.

Decades passed before a well-functioning market for oil was established. Paradoxically, it was Standard Oil, John D. Rockefeller's late-19th-century monopoly, that accelerated this process. It helped create the technology and— as suggested by the firm's name— the standards that enabled the new resource to be traded. In contrast, the evolution of data is not expected to take as long, given the decreasing time span needed for emerging technologies to become mainstream, transitioning from hype to commonplace.

One significant challenge with data has been the lack of product standards, leading to stagnating markets. The concept of Data Products, widely adopted in recent years, stands as a cornerstone of the Data Mesh. Efforts to standardize Data Products[2] do exist, and a breakthrough in this area will invigorate the data markets. We'll explore standardization further in Chapter *AI-Powered Data Product portfolio*.

The data economy lays the groundwork and outlines the structure for the creation of data products. It enables businesses and organizations to collect, process, and turn large amounts of data into valuable insights and products. Data products, in turn, fuel the growth and sustainability of the data economy by creating new business opportunities, enhancing decision-making, and improving customer experi-

---

[2]See for example https://opendataproducts.org/

ences. Overall, the data economy and data products are intertwined and mutually beneficial, with the data economy serving as the base for creating and nurturing data products. The concept of Data Products is explored in more detail in *Chapter 2 AI-Powered Data Products*.

The Rise of Artificial Intelligence

The data economy is expanding rapidly, but a relatively new player has recently entered the market - Artificial Intelligence (AI). After the launch of ChatGPT on November 30, 2022, the world witnessed significant changes. Until that point, Artificial Intelligence (AI) was a field exclusive to highly specialized experts. However, ChatGPT brought AI into the everyday lives of countless individuals. Despite its widespread use, the backstory of ChatGPT remains unknown to many. Rooted in extensive research within the AI field, let's delve into its origin.

AI traces its origins back to the early 20th century with the development of electronic computers and the concept of creating machines capable of simulating human intelligence. The term "artificial intelligence" was first used in 1956 by John McCarthy, Marvin Minsky, Nathaniel Rochester, and Claude Shannon during the Dartmouth Conference, a summer workshop aimed at exploring the computing potential for mimicking human intelligence.

Despite this, the progression of AI was relatively slow over the subsequent decades, limited by computing power and a *lack of data.* According to a 2020 survey, the top four of the five obstacles to implementing AI/ML initiatives include data accuracy (54%), data access (44%), the protection of personal and sensitive data (43%), and the time required to refresh data models (36%) [8].

It wasn't until the 1990s, with the advent of the internet, that AI began to pick up speed. Access to vast amounts of data, coupled with the development of new machine learning algorithms, paved the way for AI applications in diverse fields, ranging from image recognition to natural language processing.

In recent years, breakthroughs in deep learning, natural language processing, and computer vision have sparked significant advancements in AI, leading to revolutionary developments in domains such as autonomous driving, healthcare, and finance. For instance, a study within healthcare found that chatbot responses were preferred over physician responses and were rated significantly higher for quality and empathy[11]. In this study, which consisted of 95 questions and responses, evaluators preferred chatbot responses to physician responses in 78.6% of the 585 evaluations.

In the financial sector, the importance of leveraging Artificial Intelligence (AI) is gaining recognition. A survey from The

Economist Intelligence Unit reveals that 77% of bankers believe that unlocking the value of AI could mean the difference between their bank's success or failure[10]. Moreover, in a 2021 McKinsey survey, 56% of respondents reported using AI in at least one function within their organizations[5]. The top three applications identified in this survey are service-operations optimization, AI-based product enhancement, and contact-center automation.

Since AI relies on data to function, it's no surprise that top-performing organizations in the financial sector, which are effectively utilizing AI, also place a significant emphasis on data management. The aforementioned McKinsey survey found that leading organizations proactively scan training and testing data to detect underrepresentation of protected characteristics and/or attributes. Data professionals within these organizations are dedicated to checking for skewed or biased data during the data ingestion process, and they work to increase the representation of protected characteristics and/or attributes in training and testing data as necessary.

One of the most frequently discussed possibilities of AI is in the realm of autonomous driving. Companies like Tesla, Waymo, and Alibaba have made significant strides towards the development of fully autonomous vehicles, largely powered by AI. In these vehicles, AI is used to simulate the characteristics of a human driver, enabling the vehicle to see,

hear, think, and make decisions based on the data gathered from various onboard sensors.

Take AutoX, for instance, which creates retail-based autonomous vehicles and operates fleets for robotaxi and driverless grocery delivery services. Another example is SapientX, a company that develops white-label software for adding conversational voice and intelligence capabilities to a wide range of technological products, including those in the automotive, video conferencing, health, and air travel sectors. Then there's Tesla, renowned for its electric cars, which come equipped with Autopilot—a suite of driver-assist features that allow for automatic steering, acceleration, braking, lane changing, and parking.

In essence, although AI has been evolving for over half a century, it's only in recent years—with advancements in computing power, data availability, and algorithmic innovation—that we've come closer to realizing AI's potential to revolutionize industries and improve our lives. This AI boom that now impacts almost all facets of our existence was set into motion by OpenAI.

OpenAI was established in 2015, originally focusing on the development of AI and machine learning tools for video games. They unveiled OpenAI Gym—an open-source toolkit for developing reinforcement learning algorithms. By 2018, they had introduced the first iteration of GPT, a neural network

trained on vast datasets to produce specific outputs. In 2019, OpenAI transitioned into a for-profit entity, overseen by OpenAI Inc. Fast forward to 2021, and they introduced Dall-E, an AI model capable of generating images from natural language text descriptions.

Perhaps their most well-known product is ChatGPT, which was released in 2022. This chatbot can provide answers on a wide range of topics, sparking ongoing debates about its benefits and drawbacks.

Artificial Intelligence and the Data Economy

According to McKinsey, the use of AI in data-driven applications is projected to generate around $13 trillion in new global economic activity by 2030.[3] AI holds substantial potential to revolutionize the data economy in numerous ways. Let's delve into some potential benefits of incorporating AI in the data economy.

Firstly, it's all about enhanced data analysis. AI can aid organizations in analyzing vast amounts of data in real-time, highlighting patterns and insights that might otherwise be hard or even impossible for humans to identify. This leads to more informed decision-making, increased efficiency, and cost savings.

Secondly, AI paves the way for personalized products and

services. By analyzing customer data and behavior, organizations can tailor their products and services to meet individual customer needs and preferences, leading to improved customer loyalty and revenue growth.

Thirdly, AI brings automation to the table. It can automate numerous manual tasks, such as data entry, cleansing, and processing. This liberation of human resources allows them to focus on tasks that hold higher value.

Fourthly, AI can be employed for predictive analytics. By constructing predictive models, AI can forecast future trends and behaviors, enabling organizations to plan and strategize effectively.

Lastly, AI can augment cybersecurity measures. By analyzing large volumes of data, AI can identify potential threats and vulnerabilities, as well as detect anomalies and patterns that may be indicative of malicious activity.

These potential benefits are just the tip of the iceberg; the AI revolution is just getting started in the realm of the data economy. We anticipate witnessing profound changes in data governance, monetization opportunities, customer experiences, and value creation. We foresee the rise of almost fully autonomous, AI-powered data products that leverage AI across all facets of the product lifecycle.

As AI transitions from being a co-pilot that assists humans

in performing better, to taking a more commanding role, it raises an important question—what will the role of humans be in this new landscape? We'll delve into this topic in the final chapter of the book. But before that, we'll explore how you can harness AI with your data products and value creation in the here and now. This book aims to offer immediate value and provide strategic guidance on leveraging AI now and in the future.

The book contains 6 chapters

The Introduction (Chapter 1) may have struck an academic chord, mainly due to its grounding in data economy-related literature. But let's look at what's ahead. The subsequent chapters share a dualistic structure where we first establish a foundation for data products and related functions, and then apply AI to it.

In Chapter 2, "*AI-Powered Data Products*," we'll explore the concept of data products and the importance of focusing on data productization. We'll delve into the critical role of data products in data exchange and monetization in detail. You'll learn the differences between datasets and data products. In the second part of this chapter, we shift our focus to AI, and how it is being used today to create value-driving data products.

Chapter 3, "*Objective-Driven Quality Data Products*," is all

about rethinking your data. To be able to create data products or services, you need to understand the data you have—its location, interconnections, associated costs, and quality. Following this, we'll once again introduce AI into the conversation, sharing insights on how AI can be applied to deliver value with data products.

Chapter 4, *"Value-based selling and Measuring Success,"* is all about re...

In Chapter 5, *"Build Winning Teams with AI copilots,"* we reveal the makings of a typical successful data product team and the necessary skills to create profitable and appealing data products sustainably. As with the previous chapters, we end with an AI perspective, providing insights on how AI can be used in team management, fostering team spirit, and maintaining a flow state.

Chapter 6, *"AI-Powered Data Product Portfolio,"* presents a practical model for managing AI-enhanced data products with clear governance, guides, and actionable advice. Managing a successful data product portfolio is a far cry from data governance. Optimal Data Product Portfolio management maximizes customer experience across all products. At the end of this chapter, we offer methods to enhance portfolio management using AI.

The final chapter, *"The Rise of Intelligent Value Delivery,"* provides a glimpse into the expected trajectory of data product

development, sales, and management. We anticipate that AI will evolve from a supporting role, aiding humans, to becoming an active participant in the data economy. However, this isn't about letting AI run rampant, but rather about harnessing its capabilities within ethical boundaries and rules.

Meet the Authors

The authors are CxO professionals in platforms, data and API economy with more than 30 years of C-level experience.

Jarkko Moilanen

Ph.D. (Information science), Chief Data Officer background with passion towards AI and data-driven value delivery

Jarkko Moilanen, Ph.D. is the Chief Data Officer of Vastuu Group Oy. Jarkko has a key role in managing the portfolio of advanced and innovative data-driven development customer cases, and managing data strategy as well as data management. Prior to his employment at Vastuu Group, he served as the Chief Operations Officer and Chief Development Officer at Platform of Trust (Data as a Service platform).

Earlier, Jarkko was for 4 years leading the digitalization in K12 education in Finland as the Senior Advisor on Digital

Strategy at the Ministry of Education and Culture of Finland. One key achievement was to lead the development of government controlled secure nation-wide identification system for K12 education. Identification system was adopted by businesses operating in education sector fast and it is now used by millions of students and teachers in Finland.

Moilanen earned his doctorate degree in 2017 at University of Tampere with his work entitled "Peer Production economy - revolution in design, development and manufacturing". He is currently pursuing his 2nd PhD in the University of Jyväskylä around the design-driven data productizement process, which binds together data products, AI agents and data strategy in companies.

Jarkko is the creator of the Data Product Toolkit, which helps organizations to take advantage of data, design data products to maximize internal reuse and partner network value, and eventually, monetize derived data-driven services. As part of the toolkit development, Jarkko has defined first in the world Open Data Product Specification which is a vendor-neutral, open-source machine-readable data product metadata model. It defines the objects and attributes as well as the structure of digital data products.

Venkata Pradeep Tatiraju

Ph.D. candidate (Operations Management and Quantitative Techniques), IIM Indore, Data Scientist

Venkata Pradeep Tatiraju is an accomplished Data Science professional with 11 years of experience in the field. He has a strong passion for utilizing data to solve business problems and has worked with Fortune 100 organizations to establish data science and analytics practices, contributing to their growth. With expertise in Machine Learning, Natural Language Processing (NLP), and Deep Learning, he has successfully developed and deployed AI models on cloud and on-premises platforms.

In his recent role as Lead Data Scientist at ADS Securities, Pradeep developed data science strategies for marketing and sales teams. His skills in statistics, machine learning, optimization, and simulation enabled the creation of impactful campaigns for cross-selling, upselling, growth incentives, and customer retention. He also excelled in integrating ML models with existing systems and establishing efficient ETL pipelines for data processing.

Pradeep has delivered successful AI modules and solutions

throughout his career across various domains, including healthcare, retail, banking and financial services, supply chain, and education. His projects have ranged from predictive modeling for healthcare products to implementing knowledge management systems with NLP capabilities. He has also deployed deep learning models on edge devices and excelled in areas such as customer call intent prediction, campaign performance prediction, sentiment analysis, and segmentation models.

Pradeep is pursuing a Ph.D. from IIM Indore and holds an M.Sc (Hons) and B.E.(Hons) from BITS Pilani Hyderabad Campus. His comprehensive knowledge spans quantitative models, data and visualization tools, programming languages, cloud platforms, and dev ops.

Beyond his professional achievements, Pradeep has actively engaged in extracurricular activities and demonstrated leadership abilities. He served as President of the Engineers Without Borders (NGO) Chapter at BITS Pilani and held significant positions in the Biological Science Association and the sports club during his college years.

With an unwavering dedication to advancing data science practices and a proven track record of delivering impactful

outcomes, Pradeep brings a wealth of knowledge and experience to the field.

Bibliography

[1] M Bansal. Council post: Flying blind: How bad data undermines business. 2022.

[2] Ujval Bucha. Dear product manager, don't fall in love with your product., 2020.

[3] Jacques Bughin, Jeongmin Seong, James Manyika, Michael Chui, and Raoul Joshi. Notes from the ai frontier: Modeling the impact of ai on the world economy, Sep 2018.

[4] ChatGPT. Task-driven autonomous agent utilizing gpt-4, pinecone, and langchain for diverse applications.

[5] Michael Chui, Bryce Hall, Alex Singla, and Alex Sukharevsky. The state of ai in 2021, Dec 2021.

[6] European Commission. Building a data economy - brochure, 2021.

[7] Diane Coyle and Wendy Li. The data economy: Market size and global trade. 2021.

[8] Delphix. Data is key to unlocking ai/ml, 2020.

[9] The Economist. Data is giving rise to a new economy, 2017.

[10] The Economist. Forging new frontiers: Advanced technologies will revolutionise banking, 2020.

[11] Ayers JW et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum.

[12] Mariia Golovianko, Svitlana Gryshko, Vagan Terziyan, and Tuure Tuunanen. Responsible cognitive digital clones as decision-makers: a design science research study. *European Journal of Information Systems*, pages 1–23, 2022.

[13] Government Data Quality Hub. Government data quality. hidden costs of poor data quality. 2021.

[14] Fortune Business Insights. Data monetization market - market research report. 2023.

[15] MIT Technology Review Insights. Capitalizing on the data economy, 2021.

[16] Joona Keränen, Dirk Totzek, Anna Salonen, and Mario Kienzler. Advancing value-based selling research in b2b markets: A theoretical toolbox and research agenda. *Industrial Marketing Management*, 111:55–68, 05 2023.

[17] Yong Kim. Customer experience design for smart product-service systems based on the iterations of experience–evaluate–engage using customer experience data. *Sustainability*, 15:686, 12 2022.

[18] Shirish Limaye. Removing friction from the customer journey. *"Imperial Journal of Interdisciplinary Research (IJIR)*, VII:70–77, 09 2018.

[19] Evynn McFalls. The neuroscience of storytelling, Sep 2021.

[20] Joon Sung Park, Joseph C O'Brien, Carrie J Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. Generative agents: Interactive simulacra of human behavior. *arXiv preprint arXiv:2304.03442*, 2023.

[21] Davenport Thomas, Bean Randy, and Jain Shail. Why your company needs data-product managers, Oct 2022.

[22] John Tumaku, Jianxin Ren, Kwabena Boakye, Kwame Ofori, and Aidatu Abubakari. Interplay between perceived value, trust and continuance intention: evidence in the sharing economy. *International Journal of Quality and Service Sciences*, 15, 02 2023.

[23] Paulus Wardoyo, Endang Rusdianti, and Sri Purwantini. How important are value-based selling and courteous selling behavior in improving salesforce perfor-

mance? *Integrated Journal of Business and Economics*, 7:295, 02 2023.

[24] Xue Yang. Consumers' purchase intentions in social commerce: the role of social psychological distance, perceived value, and perceived cognitive effort. *Information Technology & People*, 35:330–348, 10 2022.

[25] Alexander Zauner, Monika Koller, and Isabella Hatak. Customer perceived value—conceptualization and avenues for future research. *Cogent psychology*, 2(1):1061782, 2015.