

The Future is Automated

The landscape ahead for M+E looks uncertain. Automation can ease the journey.

WORKFLOWS AND THE CLOUD

From cloud-based productions to unprecedented storage demands, the supply chain has never been more challenging

SMART CONTENT

The questions around AI, metadata, and analytics are endless. But we do have many of the answers already

SECURITY SOLUTIONS:

Cybersecurity and content protection in M&E is always on defense. Can automation can change that?

23,01

GENERATIVE AI MEETS MULTI-MODAL AI FOR VIDEO-BASED CONTENT CLASSIFICATION AT SCALE

The remarkable benefits of generative AI have yet to be fully explored



ABSTRACT: Today, humans are central in screening video content for localization. Spherex AI not only automates the detection, interpretation, and classification of objectionable content, but it also processes videos at scale. This article discusses how Spherex turned a cultural play-book into a regulatory platform, opening borders for content.

By Pranav Joshi, Director of Product Management, AI, and Todd Landfried, Corporate Communications Manager, Spherex

Text-based indexing peaked in the late 1990s and early 2000s, and companies leveraged keyword optimization to promote their brands, products, and services. In 2007, the iPhone's launch enabled developers to reach new audiences with mobile apps. Shortly after that, cloud computing began with Amazon Web Services (AWS), Google Cloud, and Microsoft's Azure, leading to new accessibility of information/data anytime and anywhere. As a pattern, we have seen disruptive technologies that move first and fast survive to headline most of the innovation that happens in the ensuing 5 to 10 years. This decade — the 2020s — is all about artificial intelligence (AI).

AI METAMORPHOSIS

Although AI has existed for over 50 years, most recognize it in services recommending shows to watch, products to buy, routes to take, and more. These types of AI-based systems are a reactive form of AI and are capable of ingesting substantial amounts of data and using that combined knowledge to perform tasks at a large scale. Multi-modal AI is a paradigm shift in which image, text, speech, and audio components are combined with multiple deep learning algorithms to solve real-world problems (the world we live in is multi-modal). Examples include content classification and language translations.

THE INDUSTRY IS AT AN INFLECTION POINT on the multi-modal AI and generative AI timeline, where most software with human-computer interaction (HCI) will see considerable augmentation of these two innovative capabilities.

On the other hand, generative AI is a machine's ability to create or edit text and images with minimal human input. The new frontier of generative AI is going mainstream. Well-known examples like ChatGPT, DALL-E, Scalability.AI, and similar technologies showcase these new generative capabilities. What changed for generative AI is that open-source alternatives to the proprietary models launched in quick succession towards the end of 2022. For instance, Eleuther.ai's GPT-NeoX-20B competes with OpenAI's GPT-3 for text generation, and Stability AI's Stable Diffusion competes with OpenAI's DALL-E2 for generating images and video. In all these innovations, one thing remains constant. There's an intense focus on creating platforms that serve as the foundation for industries to develop applications, thereby creating new ecosystems and economies.

Generative AI is a remarkable new technology, and it's easy to see why people get excited when considering its potential benefits. Media entities in the value chain imagine many advantages of its use. Content creators could, for example, use it to help get past writer's block, create plot twists, compose musical scores, or assist with post-production tasks. Marketing professionals could leverage text-to-image tools to create poster artwork or generate trailers for any market and in any language. Distribution groups could employ AI to create overlays on content that aid promotions and advertising-related campaigns for local audience targeting.

The industry is at an inflection point on the multi-modal AI and generative AI timeline, where most software with human-computer interaction (HCI) will see considerable augmentation of these two innovative capabilities. The area under which these two technologies

intersect could potentially change the end-user experience for consuming media and entertainment content. Streaming platforms and devices provide asset-level overlays with features such as subtitles and closed captions, frame/scene-level cast information, age ratings and advisories, and more. Multi-modal AI explains when and where culturally sensitive events occur at various time-stamps in the media asset. For example, acts of violence occur at 5 minutes 15 seconds, and nudity occurs at 10 minutes 35 seconds. Viewers can choose what treatment such scenes should receive when presented with filter options. For example, hide the mouthed f-word in addition to muting/bleeping the spoken aspect, blur the explicit sexuality detailing graphic nudity, change the color of alcohol, or reduce the amount of blood shown on the screen.

These treatments for objectionable scenes are some of the use cases where generative AI excels, as confirmed by the initial prototypes developed by Spherex AI. The possibilities are exciting as they empower global viewers to personalize their entertainment experiences for greater cultural relevance. Similar opportunities for innovation exist in content classification and compliance.

STATE OF CONTENT CLASSIFICATION

Content classification assigns age ratings and warnings to help families make informed choices about the content they watch and protect children from harm. One approach to classifying content is managing the media asset around events that contain attributes of violence, sexuality and nudity, profanity, alcohol and drug use, discrimination, horror, politics, morality, etc. For decades, humans have manually annotated content concerning



Pranav Joshi is the director of product management of AI for Spherex. He is a visionary leader with a passion for leveraging cutting-edge technologies to drive innovation and transform the industry. With a background in computer science, business administration, and extensive experience in product management in the M&E industry, he can uniquely identify new opportunities for AI-powered solutions that can enhance the customer experience, streamline operations, and improve business outcomes. pranav.joshi@spherex.com @SpherexOfficial



Todd Landfried is corporate communications manager for Spherex. A perceptive analyst of the intersection between regulation, technology, and culture in global media and entertainment, he brings invaluable knowledge to the forefront of Spherex's operations. His insights are instrumental in shaping the outward-facing stories that highlight the ways in which Spherex's cutting-edge AI/ML technologies can empower content owners, producers, distributors, and platforms. todd.landfried@spherex.com @SpherexOfficial

classifiable events. However, these annotations were subjective and did not necessarily meet local requirements.

Additionally, it is a time-intensive process to annotate content manually. Manually annotating a full-length title takes three or four times the total runtime. Then there is the time necessary to classify events according to territory-specific rules and make editing recommendations for compliance.

There are three significant considerations to this approach when applying it to real-world, high-volume tasks, such as rating large title catalogs:

1. **What if a classifiable event is missed (involves rework)?**
2. **Apart from the events that require local compliance, what other occurrences could have been annotated to improve content recommendations, search and discovery, and advertising-related business use cases?**
3. **Lastly, but most importantly, how scalable can this process be, e.g., can hundreds or thousands of video assets be assessed daily?**

The unprecedented growth in global content distribution draws our attention to the scalability challenges of content classification. Scalability is more than the method to ensure large volumes of assets can be ingested, processed, and delivered. When a media asset seeks simultaneous release worldwide, it is no longer a phased approach. It is more analogous to the big bang—everywhere, all at once. Spherex is solving this problem by tokenizing the context of cultural events and scenes to address scalability, a multi-layer problem.

In storytelling, context can be examined by characters, plot, and setting, which, combined, form the story's narrative aspect. Location, historical, situational, emotional, cultural, linguistic, physical, and literary elements of storytelling create the context of a scene, the plot, or the overall theme that defines the content's genre.

SPHEREX AI AT WORK (DETECT, INTERPRET, AND CLASSIFY)

Spherex organizes and aligns these aspects of storytelling with contextual tokens, which are then detected and interpreted in terms of their relevance in an event and scene-based setting and classified according to a territory's regulatory policies. Spherex's approach allows detecting the not safe for work (NSFW) events using its proprietary training and fine-tuning foundational AI models. It can integrate third-party tokens/events and uses its Events Transformer to transform those tokens/events into Spherex's taxonomy. Spherex's Culture Knowledge Graph organizes these mappings, and the ML-based rules engine classifies the content based on the business use case. Spherex uses its multi-modal AI to classify content for age-rating use cases, thereby reducing the costs associated with manual labor by over 80% while processing thousands of assets daily.

Humans are still essential in AI-generated content classifications as artificial intelligence is a work in progress, and accuracy is not guaranteed in interpreting context. Spherex's multi-modal AI has deployed expert-in-the-loop processes to fulfill the exceptions when the output does not provide the necessary confidence and processes the relevant feedback to the AI for improvements.

THE FUTURE IS NOW

Global content demand pressures media and entertainment companies to ensure that titles are linguistically correct, culturally relevant, locally compliant, and age appropriate. Humans have been central to those efforts because of the careful observation needed of events within a title, understanding nuance, and the ability to assess country-specific regulations. That is about to change. The promise of generative AI is making complex, data-centric tasks easier and faster. Spherex AI not only automates the detection, interpretation, and classification of objectionable content, it does it at scale. For the first time, proper classification of any content, whether first-run or catalog content, film, or TV, long- or short-form, is affordable and fast. This is only possible with multi-modal and generative AI's intersected capabilities. ■



The Smartest Way to Prepare Content for International Distribution

Spherexgreenlight™ is expert-in-the-loop AI/ML technology that analyzes movies and TV shows for objectionable content, timestamps each event, and provides appropriate age ratings for countries worldwide, ensuring regulatory compliance and brand safety for content providers, distributors, and streaming platforms.



Avoid censorship and cultural missteps



Save time and money on localization



Reach the broadest audience in every market



The screenshot shows the Spherexgreenlight interface for analyzing the movie 'Divergent (2014)' in Brazil. The interface includes a search bar, a video player with a 'Censor Copy' overlay, and a 'Classified Events' table. Below the video player is an 'Event Timeline' table.

Category	Count
Alcohol & Drugs	1
Drugs	2
Blasphemy	3
Cursing	1
Cinematography	1
Cinema	1
Discrimination	1
Behavior	1
Horror	1
General	1
Irritable Acts	1
Behavior	1
Self-Harm	1
Nudity	2
Nudity	2
Profanity	1
Profanity	1

Age Group	Flagged
Adults	18
Young Adults	16
Young Adults	14
Teens	12
Older Kids	10