M+E JOURNAL

AUTOMATION INTEGRATION

The media and entertainment industry implementing artificial intelligence throughout the global entertainment supply chain.

It's not just artificial intelligence but a real tool for complete automation and integration

SMART CONTENT
Speech-to-text solutions, translation tools, live captioning and dubbing … AI localization applications are broad, with proven efficiencies

SECURITY & WORKFLOW
Server blocking, zero-trust cybersecurity strategies, future-proofing contracts and rights management systems, all are seeing advances

24.01

AI IS CREATING NEW PARADIGMS

ABSTRACT: The value of data has risen proportionately as the media industry looks to artificial intelligence (AI) to streamline workflows, improve operational efficiency, and subsequently reduce costs. However, AI depends on data even as it creates new data in unprecedented volumes. As a result, data integrity must be the new imperative.

Data integrity is the new imperative

By Peggy Dau, Marketing Director, MetaBroadcast

Since the dawn of technology, persistent change has been the norm. The CEO of a Silicon Valley IT company once said that if you are uncomfortable with change, you are in the wrong business. Technology has improved productivity, resulting in greater efficiency. It has also disrupted existing processes, giving businesses and consumers new alternatives for creating and receiving value.

The latest technology to upend and reframe business methodologies is artificial intelligence (AI). It's not new. Consumers have been exposed to AI concepts for years through autocorrect, mapping services, and facial recognition. Businesses have adopted low-level AI in their use of automation. So, why is AI suddenly attracting the attention of industry leaders? Well, the underlying technology has gotten smarter and faster. Chips, or AI accelerators, are specifically designed to execute AI workloads efficiently. They can significantly improve the performance of AI algorithms compared to general-purpose CPUs or GPUs. This performance improvement allows faster training and inference times, enabling real-time or near-real-time AI applications.

With the proliferation of data from various sources, including sensors, IoT de-

TO BE 'SUPERINTELLIGENT' AI MUST FIRST HAVE HUMAN INTELLIGENCE.
And since human intelligence requires self-awareness, then the understanding that there is a 'self' is crucial.

vices, social media, and digital platforms, advanced AI algorithms are necessary to process, analyze, and derive insights from large datasets efficiently. AI is now more accessible to anyone at home or in the office through various generative AI platforms. So, with this heightened awareness, it's time to clarify the importance of the relationship between data and AI.

DATA DRIVERS

The rise of data-driven decision-making not only influences the adoption of AI but also reinforces the importance of data itself. The DPP 2024 Predictions highlighted the importance of media organizations defining their data strategies. Throughout the media supply chain, data is critical to the effectiveness of various integrations and workflows. At the same time, high volumes of data are created at every stage, from creation to consumption.

The demand for cost reduction, streamlined workflows, and operational efficiency are the primary drivers of AI adoption across the media and entertainment industry. According to Grandview Research, the market value of AI in this sector is projected to reach $124.48 billion by 2028, a compound annual growth rate (CAGR) of 31.89 percent.

The application of AI will not only create efficiencies but also more data. The media and entertainment industry must ensure that the underlying data used to train AI engines is complete, validated, and unbiased. If the quality of the underlying data is in question, then the output is also problematic.

THE IMPORTANCE OF CLEAN DATA

With the rising use of generative AI (GenAI), media organizations must ensure that the data underlying GenAI platforms integrated with media workflows is accurate, comprehensive, and complete. GenAI incorporates algorithms and data models to create new data. If errors exist in the data used in these models, the outcome will magnify those flaws. On the other hand, clean data ensures that the generative model learns meaningful patterns and produces accurate and relevant outputs.

Generative AI has many applications, including image generation, text generation, music generation, and more. In each application, clean and high-quality data is essential for training accurate and realistic generative models. For example, in image generation, if the input images are blurry or contain artefacts, the generated images may also suffer from the same issues.

Preparing data for use in a GenAI model requires ensuring the data is in a consistent format and data schema. Understanding the source, provenance, and

Continued on page 106

Peggy Dau, *currently the marketing director at MetaBroadcast, has been connecting the dots for tech companies enabling the media supply chain for 20-plus years. Her experience spans roles and projects with companies such as Hewlett-Packard Enterprise, Amino Communications, Avid, Grass Valley, Quantum Corporation and Microsoft. peggy@metabroadcast.com*

METABROADCAST *Continued from page 58*

copyright of the original data and any updates to that data is an element that is often overlooked. As GenAI advances, tracking data lineage will increase in importance. Upon validation of data formats, schemas, accuracy, completeness, and consistency, the data is ready for normalization. This process ensures that data is suitable for use in a GenAI, machine learning, or more advanced AI training environment.

AI ENHANCING METADATA MANAGEMENT

When it comes to monetizing content catalogs, AI algorithms will analyze user behavior, consumption patterns, and preferences, enabling them to deliver personalized content recommendations. By analyzing the descriptive and collaborative data associated with content, content recommendation engines can strengthen personalization beyond genre or cast members. Using a broader data set when training AI algorithms, will provide greater relevance, thereby improving user engagement and satisfaction.

Another area ripe for AI is that of image generation. Images play a significant role in enticing consumers to watch a series or movie. Audiences with different backgrounds or cultural beliefs are attracted to different types of imagery. Generating a wider range of images is appealing to many video service providers. However, using AI-generated images raises legal and ethical concerns, particularly regarding copyright and ownership. Video service providers must navigate these issues carefully to ensure they have the rights to use and distribute AI-generated content legally.

There are several ways in which a pragmatic approach to AI can add value to metadata management. Starting with quality control, AI models can identify inconsistencies or errors in metadata. In learning patterns such as date or text formats, AI can detect anomalies and flag data fields that are likely to be incorrect. Then, GenAI can make recommendations to improve data quality. AI can also identify opportunities for metadata enrichment by assessing existing descriptive metadata and suggesting relevant keywords, tags or captions.

As AI is trained to understand the content and context of data fields and content records, many use cases may be defined where it can standardize and accelerate outcomes. For example, as video service providers aggregate data from different sources, machine learning can streamline the matching and linking of metadata across different datasets or sources. Another critical use case for broadcasters and streaming services is that of genre classification. Clustering algorithms help identify common characteristics within each genre and differentiate between different genres, even in cases where traditional genre definitions are ambiguous or overlapping. It is in these types of narrow use cases that AI can provide measurable value. However, this is only possible with a foundation of clean data.

MetaBroadcast has been consolidating and cleansing data for over ten years. It is a critical prerequisite to the automated equivalence process enabled by Atlas, our metadata management platform. Our metadata repository of over 140 million master MBIDs and their associated content records reflects millions of data fields ingested, cleansed, and equipped from many sources (e.g., ITV, Gracenote, IMDb, Press Association, Wikipedia, broadcaster CMS, etc.). The records have been persistently updated as existing data changed or new data became available, giving MetaBroadcast enhanced capabilities to match content records successfully and deliver clean, consolidated metadata to our customers.

There is no argument that the data to train all AI models is present and available through the media supply chain. Yet, as all forms of AI are implemented across the media and entertainment industry, we must acknowledge that clean, validated, and verified data is imperative. ⊞

INDEE *Continued from page 72*

behavior. Machine learning (ML) algorithms can discern such patterns easily and alert the project owners of any unexpected changes in behavioral patterns. This forms the basis of our anomaly detection system - an intelligence layer that monitors user actions for significant change in patterns, raising alerts as necessary.

We see multiple use cases for such an anomaly detection system:

- ***Viewers:*** Changes in viewing behavior, including device switches, rapid IP address changes, altered viewing times, and content consumption speed, will be flagged.
- ***Administrators:*** The number of screeners being issued in a week, roles and their actions, number of videos uploaded, and so on can be monitored as well.

With an ML algorithm, we envision a greater understanding of customer behavior over time as well as providing a strong backbone for anomaly detection.

AN API-ENABLED BUSINESS

As part of our roadmap for 2024, we are looking at APIs enabling every stage of our system. The goal is to provide a seamless way by which customers can replace a particular stage with a custom solution and leverage existing applications and solutions within the Indee platform for managing their workflow.

Indee has been a high-touch business where our success has come from our customer success team. At the same time, an API model predicates that the customers build their own solutions and run and operate their solutions. While that may seem antithetical to one another, we see that dichotomy as being crucial to providing exceptional customer service to both business and technology teams.

We foresee customers picking a part of a stage or possibly one stage and building their own solutions, such as say ingestion and still leveraging the rest of Indee and our managed service offerings to complete their screening needs. Our aim for the API-enabled business deepens our integration with customer systems and enables customers to build what they want faster. ⊞