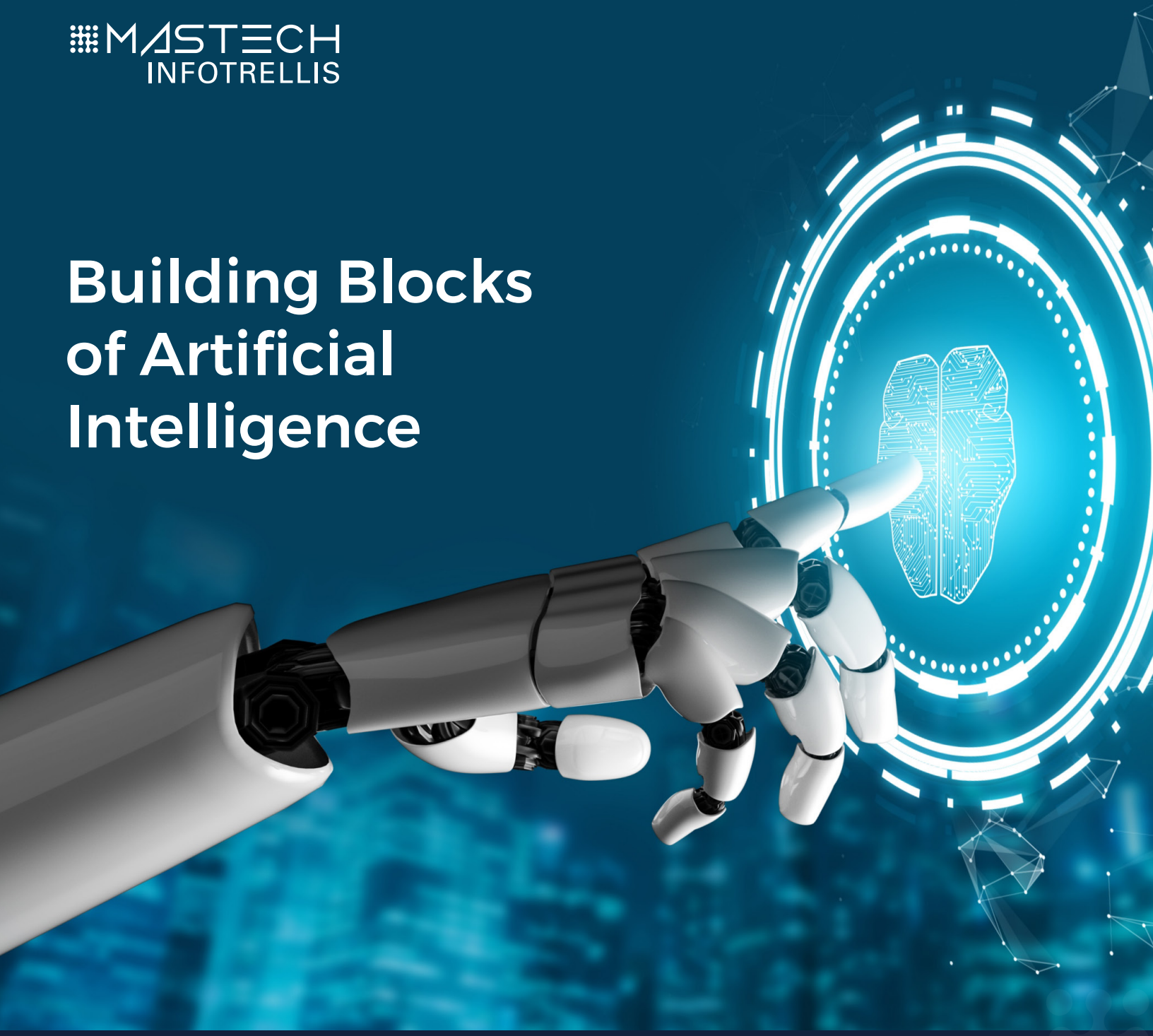
MASTECH INFOTRELLIS

Building Blocks of Artificial Intelligence



Artificial Intelligence describes the ability of machines or software to demonstrate reasoning and decisioning capabilities like humans or animals. It is based on the concept that a machine or software can make reasoned decisions based on events and information at its disposal. Artificial "reasoning" uses computer models combined with information to make decisions.

Artificial Intelligence needs data and information to make decisions and train the underlying reasoning models. The more data it can access, the higher the data quality, the better its models and decisions will be. Artificial Intelligence is a data hog.

Data is the Backbone of Any AI Initiative

An effective AI strategy comprises various components such as infrastructure, technology stack, data and architectural strategy, and organizational changes. Among these, the data strategy plays a crucial role in driving the effective adoption of AI. It's the data that makes AI work.

A meticulously crafted data strategy is the cornerstone for successfully implementing AI. Whether streamlining operations, enhancing customer experiences, or making strategic decisions, the AI strategy is directly linked to the data strategy.

The starting point for implementing any Artificial Intelligence capabilities within a broader IT landscape is the development of the underlying models and machine learning algorithms. The development of these models is critical and requires a well-organized and curated data landscape. Philip Russom of TDWI emphasizes this, stating, "The development of machine learning algorithms depends on large volumes of data, from which the learning process draws many entities, relationships, and clusters. To broaden and enrich the correlations made by the algorithm, machine learning needs data from diverse sources, in diverse formats, about diverse business processes."

AI algorithms derive insights and recognize patterns from data, underscoring the importance of data quality, quantity, and accessibility. Consequently, the triumph of any AI initiative is contingent upon an enterprise's capability to seamlessly access, process, and analyze data on a large scale, navigating through the entire data lifecycle — from initial collection and storage to subsequent data engineering and integration.

Navigating Data Challenges for Successful AI Implementation

The efficacy of AI systems hinges significantly on the quality and precision of the data employed. When training an AI model, if the data is flawed due to inaccuracies, incompleteness, inconsistencies, or biases, the model's predictions and decisions will inevitably mirror these deficiencies and potentially create AI "hallucinations".

Conversely, high-quality data serves as the bedrock for AI systems to generate more precise predictions, offer relevant recommendations, and streamline processes effectively. Simply put, superior input yields superior output. For instance, providing clean sales information tied to high-quality 360 customer data will provide a much better base for developing accurate customer profiles and determining the next-best-action recommendations than relying on disparate data sources with varying degrees of quality.

Common data quality challenges enterprises encounter include inconsistency, inaccuracy, incompleteness, and duplication. Enterprises also grapple with issues like outdated information, irrelevancy, poor data governance, lack of standardization, compromised data integrity, and concerns about data security.

Resolving these issues requires a multi-faceted approach, combining data management solutions with organizational changes. For instance, enhancing data quality for optimal AI outcomes involves employing data-cleansing techniques to eliminate errors, duplicates, and inconsistencies. Ensuring data completeness and addressing missing values is crucial to preventing model skew.

Furthermore, poor data quality may also result in biased AI decisions, often originating from biased training data. Biases present in the training data can be replicated and potentially amplified by the AI system, posing significant regulatory risks in areas such as lending and law enforcement. Using representative and unbiased data is key to mitigating such risks.

Careful consideration should be given to the feature engineering phase of AI systems. This would include examples like avoiding the direct inclusion of protected attributes, such as gender, age, ethnicity, etc., in the training process of ML models. Recognizing the human aspect of data and addressing privacy concerns while understanding historical biases are essential steps in any AI development process. Employing diverse perspectives in dataset creation helps prevent the propagation of societal biases.

Leveraging methods like data augmentation enhances data diversity and representativeness, thus helping to reduce these biases. Techniques like statistical parity tests help to ensure that model decisions remain free from unfair influences tied to these protected features. Additionally, robust data governance strategies, including transparent data collection, storage, and use procedures, contribute to long-term data quality maintenance and reliability.

Maximizing AI Potential: The Role of Metadata

Equally important as providing large quantities of high-quality data is capturing and maintaining metadata for your enterprise. Metadata is pivotal in ensuring data quality and accuracy, which is crucial for training AI models. Metadata can be categorized into three groups – structural, administrative, and descriptive metadata. Well-curated metadata, including information about data sources, formats, and timestamps, aids in validating and verifying data integrity.

This minimizes errors and biases that could adversely affect AI-driven insights. Additionally, metadata provides context about data sources, collection methods, and relationships, facilitating effective model development by aiding in data understanding and pattern identification, making the data more discoverable and functional overall. Further, metadata empowers AI systems to learn and evolve over time. Detailed metadata facilitates tracking changes in content, which enables AI algorithms to adapt and improve over time.

Enriching data with descriptive metadata empowers AI systems to swiftly and accurately process, categorize, and analyze information. This, in turn, enhances the capabilities of AI-driven applications and facilitates AI model interpretability.

It helps data scientists gather, work with, and utilize enterprise data effectively. While some of this data can be gathered manually, an active tool-based metadata management approach will assist current AI efforts and pay future dividends in a world where AI models and algorithms will be increasingly used to identify and prepare the data needed for AI-driven systems and to assist with the development of new models and algorithms.

End-to-end data visibility, quality, and metadata management programs are essential to optimizing AI development and results efforts. A holistic data strategy will also provide better results than infusing data quality as a last-minute data science ingestion cleansing effort, which will be less effective than cleaning at the enterprise data or operational process level.

Governance also plays a vital role in collecting and managing data used in AI. Beyond having quality metadata that describes the actual data being consumed, having lineage and tracking information on the use of data in AI initiatives is imperative for broader enterprise AI Governance processes and operations. An accurate picture of what data is being used, how that data was selected, and the overall processes for sourcing, use, development, and deployment of AI-enabled solutions is critical to broader enterprise risk management functions.

Operationalizing AI

Beyond generating AI models, companies aiming to harness AI-driven insights must operationalize AI finding and automated decisioning capabilities. A well-structured data and application architecture employing AI service interface concepts allows companies to develop and leverage their AI work for better business outcomes.

Increasingly, a company's ability to generate compelling, efficient AI-enabled systems will depend on its data engineering, management, and governance capabilities. Sourcing, managing, governing, and storing data are essential building blocks in any company's AI journey. Data scientists and business intelligence analysts will still be critical, but the goal needs to be to enable self-sufficiency in everyday data users.

Long-term successful adoption of AI will depend on an enterprise's ability to grow a cultural appetite for data and AI, build a well-structured data and application architecture, and foster a willingness to invest in and deploy new data management and AI technologies.

Unlocking Data Excellence with Mastech: Real-world Stories

As a leading Data and Analytics company, Mastech empowers enterprises to maximize their data's value through a comprehensive suite of services tailored to meet evolving business needs. With a holistic data management and analytics approach, Mastech empowers organizations to leverage their data assets effectively, uncover valuable insights, streamline operations, and make informed decisions.

Whether it's modernizing Master Data Management (MDM) solutions, enhancing data quality, or implementing advanced analytics techniques, Mastech is committed to helping clients harness the full potential of their data to achieve their strategic objectives in today's competitive landscape.

Shipping Optimization for a North American Manufacturing Firm by Unifying Data from Nine Different Distributed Systems

Mastech InfoTrellis revolutionized a North American manufacturing company's shipping operations, leveraging advanced analytics solutions for efficient industrial product distribution via multiple channels.

We established a robust data collection, storage, and management foundation, ensuring seamless integration with future AI and analytics needs. Our solution optimized shipping processes, improving the company's logistics capabilities for enhanced performance and strategic decision-making.

Highlights

01 Unified data from 9 sources for the first time ever, enabling the implementation of a single enterprise shipping process

02 Leveraged shipment data, advanced analytics, and AI to accurately predict the company's ability to achieve its annual performance target of 17 million tons of shipments, representing a 6% growth

The Challenge

- The client struggled with scattered data across 9 systems, hindering strategic decisions and demanding centralized solutions for seamless insights
- Varied formats in legacy systems delayed data consistency, necessitating a standardized approach for improved analytics
- The absence of unified shipping data collection tools posed a challenge, requiring streamlined processes for enhanced efficiency

The Solution

- Implemented Azure Medallion architecture for efficient shipping data management
- Built an automated data pipeline using Azure ADF for legacy shipping data sources, including custom solutions for older legacy systems
- We streamlined raw data processing using parquet-based formats for increased efficiency
- Successfully established Synapse-based storage layer for reliable management of shipping domain data

The Outcome

- Integrated shipping data for cost-efficient management and immediate routing and carrier selection savings
- Successfully tested and implemented an AI/ML analytic solution for shipping optimization recommendations
- Established Azure Medallion Lakehouse, leveraging advanced analytics to predict annualized performance accurately

Leveraging AI and ML to Understand Customer Behavior in Banking

Mastech InfoTrellis spearheaded a transformative shift in banking operations for our client, revealing nuances in customer relationships between small businesses and retail clients. We streamlined this information's storage, management, and distribution by implementing an innovative Cloud-based Data Lake.

This optimization enhanced customer-facing activities by accurately identifying new sales opportunities that powered the bank's strategic growth initiatives.

The Challenge

- Hindered by the absence of C2C relationship tracking, the bank struggled to maintain a comprehensive view of client interactions
- Small business customers needed more coverage of B2B and B2C relationships from traditional external vendors (e.g., D&B)
- Incomplete customer information impeded integral services and obstructed the identification of valuable cross-sell and up-sell opportunities

The Solution

- Integrated Graph DB for comprehensive customer data from MDM and secondary sources
- Developed AI/ML algorithms to unveil non-obvious B2B and B2C customer relationships
- Enhanced Data as a Service architecture, ensuring storage, distribution, and ongoing synchronization of relationship data
- Successfully established Synapse-based storage layer for reliable management of shipping domain data

Highlights

01 An improved 360-degree customer view enhanced the customer-facing support processes

02 Amplified new sales opportunities by leveraging new relationship data that was unearthed

The Outcome

- A continuous emphasis on AI/ML in our approach uncovered non-obvious customer relationships and seamlessly delivered results to the larger enterprise for enhanced process and decision-making support
- Our solution optimized customer-facing activities and identified new sales opportunities with more insightful information on customer relationships
- Customer relationship information was delivered to branches in real time, boosting overall efficiency

112% Revenue Growth for a Tech Company with Smart FMP Implementation

Mastech InfoTrellis helped the federal marketing team of a tech company with an innovative solution to enhance campaign efficacy and audience creation.

Our expertise facilitated seamless data ingestion from the Customer Engagement Platform (CEP) into the Federal Marketing Platform (FMP). This implementation optimized processes, improved data handling capabilities, and streamlined operations.

Highlights

01 Exceeded revenue goals with an outstanding 112% attainment, growing +4% YoY, surpassing the plan by $225 million

02 Innovatively crafted 16 new audience segments through Adobe for impactful federal marketing campaigns

03 Achieved a stellar 99% email deliverability rate, surpassing the business goal of 97% for enhanced campaign effectiveness

The Challenge

- This company's federal marketing faced challenges with manual segmentation using excel, which caused delays in sending data to SFMC for email deployment
- The marketing teams lacked immediate access to data, hindering their ability to experiment and create effective segmentation, affecting the campaign performance
- Manual file creation for segmentation caused scalability and agility issues, especially when coordinating with SFMC for large-scale email deployment identification of valuable cross-sell and up-sell opportunities

The Solution

- Established FMP's Consumption Layer, optimizing Adobe Campaign for efficient campaign and audience creation
- Enabled user-friendly management within Adobe Campaign, facilitating audience file delivery to SFMC for email distribution
- Streamlined workflow for tracking email delivery and interaction data, seamlessly integrated into the Consumption Layer
- Developed Power BI dashboards for comprehensive tracking of campaign performance

The Outcome

- Improved engagement metrics with enhanced click-to-open rates and reduced unsubscribe rates
- Streamlined campaign execution by replacing manual processes with efficient Adobe Campaign deployment
- Tailored campaigns by implementing custom federal schemas in Adobe, that aligned seamlessly with the Consumption Layer

Authors

Michael Ashwell

Michael is a seasoned professional with over 35 years of experience in enterprise architecture, solution developmer cloud offerings, global sales, and consulting. He spent 30+ years at IBM where he held various roles, including leading the Data and Analytics Lab Services Cloud COE, and developed several key offerings. Michael has worked internationally across multiple industries and holds a BSBA degree in Information Systems from Xavier University. In his free time, he enjoys home theater, making wood-fired pizza, and driving his restored 1962 TR3.

Deepti Soni

Deepti Soni, a seasoned data scientist with over a decade of experience, leverages her expertise to design and implement impactful solutions for complex business challenges. Her forte lies in architecting innovative knowledge graphs and AI/ML-based approaches, empowering clients to achieve rapid ROI from their data-driven initiatives. Deepti holds a Ph.D. in Machine Learning-based High-Dimensional Data Analysis from IIT Roorkee and a Master's in Computer Engineering from Delhi College of Engineering.

Let's Get in Touch

+1 412.787.2100

Pittsburgh, PA

experience@mastechinfotrellis.com