International Journal of Artificial Intelligence, Data Science, and Machine Learning
Grace Horizon Publication | Volume 3, Issue 4, 38-46, 2022
ISSN: 3050-9262 | https://doi.org/10.63282/3050-9262.IJAIDSML-V3I4P105

Original Article

Redefining Data Products in the Age of Artificial Intelligence and Deep Learning

Lalmohan Behera
Senior IEEE Member and IETE Membership.

Abstract *- AI and deep learning technologies are rapidly evolving and these advanced technologies are altering the approaches towards generating, maintaining, and using data solutions. The traditional data products were thus initially catered for descriptive analytics especially on past data. However, the integration of AI and DL has caused a shift to normal from where the organization can achieve predictive and prescriptive analytics and decision-making, as well as learning systems that can adapt themselves. This paper focuses on how technological advancements and their subtopics concerning the data docket have changed due to the rise of AI-powered data products. With the help of indexing, issues related to data quality and bias, interpretability of AI, and data security are discussed in the context of the AI ecosystem. Cloud computing and edge AI, as well as federated learning, are examined in order to understand the effect they have on today's data product architecture. Moreover, there is a transition to real-time data processing and intelligent automation, and the need for organizations to employ the AI-native architecture is outlined. Thus, with the help of redefining data products, organizations can make more business sense, deliver exceptional customer experience, and facilitate the creation of new data solutions. This paper discusses the emerging trends and the guidelines needed to support organizations to achieve the optimal functionalities of AI and Deep Learning in developing the next generation of intelligent data products.*

Keywords *- Artificial Intelligence, Deep Learning, Data Products, Real-time Data Processing, Machine Learning.*

1. Introduction

Artificial intelligence (AI) and deep learning (DL) advancements have accelerated and brought significant changes to data products regarding understanding, management, and data utilization. Typically, data products were used to contain structured and semi-structured information for descriptive analytics and mainly for reporting past events. [1-3] These products allowed organizations to draw from past experiences to direct the business decisions from future contingencies analysis. However, with the help of AI and DL, the role of data products has changed significantly, and the old paradigm of data warehousing has transferred into newer generation active data warehouses and ecosystems, forming real-time smart and learning systems able to carry out even predictive and prescriptive analytics. Real-time data analysis based on such products fully applies artificial intelligence with machine learning to analyze large volumes of structured and unstructured data to generate fresh patterns to be adopted in different stages. The use of deep learning models adds to these capabilities, which entail decision-making without human intervention, detecting anomalies in the input data, giving recommendations depending on the input data, and self-learning systems that adapt to new data inputs. This brings some major changes in the business, government, and research areas because the data products are static, evolving, and growing.

However, several challenges exist when it comes to AI-driven data products, and they include the following. Some of the important challenges in deploying AI solutions include aspects of the AI system, such as data bias, model interpretability, scalability, and security considerations. Moreover, with the current advancement in AI implementation, there is a need to put in place data management measures to enhance AI implementation in an ethical way. Business entities must reconsider the information architecture, transitioning towards cloud and edge computing to address the challenges. As for monetization, a rather new and promising trend is seen in the monetization of AI-driven data products and services, as more and more companies are searching for new kinds of revenues based on data and intelligent data services. Through this paper, an attempt is made to redefine the data products in the context of AI and DL, examining the changes that may occur in its role, the technologies that may have initiated the change, and finally, identifying associated challenges and opportunities. Through such insight, an organization can place itself in the right strategic place to incorporate AI AI-driven data products and hence have an added vantage point in an ever-expanding data-oriented market.

2. Evolution of Data Products in the AI Era

The comprehension of data products has steadily developed and opened up in response to the changes in the approach toward manipulating data. [4-7] First, data products were developed as descriptive ones, and their goal was to make simple and structured reports, dashboards and business intelligence tools for analyzing past trends. However, according to the evolution of artificial intelligence (AI) and deep learning (DL) abilities, data products have become enriched, contextually intelligent, analytically innovative, and capable of using analytical types, including predictive and prescriptive. This has resulted in the origins of innovative data commodities in the application of Artificial Intelligence, where data can learn from itself, make decisions and improve organizational functions.

2.1. Traditional Data Products

Legacy data products were initially designed to deal with and warehouse data from traditional transactional, ERP, and CRM systems, mostly structured data types. They include the following products: relational databases, the umbrella under which most products stand; data warehouses; and business intelligence (BI), which help generate reports and depict historical trends. These products' main purpose was to combine data, keep it consistent, and perform analysis by running predefined queries. At the initial steps of the digital transformation, the companies used extract, transform and load (ETL) tools for data accumulation. Data products in this era were mainly informational and postmortem, providing the information that a business needed to perform better and insights into its customers, processes, and operations. However, these systems did not incorporate dynamic updating of models, improvement of queries or changing data paths when new information occurred. Moreover, old data products broke down rapidly when dealing with unstructured formats of information, for example, images, videos, social media, and so forth, rendering it impossible for them to analyze different data types.

2.2. AI and Deep Learning-Driven Data Products

All data products have been transformed from simple data tools that can be used to find and review data to smarter and self-learning systems enabled by AI and deep learning. While most of the data products developed up to this point are based on set rules and predefined queries, AI data products use Machine learning algorithms to parse huge volumes of structured and unstructured data, superimposing on-demand patterns, trends and discrepancies. These products employ techniques like NLP, computer vision, and reinforcement learning to obtain more insights and automate decision-making. AI-supported data products keep improving their models by updating the knowledge from new data, thus increasing the correctness rate.

This makes it possible to perform prescriptive analytics where one can forecast the customer needs, identify the risks and prevent them, and optimize the operations. For example, recommendation systems based on deep learning bring virtual life suggestions to e-commerce and media streaming services. In contrast, AI-based fraud identification systems analyze ongoing financial operations to identify potential scams. On the same note, edge AI and federated learning are beneficial because they enable data products to analyze information with minimal delays and ensure high data privacy. Therefore, it is crucial to understand the vectors of AI adoption and expand it to new data products. However, we also need to address the new issues, including the explicability of the model, the question of bias, and the data protection problem. With advanced and progressive AI solutions being adopted today, organizations should have proper guidelines to govern AI to avoid the negative impacts of AI. Nevertheless, applying data products based on AI technology is a powerful opportunity for companies to enhance their revenues, decrease costs and foster innovation as the role of data grows more vital.

Table 1: Comparison of Traditional vs. AI-Driven Data Products

Feature	Traditional Data Products	AI-Driven Data Products
Data Processing	Batch processing	Real-time, dynamic learning
Personalization	Rule-based recommendations	Adaptive, behavior-based predictions
Scalability	Limited by predefined rules	Scales dynamically with more data
Decision-Making	Human-driven analysis	AI-driven automation and insights
Adaptability	Static requires manual updates	Self-learning and evolving models

3. Key Components of AI-Enabled Data Products

Artificial Intelligence depends on a hierarchical web of components that interact to take data, collect it, transform it, analyze it, and provide meaningful information. In contrast to most new data products, where the principal mode of data utilization is the storing and searching for structurally formatted information, new AI data products need exceptional data pipeline structures that allow for constant learning, real-time data processing, and decision-making. [8-11] Some main steps involved in such systems are data ingestion and manipulation, model creation and training, model deployment and monitoring. Among these, data ingestion and processing act as the core, and model development and training are the intelligence of the data products.

3.1. Data Ingestion and Processing

Data ingestion can be defined as the process of collecting data in the context of the development of AI-enabled data products, and it is the first and most important step. It includes collecting data from various structured databases, semi-structured logs, unstructured texts and videos, image data, audio data, live sensor data and many more. AI-based data products must process large volumes of diverse data from IoT gadgets, web interfaces, social media, and business applications. This was designed for high-velocity data and the means to handle high volume and varying types of data streams.

Data processing, as the term suggests, involves purification, formatting, and cataloguing of the data to be fed to the data analysis system due to ingestion. Preprocessing is a vital step towards ensuring that the models are fed with the most relevant AI data, and maintaining the quality of the data input for the AI models is critical since it can significantly impact their performance. The current application of Artificial Intelligence entails data-based applications that rely on distributed computing environments such as Apache Spark and big data warehouses for data crunching in large data. Further, real-time data streaming platforms such as Apache Kafka make the process more efficient, allowing AI to utilize the results more or less in real time. As mentioned previously, data annotation and labelling also play a significant role in ML, especially in supervised learning model processing. HITL strategies and AI-based assistants facilitate better data labelling to enhance the model learning quality and a clear, meaningful set of annotations. Furthermore, organizations need to implement effective data structures to meet the required data protection standards posed by the GDPR and the CCPA.

3.2. Model Development and Training

The most important step after data ingestion and preparation is the model creation and training. They entail developing and constantly deploying machine learning and deep learning solutions capable of making sound decisions based on featured patterns. The model selection is based on the specific type of data product, from a simple linear model to analyze trends and complicated artificial neural networks for NLP and computer vision. Training a model is another complex process which demands adequate computer power to accomplish the task. Data products are developed with the help of high-performance GPUs, TPUs, and distributed computing frameworks to speed up the training process. Pre-trained models whose weights are tuned are examples of transfer learning. In contrast, models trained by several but decentralized devices are categorized under federated learning, all of which respect client privacy.

Several significant steps are often employed to ensure that the AI models work well when the regular parameter examination model testing and validation process occurs. The K-Fold cross-section, scoring matrix, and measures of accuracy, precision, recall, and F1-measure are some ways to check models' performance. Similarly, SHAP (Shapley Additive Explanations) and LIME, Local Interpretable Model-agnostic Explanations are gaining significance in supporting the explanation ability and reducing bias in the model. Therefore, continuous learning is needed in the data products for the model's generalizability. These include retraining models with new data, integrating AutoML, using pipelines, and integrating model monitoring to monitor model shifts. In this way, these components can help organizations create intelligent data products for various organizations, which can solve problems in the internal environment, support decision-making, and generate enhanced business values.

Machine Learning and Artificial Intelligence data product pipeline, indicating how data passes through consumption and processing up to the model deployment and model-making process. To begin with, the edge devices provide the input data, and the latter gets collected from a messaging queue. This queue serves as a dam holding data to be processed in the data processing line, where data is cleaned, analyzed, and turned into information. It also interfaces with a cloud computing system required for performing large data analysis and training of AI models. The data is stored in Data Lake, which will be used for analysis and predictions. This data is provided to the AI/ML model training segment for creating and optimizing the models. These models rarely get retrained by employing feature-engineered outputs stored in the data processing pipeline. The process concludes with saving the model in the model repository for future use. The metadata store manages information related to the model performance, versioning key and configuration to enhance the overall management of the model.

The inference engine becomes useful for real-time outcome prediction using visualization and analytics tools. This engine retrieves models from the model repository, applies them to the incoming data, and, in turn, gives output to the user. The inference engine also extracts the data from the data lake and works on the data processed, so the prediction is made on the up-to-date information. The whole cycle affords constant evaluation of the model’s performance, assimilation of new data for performance improvement, and the facilitation of deployment. This makes creating highly adaptive and scalable AI-driven data products that provide real-time predictions possible. It also presents how these components of an AI-empowered data product relate to a data model's storage, learning, and deployment.

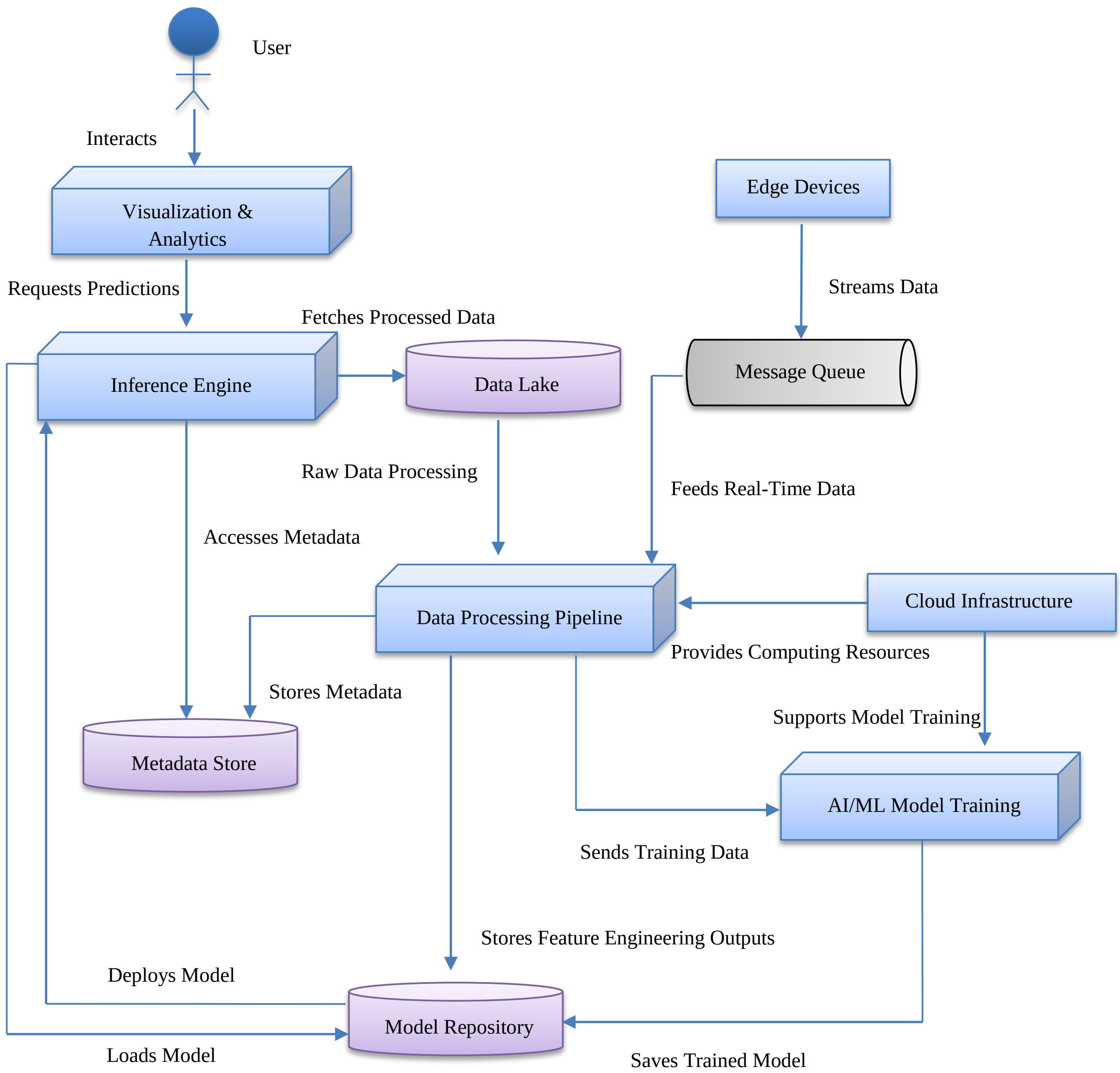

Fig 1: AI-Driven Data Processing and Model Deployment Workflow

Table 2: Comparison of AI Models for Data Products

AI Model Type	Strengths	Weaknesses	Example Applications
Machine Learning (ML)	Interpretable, requires less data	Limited in handling unstructured data	Fraud detection, recommendation engines
Deep Learning (DL)	Can process complex, unstructured data	Requires large datasets and high computation	Image recognition, NLP, autonomous systems
Reinforcement Learning (RL)	Learns from dynamic environments	Computationally expensive, slow learning	Robotics, game AI, self-driving cars

3.3. Explainability and Interpretability

As data products increase in sophistication, it is considered that they should and must also be explainable. While explainability is the ability to account for given decisions made by an AI model, interpretability comes down to identifying how inputs are related to an output. [12-14] These aspects are critical for trust building, compliance with the laws, and promoting ethics in developing AI solutions, especially for sensitive sectors, including medical, financial, and law enforcement. Models like decision trees and linear regression are part of the family of traditional machine learning models, which are highly understandable in arriving at the given results. However, models such as neural networks and large language models (LLMs) do not have a way of making their decision-making transparent. In order to tackle this issue, many approaches have been engineered to improve model interpretability. SHAP and LIME explainers are among the most popular methods widely used to analyze certain inputs' feature importance and sensitivity.

Accountability is a technical problem and a business and regulatory requirement. This was the focus of the recently enacted General Data Protection Regulation and the soon-to-be-enacted Artificial Intelligence Act, where organizations were expected to justify and audit the decisions made by AI. However, some of the important elements that should not be absent while developing data products with the help of artificial Intelligence include. Moreover, the following approaches should be considered to avoid discriminating AI results. These features can be incorporated into the models to make AI systems more transparent to the end users and help comply with the best available ethical standards.

3.4. Scalability and Performance Optimization

The elements of data products with AI integration, scalability, and performance are two significant factors that must be considered to accommodate large data volumes, growing user load, and real-time processing. Compared to other data products based on the batch processing approach, AI solutions must have updated and expansive structures suitable for always-on training inference and decision-making. Certain factors are paramount when developing any software, and one of these factors is infrastructure. The architecture organization is cloud-native and has a serverless style. It has distributed data processing frameworks such as Apache Spark, Kubernetes, and TensorFlow, which support executing AI models in multiple nodes. The two basic types of scaling include vertical scaling, where the hardware's capabilities in AI-driven data products are enhanced, and horizontal scaling, whereby more machines are incorporated to handle the increasing workloads. Furthermore, edge computing and federated learning lower the concentration of computing tasks in the cloud by processing the information nearer to the source, cutting on time, and being privacy-preserving.

Another important factor is performance optimization, especially when the application is a real-time AI application. For effective implementation of deep learning models at the edge, strategies like quantization, pruning, and distillation are used to optimize them, hence providing a solution to the computational limitations of devices like smartphones and IoT sensors. Triangulations of load balancing operations and levels of cache depth even work to bolster response time concerning AI recommendations, anomalies, and other predictions. Therefore, with advancements in AI-enabled data products, organizations need to set good scalability and computational efficiency standards to retain acceptable system response time and costs. With the help of modern AI solutions and optimal configurations, it is possible to implement high-quality data products equipped with AI and provide users with the necessary analytical data.

4. Challenges and Ethical Considerations

Implementing data products with the help of Artificial Intelligence comes with multiple potential advantages, yet it entails certain risks and concerns. Since adopting artificial Intelligence in organizations due to its vast value, organizations need to determine the Problems and solutions of Artificial Intelligence in integrating data security, data privacy, and AI bias and fairness with current frameworks. [15-18] Overcoming these hurdles can be complex. If these are not managed adequately, they can make it possible for a firm to violate certain regulations, harm its reputation, and possibly negatively impact users and stakeholders.

4.1. Data Privacy and Security

The most critical issues still pertinent to AI-driven data products include data security/privacy. Deep learning models call for huge amounts of data for training and testing, which, in most cases, may be provided from highly confidential arenas such as medical, financial, or social media. Unauthorized access, data breaches, and misuse of personal information pose significant risks to users and organizations. Geopolitical and industry-driven regulations like GDPR, CCPA, etc., provide a clear-cut code of conduct regarding data and how it is gathered, stored, and/or used. The need to implement stringent regulations means that organizations must adhere to standards such as encryption, anonymization of data, and differential privacy techniques. Secure multiparty computation (SMPC) and federated learning have become the solutions that allow AI model training without compromising sensitive data. Furthermore, safe and secure AI implementations need strict authentication mechanisms and access controls accompanied by the perpetual monitoring of the final models developed for threats such as adversarial attacks, model inversion and data poisoning.

4.2. Bias and Fairness in AI-Driven Data Products

Bias in AI models is one of the most severe emerging ethical issues which affect the decisions and results of models' actions. This is because data products of artificial Intelligence learn from past data and, therefore, adopt the qualities and characteristics of the training data. They also may be of racial, gender, or socioeconomic nature, to name but a few, and lead to a reinforcement of the bias in the model and, therefore, in the predictions made throughout the system. To prevent bias, the data used should be diverse, the algorithms used should also be fairly developed, and there should be regular model reviewing.

There are ways to rectify this, such as using adversarial debiasing, reweighing training samples, and using explainability tools such as SHAP and LIME. Also, data scientist-ethicist collaboration with other domain specialists is required to have more ethical procedures and regulate AI decision-making. Market players and regulatory authorities are escalating their call for fair and/or transparent AI requisition. Businesses need to implement ethical AI strategies, achieve the identification and assessment of the potential negative effects of AI bias, and guarantee that emerging AI data solutions meet social and legal requirements. The bias should be addressed to increase trust and accountability within the industry and to increase the acceptance and efficiency of artificial intelligence.

4.3. Integration with Existing Systems

Integrating AI data products with the organization's information systems brings other technical and operational difficulties. It is worth noting that most organizations use networks that cannot effectively accommodate AI and machine learning processes. They might have a very rigid architecture, stored data are kept in separate compartments, and the technology it uses is quite old, which does not easily support AI integration. Such problems could be solved by changing the existing data environment through implementing the cloud-native, API-driven, and microservices approaches. Integrating Data Lake and real-time data ingestion and using an AI orchestration framework will go far in creating the missing link between traditional programming techniques and AI.

Moreover, the AI model should be further designed to maintain compatibility with on-premise databases, Enterprise Resource Planning (ERP) systems, and business intelligence tools. Managing change is one of the key issues that arise during the integration process. People may have some form of fear or reject the use of AI in organizations due to issues such as redundancy, difficulty, or comprehension. Indeed, to avoid these negative consequences, organizations must endeavour to enhance AI literacy, offer sufficient information on the roles of AI in organizations and encourage human conformity with AI systems. There is the need to understand that the transition to data products based on AI also requires organizational change management approaches that provide a connection between AI adoption and business and user expectations.

5. Case Study: Media Stream – Content Recommendation

5.1. Task or Conflict

Consequently, there were emerging issues in delivering highly targeted media streams to Media Stream's clients. The aforementioned methods, like collaborative filtering and rule-based mechanisms, failed to work as proactive recommendation solutions when the user base grew. [19-21] Most of them were providing repetitive information, incapable of altering over the shift of preferences, and unable to detect changed behavioural trends. Hence, most users got annoyed with repetitive suggestions and, thus, watched or listened to content less often and finally cancelled subscriptions. The company's growth in the fast-growing streaming market required a better recommendation system to meet customers' needs and deliver personalized content from Media Stream.

5.2. Solution

Media Stream had to embrace a deep learning approach to content recommendation to counter these challenges. The company created a neural network model that is adjusted to pertinent data on the user's interactions, such as the videos the user has watched, the search queries, the duration of the videos the user has watched, the ratings given, and any other related preference. With the help of architectures like RNN and transformer models, the system could identify the specific patterns the users go through and predict their tendency towards certain content.

5.3. Overall Impact

The introduction of deep learning in Media Stream's recommender system proved beneficial as it improved the engagement of users and the business outcomes. Key outcomes included:

- With a 25% rise in viewers' retention, as a result of making recommendations more specific, the users were inspired to watch more genres of videos.

- An increase in renewal rates is due to increased customer satisfaction with the suggestion of content to be subscribed to.
- The content consumed increased as more users came across relevant content, ranging from series to movies they preferred.
- This feature reduced the content fatigue rate since the model came up with different articles and subjects to offer rich content variation.
- Through the effective use of deep learning to provide recommendations for content, Media Stream was able to set itself apart from stiff competition in the streaming industry while promoting the overall satisfaction of both the users and the business.

This paper provides the following lessons about data products in organizations with a focus on artificial Intelligence products, as seen in the Media Stream case:

- Deep learning allows for an individualized content recommendation – generic recommendation methods do not always consider users' behavior and tendencies; therefore, AI systems can propose the most appealing option.
- Constant training with real-time data is needed – customers' preferences change with time, so the AI models should also be updated live. The real-time approach increases the dynamism since there is always an opportunity to make the required tweaks in the form of recommendable content on the platform to keep the audience engaged.
- Scalability and computational efficiency are important – simple high-volume recommendations of millions of users would need the hardware and software solution to work in the cloud, take advantage of distributed computing, and employ all sorts of techniques to scale the AI model.
- The findings suggest that diversity in recommendations brings about higher user satisfaction. Thus, an AI system should always provide diverse but not repetitive recommendations to ensure that the user can also search for some other content.

6. Future Directions and Research Opportunities

Given that the trends of refining and commercialization of AI data products are gradually emerging, new issues and opportunities are appearing in various fields. Over the years, more developments in Artificial Intelligence, deep learning, and data science will enhance the development of improved, larger, and ethical data products. Future research in this area would be concerned with increasing the speed of models, model interpretability, model security, and applying AI to areas other than the traditional ones.

6.1. Advanced Personalization and Context-Aware AI

The most important trend concerning the overall development of AI in the future is personalization, which utilizes context-sensitive AI-driven decisions. Today, based on usage history, data services make recommendations about the user's preferences, but tomorrow, there will be contextual recommendations based on the user's current location, emotions, environment, and social circle. For example, streaming services can change the playback list depending on the user's activity (for example, offering cheerful music for exercising or anime for resting). Incorporating textual, auditory, visual, and sensor data in multi-modal AI will add depth to the extent of understanding the users' needs.

6.2. Federated Learning and Privacy-Preserving AI

There are growing concerns about data privacy and regulations, and thus, there tends to be a focus in research on federated learning and privacy-preserving Artificial Intelligence. It allows an update of the AI algorithm to be carried out across multiple and different devices without having to exchange the original data, hence improving their privacy while at the same time improving the model's performance. This is especially useful in healthcare and finance, where the privacy of information is of utmost importance. Techniques that would still be relevant and critical include differential privacy, homomorphic encryption, and secure multiparty computation to ensure that the resultant AI-driven data solutions and products strictly adhere to aspects such as GDPR and CCPA.

6.3. Explainable AI (XAI) and Ethical AI Frameworks

As the complexity of AI models increases, there is also the demand for the ability to understand the model's logic and at which stages it might be wrong. The next steps include increasing the work towards using clear and understandable methods of functioning AI that could be explained. Some techniques, such as counterfactual reasoning, saliency maps, and causal inference models, will be useful in explaining its effects to users and regulators. Also, AI ethics research will go on to improve the fairness auditing techniques, bias discovery and mitigation instruments, and rules and regulations governing mechanisms for AI data products to be fully transparent, responsible, and impartial.

6.4. Real-Time AI and Edge Computing

The future of AI-enabled data products is real-time AI, wherein AI aspects are cast off edge devices instead of centralized computing. Today, most AI algorithms are implemented with the help of cloud computing, which has some delay and requires a network connection. It is just a short step forward to put the described artificial intelligence models on edge nodes or in individuals' mobile devices, sensors, vehicles, etc. Portable or low-power weight deep learning models, new quantization techniques and other AI accelerators such as neuromorphic chips will make data products from AI even faster and more energy efficient, which could unlock more possibilities for more real-time solutions.

7. Conclusion

The growth in sophistication of artificial intelligence, which is accompanied by deep learning, has shifted data products from traditional, rule-based structured instruments into intelligent systems that can evolve in their function and operation. With the help of AI-enabled data products, organizations and businesses can access big data artificial intelligence models that refine, optimize, and personalize the data and use it as a tool for decision-making and automation. However, they also bring certain issues, such as Privacy, Controlling Bias, Compatibility and Scalability, and High Velocity and High Throughput. To solve such issues, there is a need to develop a clear approach that focuses on explainability, the AI ethical practice approach, and proper security measures. In the future, there will be advancements built on concepts such as federated learning, real-time AI, cross-domain, and incorporation of quantum programs. Organizations must implement themselves in these advancements while being open, non-lying, and legal. Through research and development of sound ethical practices and building up of agent structural infrastructure that can be scaled up, then it is possible for businesses to manage to actualize the raw potentials in the use of data products by way of AI to deliver better, efficient, user-friendly solutions to technology solutions that would enable the growth of other related industries.

References

[1] Blasch, E., Pham, T., Chong, C. Y., Koch, W., Leung, H., Braines, D., & Abdelzaher, T. (2021). Machine learning/artificial intelligence for sensor data fusion–opportunities and challenges. IEEE Aerospace and Electronic Systems Magazine, 36(7), 80-93.

[2] Lu, Y. (2019). Artificial Intelligence: a survey on evolution, models, applications and future trends. Journal of Management Analytics, 6(1), 1-29.

[3] Bugembe, M. (2022). *Data As A Product, Redefining Our Approach To Producing Value From Data. Forbes.* Showcases a shift from treating data as a defensive project asset to treating it as an accessible, intuitive product that drives adoption and business value.

[4] Top 20 Deep Learning Case Studies [Detailed Analysis], digitaldefynd, online. https://digitaldefynd.com/IQ/deep-learning-case-studies/

[5] Woodie, A. (2023). *In Search of Trustworthy Data Products. Datanami.* Traces the concept back to DJ Patil (2012) and tracks its evolution through data mesh and the differentiation between data products and analytics/AI products; also emphasizes the emerging role of the data product manager.

[6] "Data products, data mesh, and data fabric." (2024, but reflecting earlier research). *Business & Information Systems Engineering.* Discusses how AI (especially generative AI) is starting to change the nature of data productsincluding development processes, organizational models, and lifecycle management.

[7] Yan, F., Ruwase, O., He, Y., & Chilimbi, T. (2015, August). Performance modelling and scalability optimization of distributed deep learning systems. In Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1355-1364).

[8] 10 Examples of Deep Learning Applications, Coursera, online. https://www.coursera.org/articles/deep-learning-applications

[9] Benson, S. J., McInnes, L. C., & Moré, J. J. (2001). A case study in the performance and scalability of optimization algorithms. ACM Transactions on Mathematical Software (TOMS), 27(3), 361-376.

[10] Cecchet, E., Marguerite, J., & Zwaenepoel, W. (2002). Performance and scalability of EJB applications. ACM Sigplan Notices, 37(11), 246-261.

[11] "Data as a Product: An AI-Readiness Strategy." (Grid Dynamics blog, date implied 2022 or earlier). Explores how AI can enhance DataOpsthrough AI-powered metadata/catalog management, auto-quality checks, anomaly detection, orchestration, and access controlto elevate data products.

[12] Mantri, M. (2023). *Do you know 5 types of Data and AI products? Medium.* Identifies categories of data productsfraud detection tools, customer segmentation, real-time analytics platforms, AI-driven assistants, etc.highlighting how AI/ML powers different data product types.

[13] "What is Data Product: Definition & Examples." *GeeksforGeeks.* Defines various data product archetypesrecommendation systems, predictive analytics tools, dashboards, AI assistantsand explains how they embody evolving data product forms driven by ML/AI.

[14] Cui, Z., Gao, T., Talamadupula, K., & Ji, Q. (2022). *Knowledge-Augmented Deep Learning and Its Applications: A Survey*. While not specifically about "data products," this survey introduces knowledge-augmented deep learning methods that can enrich AI-driven data products with domain knowledge, enhancing generalization and interpretability
[15] Ding, J., Li, X., Kang, X., & Gudivada, V. N. (2019). A case study of the augmentation and evaluation of training data for deep learning. Journal of Data and Information Quality (JDIQ), 11(4), 1-22.
[16] Lu, Y., Shen, M., Wang, H., Wang, X., van Rechem, C., Fu, T., & Wei, W. (2023). *Machine Learning for Synthetic Data Generation: A Review*. Although published in early 2023, it reviews techniquesprior to 2023for synthetic data generation, enabling data-product designers to supply AI and ML models with robust, privacy-preserving datasets.
[17] Kim, Y., Lee, H. J., & Shim, J. (2021). Developing data-conscious deep learning models for product classification. Applied Sciences, 11(12), 5694.
[18] 15 Most Common Deep Learning Use Cases across Industries, softengi, online. https://softengi.com/blog/15-most-common-deep-learning-use-cases-across-industries/
[19] Cheng, Y., Liu, Y., Chen, T., & Yang, Q. (2020). Federated learning for privacy-preserving AI. Communications of the ACM, 63(12), 33-36.
[20] Dehghani, Z. (2019). *How to Move Beyond a Monolithic Data Lake to a Distributed Data Mesh.* ThoughtWorks (blog/whitepaper). Introduced the data mesh paradigm and the concept of autonomous data products domain-aligned, self-managing data artifacts embodying governance, discoverability, and access principles suitable for AI environments.
[21] Sodhro, A. H., Pirbhulal, S., & De Albuquerque, V. H. C. (2019). Artificial intelligence-driven mechanism for edge computing-based industrial applications. IEEE Transactions on Industrial Informatics, 15(7), 4235-4243.