Whitepaper

censhare

Thinking
About Content

Artificial Intelligence and Machine Learning
in Content Management

*freedom
to create*

Contents

The Challenge of Expanding Data in Content Management

Many organizations seek to grow by widening their product and service portfolios, and attempting to reach more markets.

Many organizations seek to grow by widening their product and service portfolios, and attempting to reach more markets. These may be entirely new products – for example, GoPro generates most of its revenue by producing and selling action cameras, but in 2015 launched a premium licensing portal to re-sell user generated content as a second business line. Alternatively, businesses can simply create variants of a product to reach smaller groups of more targeted customers.

This growth means the number of content items requiring management also grows, and rapidly, including product data, descriptions, images, media assets, respective translations, and target group assignments. The list goes on.

Some organizations decide to grow by moving into previously unserved markets. This poses its own challenges for content management, as literature and websites must be adapted to and translated for these new markets. Products or services may be restricted, subject to extra standards, or face considerable legal barriers to entering a market, requiring multiple different versions of websites and marketing material.

Further complexity arises with the need to produce personalized messages. Consumers are bombarded with information every day. They find it difficult to focus on those most relevant to them or indeed to find the content they need. any ignore most of what they see, potentially wasting marketing effort due to lack of focus. As such, consumers are increasingly demanding messages relevant to their needs and lifestyles, and accessible in a way that is convenient and easy to use.

This need for personalization adds to the complexity already at hand in requiring a more individualized composition of messages. Add in the growing number of marketing channels and it becomes almost impossible for any realistically sized team to produce and manage the number of different communications needed.

Manual methods may work for a certain level of complexity, but at some point, an extra language or piece of collateral will make these no longer practical.

Many businesses may conclude that extensive personalization is not worth the effort. They may choose instead to concentrate on those channels that reach the largest proportion of their customers. However, even the largest channels are likely to diminish in importance over time, while simultaneously, the number of channels continues to increase, adding further complexity to content management and channel execution.

Consumers also want a rapid response to any interaction with the business. Being always online and mobile, they are used to having information constantly available at their fingertips.

For example, they expect the reassurance of an immediate email confirmation of their purchase. Delays in confirming a purchase lead to consumer suspicion that they may be dealing with a fraudulent site or vendor, and discourage the customer from future interaction with the organization.

Another example is that consumers also expect every function in the organization to know the details of their purchase right away.

By contrast, a joined up approach will give the customer a more integrated experience, for example they will not receive adverts for a product they have just purchased. More positively, it could allow the customer to get instruction manuals, online warranties or adverts for accessories.

Managing the huge amounts of data required for multiple challenges and markets has been made easier with the adoption of advanced, all digital content management systems that make use of Artificial Intelligence (AI) and Machine Learning (ML) techniques.
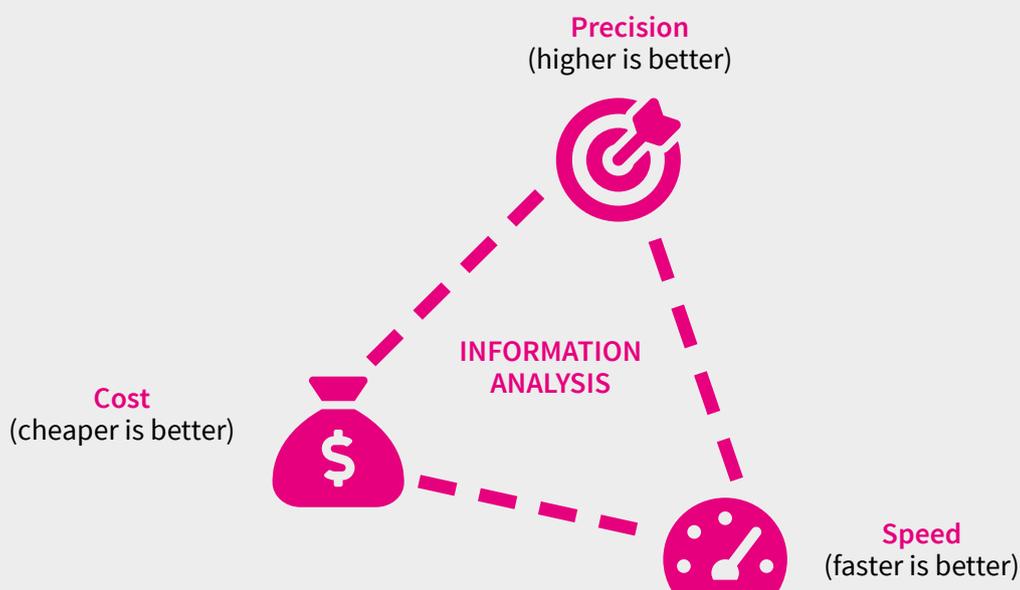
Precision, cost and speed – a difficult balance

As demands increase and complexity grows, there is an ever greater need to balance this against some business aspects that cannot easily be reconciled – precision, cost, and speed.

Typically, only two of the three objectives in this triangle can be achieved – precision and speed, cost and speed, or precision and cost – with the missing third objective presenting limitations.

- You may aim for precision analyses at high speed, but that won't be cheap
- You may aim for fast, affordable analyses, but that won't be very accurate
- You may aim for precise, affordable analyses, but that won't be very fast

Through their ability to improve productivity, reduce errors, and format and arrange content rapidly, AI and ML can help achieve this balance.

Precision
(higher is better)

INFORMATION ANALYSIS

Cost
(cheaper is better)

Speed
(faster is better)

AI and ML help achieve a balance between precision, cost, and speed in content management

How Artificial Intelligence and Machine Learning Can Help in Content Management

The huge and growing amount of data involved in the content management challenge means it cannot be handled manually.

Automated processing is required and AI/ML, which cuts the costs and time needed to manage large volumes of content, is a good solution. Also, because systems enhanced by these technologies learn as they go, precision also increases with the greater amount of data available for analysis – marketers see better results the more they use it and the more data they feed in.

The drawback is that to process all this data, AI and ML require a lot of computing power. Fortunately, this has increased tremendously over the past decade, with microprocessor clock speed now above 10 Gigahertz and continuing to increase.[1]

As well as raw computing power, there have also been developments in techniques to store and analyze data. Chief among these are cloud services that make it easy for developers to store almost unlimited amounts of data and allocate servers as and when required. The leader in the field is Amazon Web Services (AWS).

In 2018, AWS had 40% of the global market for public IaaS (Infrastructure as a Service) and PaaS (Platform as a Service), and has achieved a 45.8% increase in revenues year on year.[2,3]

AI/ML services are also now available as cloud services, for example, automated image analysis by Google Vision API or smaller innovative companies such as Acrolinx.

The obvious attractions of AI have seen a huge upsurge of interest in the technology, with nearly $29 billion expected to be invested worldwide by companies in all sectors by 2021. There is also a growing appreciation of the impact of AI among marketers – 80% believe AI will revolutionize the marketer's role, 82% believe AI allows staff to focus on value generating tasks because of its role in automation, and 86% say AI makes marketing teams more efficient and more effective. Some 60% of marketing leaders already believe AI can help them run more effective advertising and media campaigns.[4]

1 Roser, M. and Ritchie, H. (2019). Technological Progress. [online] Our World in Data. Available at: https://ourworldindata.org/technological-progress [Accessed 7 Nov. 2019]

2 Carey, S. (2019). The history of AWS: defining moments from 2002 to now. [online] Computerworld. Available at: https://www.computerworld.com/article/3412382/the-history-of-aws--a-timeline-of-defining-moments-from-2002-to-now.html [Accessed 7 Nov. 2019]

3 The Next Platform. (2019). Navigating The Revenue Streams And Profit Pools Of AWS. [online] Available at: https://www.nextplatform.com/2018/02/05/navigating-revenue-streams-profit-pools-aws/ [Accessed 7 Nov. 2019]

4 AppNexus (2018) The digital advertising stats you need for 2018. AppNexus

The flipside is that many marketers are naturally cautious about adopting AI and ML technologies across their operations. Their trust must be earned over time. Therefore, it is important to introduce the technologies only where they will be beneficial and transparent, avoiding their implementation in areas where there will be little gain.

AI and ML techniques are still at an early stage in their development, so it is advisable to manage the expectations of decision makers, who can quickly become very enthusiastic about the perceived possibilities of the technology. As with all new developments, overpromising may lead to disappointment, with future opportunities also being discounted too readily.

AI solutions also require a certain amount of supervision, with marketers feeding them with information. This supervised learning does not match the way humans learn and is a barrier to making AI more human like.

What is Artificial Intelligence and Machine Learning? And what's the difference?

Artificial Intelligence (AI) is the discipline of creating intelligent machines, while Machine Learning (ML) refers to systems that can learn from experience.

AI covers anything related to making machines smart. The machines in question can be a robot or piece of industrial machinery, an everyday domestic item such as a dishwasher or home management system, or even a software application.

ML is often discussed in association with AI, but they are not the same and the two terms are not interchangeable. ML refers to systems that can learn by themselves, working on data to recognize patterns and draw inferences about how things work. Essentially, ML systems can become smarter without the need for human instruction or guidance.

Most AI systems require a ML component. This is because the intelligent behavior that AI seeks to emulate is based on knowledge, which in turn requires an ability to learn.

Improving Automated Image Processing

In today's digital asset management, one of the most useful and productive abilities of AI is the automatic recognition of images.

This is achieved via image metadata, i.e. text information relating to an image file. The text is either embedded in the file or contained in a separate file associated with it. Image metadata includes details about the image itself and about how it was produced.

AI can enhance this metadata by automatically opening each image, analyzing its content and finding known patterns in the mass of pixels. Once a pattern has been identified it can be used to generate a list of additional descriptive keywords or tags.

Because of its automated nature, the quality of the tagging can be more precise, although there are limitations to the accuracy that can be achieved. More important to its use in marketing is the higher level of consistency it can accomplish compared to a human.

This process can be very powerful when combined with automated translations of these tags, for example. If the content owner needs to enter a new market with a new language, tags can be translated automatically, allowing them to be used immediately by local staff. Imagine the speed gains for a newly formed local marketing team that can, for example, search all centrally available images in their own language. Of course, for automated content selection process, e.g. for ad campaigns, content tagged in the local language will also be a big help.

There is the danger that AI could get large parts of the tagging wrong. To manage this risk, the original keywords produced by a human should be preserved so that the AI selected keywords do not overwrite them.

The ability to recognize images makes AI a very valuable tool for several use cases:

Automatic classification of products

Automatic classification of products based on product images is valuable for retailers needing to manage product information for a large product portfolio. They can use it to close gaps in the product information, for realigning entire product portfolios or product catalogs automatically, or for identifying and removing duplicate data.

Automatic composition of product descriptions

With consistent metadata, it becomes easy to compose detailed product descriptions – based on templates with placeholders – that reference information that is present in an image, expressed by its metadata.

Finding images based on sentiment and colors

Many marketers need to illustrate a text feature or an advertisement with an image that expresses or encourages a certain emotion in the audience using colors or a human subject's expression. With an AI based image recognition system, they are spared the need to wade through countless unsuitable images and can instead apply filters to show only the colors and sentiments they are looking for.

Following is an example screenshot for automatic detection of mood:
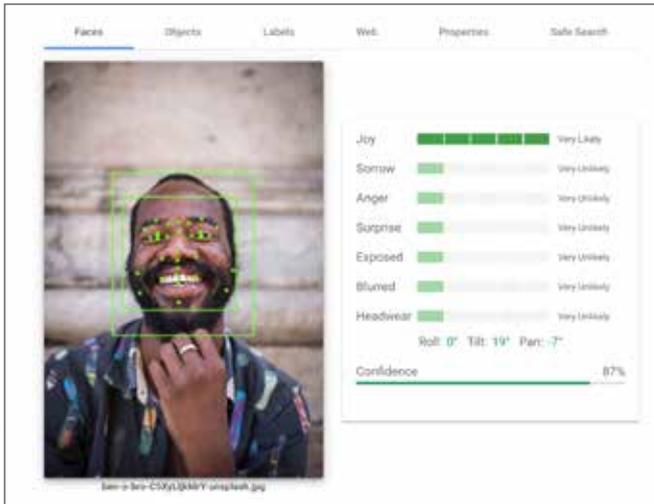
AI can analyze images to identify particular moods or expressions

Finding identical or similar photos and images stored in the internal Content Hub

This capability can be used to eliminate duplicate images from the content store, avoiding double payment of royalties. The goal is a 'single source of truth', with each image being stored just once in the best resolution and with all legal restrictions captured, for example, image source, licensed when, licensed by whom, and expiration date.

The ability to find identical or similar images online is useful in several ways. These include discovering how widespread a piece of content is, for example, a promotional photo showing a product or to identify illegal or unwanted uses of content, such as copyright violations, to protect a brand from damages. This information can be used to trigger legal action against offenders. On a more positive note, this ability could be used to see whether a specific photo makes a difference in the market. For example, if you are marketing holidays, do you show a certain place like everyone else, or is your view unique?

Automatically determining text in images

In certain circumstances or marketing activities, it may be necessary to recognize text in imagery, perhaps to identify those images which contain a certain brand name or a specific number.

Imagine, for example, a sports event. For larger meetings, the organizer may use professional sports photographers to take photographs of individual participants, with the idea of offering them a photo of themselves as a souvenir.

This produces thousands of photos that would usually take days, if not weeks, to properly tag with the correct starting numbers. Some photo services ask the user community to tag photos, effectively outsourcing the task to them. This may work, but it results in random and inconsistent tagging as it cannot guarantee that all runners will be identified by other users correctly. Using an AI based system to automatically tag the photos with the runner's name and number is much quicker and reliable, providing the successful participants with a perfect customer experience.

Automatically determining places where a photo might have been taken

This ability is particularly useful for user generated content, as it can help identify user preferences and needs and allow a marketer to more accurately target its offers to meet those needs. For example, if a user uploads several beach photos taken in Thailand, we might conclude that the user enjoyed visiting the country and might be receptive to offers involving travel to the region.

The value of these use cases is backed by leading analysts who regularly screen the market for content management platforms offered by the industry's major players. They report that AI adds value to content, with many vendors having artificial intelligence capabilities as a fundamental part of its platform. Most solutions use technology from Amazon, Google, or Microsoft, with the exception of Adobe, which apparently uses proprietary technology. As in our examples, they see a key use case for AI in applying generic tags instead of asking people to go through large volumes of content.

Despite these abilities, we must keep in mind that AI lacks true creativity, which remains a vital part of a marketing campaign. Although AI is fast and capable, it cannot yet supply the originality and creative thinking offered by humans and we therefore need both people and AI solutions to provide a complete answer to marketing challenges.

The censhare way

To provide these AI facilities, the censhare platform integrates Google Cloud Vision, a cloud service for image analysis. For example, Google Cloud Vision identifies brands, landmarks, keywords, or web content related to an image. The Google Cloud Vision functionality is available for JPEG, PNG, or GIF images.

censhare offers the following functionality from the Google Cloud Vision API:

• Assign keywords – shows all keywords that have been identified and assigned to an asset, along with their relevance

• Identify the main colors of an image – shows which colors from the color palette in censhare were detected, as well as calculated coverage of that color in the image
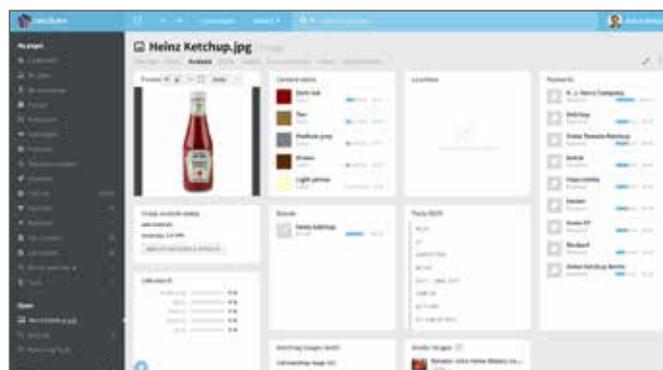
censhare can recognize brands based on colors and text in the image

- Related web pages – Google Cloud Vision searches in three web related categories:
  - Images that have a full match with the analyzed image
  - Images that have a partial match with the analyzed image
  - Web pages that contain a full match of the analyzed image
- Identify locations shown in the image – shows all landmarks that were detected, as well as a score for the location's relevance
- Identify brands related to the image – shows all logos detected. Google also calculates a relevance score for the brand logo
- Detect text shown in the image – Google Cloud Vision uses Optical Character Recognition (OCR) to check for text in the analyzed image. If text is detected, censhare stores the result as plain text in a storage item attached to the image
- Check for inadequate content (safe search) – identifies images that may be inadequate or inappropriate, assessing them in five categories – adult content, spoof (content that has been modified to make it funny or offensive), medical topics, violence and racy imagery.

The limits of AI

Although AI technology is developing rapidly in its ability to recognize the content of images, it still has its limitations. In a recent test of the Google Vision API, the system correctly identified Helsinki Cathedral with a 57% confidence score. It also claimed to have found one "fully matched image", although this proved to be a similar building in Lisbon, Portugal.

The Google API also failed to identify a general shot of Kiel in Germany, despite Google Streetview showing a similar shot and Google listing many similar photos.

Another drawback is that auto tags may be useful in a machine context but may not be from the expectations and perspective of a person. For example, a machine might identify the contents of a photograph as a white ceramic cup. However, other clues in the picture, such as water vapor rising from the cup and a pile of beans, would allow a human to identify the subject as a cup of coffee.

A machine may recognize this image of a white ceramic cup, but a human will know that it shows a cup of hot coffee

Analyzing Text Assets

Accurate text, that provides a clear, unambiguous summary of features and benefits, or by contrast achieves an emotional connection with an audience, can be very persuasive and encourage people to learn more about the company's products and services.

Depending on the audience, the text may need to meet various criteria for style and keywords.

Artificial Intelligence and Machine Learning can already help content creators by providing a real time check of text as it is being written, assessing it against content goals and brand guidelines. As the writer works, the system can start analyzing and measuring the text against goals based on several criteria:

• Terminology
• Spelling and grammar
• Style and clarity
• Conversational tone
• Acronyms
• Scanability
• Gendered words
• SEO

For example, the censhare Universal Content Management platform integrates the AI analysis service from Acrolinx. This checks the copy in real time and provides the writer with a comprehensive list of potential issues, including suggestions for resolving them.

"Create better content, faster"

Andrew Bredenkamp, founder and CEO of Acrolinx, outlines the benefits that the Acrolinx platform brings.

censhare: What trends are you seeing in the industry?

Bredenkamp: We work with some of the most innovative brands in the world to address the content creation challenges they have. Recently we've noticed the trend towards data driven decision making around content. Increasingly, everyone is being asked to provide data for everything as a rule – after all, we're a data driven industry – and the last frontier is the data around the actual content.

censhare: How does Acrolinx support marketers in this respect?

Bredenkamp: We help marketers create better content faster by measuring every aspect of the creation process. Up to now, there was a lot of guesswork and gut feeling involved when creating content. This can be reduced significantly with Acrolinx as we bring more automation and more analytics to the content creation process.

censhare: But don't you experience serious backlash from marketers? Many of them might be skeptical about introducing automation to the creative process.

Bredenkamp: I think technology can be your friend. You see, we're not one of those AI companies that comes and starts to replace humans in the process or starts to automate the creation of content. There's nothing we're looking to fundamentally disrupt or alter the way things get done. Essentially, we just remove some of the pain that large companies and organizations experience when they try to work on content.

As with a standard word processing package, simple issues such as spelling mistakes can be corrected immediately. More complex issues, like style, can only be resolved by a skilled writer but help is provided in the form of suggestions and questions which guide action, such as, "Shorten this sentence" or, "Do you really need this phrase?"

For instance, sentence length can be a good indicator of content complexity. Shorter sentences are generally easier to understand – the system can use the number of these to indicate the readability of a text.

The integration also offers a scorecard that summarizes the findings and provides metrics that give a guide to the clarity, informality, liveliness and readability of the copy. Marketers can use these metrics as additional guidance to complement their own subjective impressions of a text.
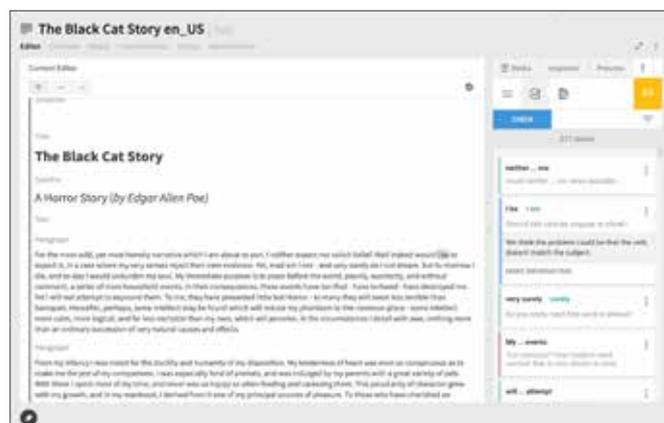
censhare's Acrolinx integration provides useful real time feedback to the author during the creation of an article

And now: AI generated content creation

News agencies and newspapers are already exploring the use of AI technology to automatically create meaningful articles for human readers and to assist journalists in finding trending topics. For example, news agency Reuters has launched Lynx Insight, a tool that "can augment human journalism by identifying trends, anomalies, key facts and suggesting new stories reporters should write."[5]

The Washington Post uses a robot called Heliograph to cover congressional and gubernatorial races on Election Day as well as DC area high school football games.[6]

PA (the Press Association), the national news agency for the UK and Ireland, has received funds from Google to build "RADAR – Reporters And Data And Robots – a new service which will create up to 30,000 localized stories each month from open data sets. RADAR is intended to meet the increasing demand for consistent, fact-based insights into local communities."[7]

5 *Reuters (2019) The cybernetic newsroom: horses and cars. [online] Available at: https://www. reuters.com/article/rpb-cyber/the-cybernetic-newsroom-horses-and-cars-idUSKCN1GO0Z0 [Accessed 27th November 2019]*

6 *Digiday (2017) The Washington Post's robot reporter has published 850 articles in the past year. [online] Available at: https://digiday.com/media/washington-posts-robot-reporter-published-500-articles-last-year/ [Accessed 27th November 2019]*

7 *The Press Association (2017) PA awarded €706,000 grant from Google to fund a local news automation service in collaboration with Urbs Media. [online] Available at: https:// pa.media/2017/07/06/pa-awarded-e706000-grant-google-fund-local-news-automation-service-collaboration-urbs-media/ [Accessed 27th November 2019]*

The Need for a Flexible Database

AI and ML are powerful tools, but to make the most of them, vendors also need to adapt and improve some existing technologies and techniques. One of the main technologies is the database an organization chooses to use with these advanced tools.

With AI able to handle vastly more information about content, it is easy for a database to become overwhelmed. Traditional database structures do not have the flexibility and speed needed – if new data fields are required, this must usually be performed by IT specialists, which can be complicated and takes time.

By contrast, a semantic database is fully flexible, allowing new metadata fields to be added easily. Users can attach new information and links to any other information, naturally and 'on the fly' without support from IT specialists. For example, a sentiment field could be added to the database of images and populated automatically.

A semantic database is also more comprehensive than traditional database structures. It is powerful because it uses all information about a piece of digital content to create relationships with other assets. These include the photographer, usage rights, an associated article, the model pictured, previous campaigns, the photographer's other work, other images of the model – in fact almost any relevant information.

Using semantic technology, the censhare Universal Content Management platform understands these relationships. It can quickly search through them to provide insights into any aspect of the content and how it is used. This speed means it can respond almost instantly to a processing request, even for the very extensive and diverse content types used by the largest global organizations.

censhare's content management platform uses such a database, which provides a solid foundation for future AI implementations.

censhare uses a semantic database that can also visualize an asset's relations to other assets and information.

Conclusion

The desire to grow markets and revenues inevitably means that companies and organizations must manage a rapidly increasing number of content items. These content items include product data like descriptions, images, media assets, target group assignments, and respective translations.

Moving into new markets with new regulations and the need to produce personalized messages compounds the content management challenge.

Given the amount of material to be handled and processed, meeting this challenge with purely manual methods is no longer possible. However, it has been made easier with the adoption of advanced, all digital content management systems enhanced by Artificial Intelligence (AI) and Machine Learning (ML) techniques.

Due to the high speed and inherent automation of these techniques, AI and ML cut the costs and time needed to manage large volumes of content. More specifically, because systems enhanced by these technologies learn as they go, precision also increases as more data is analyzed.

Although it has not achieved the levels of accuracy that a human can bring to the task, AI has a big role to play in managing and assessing images – it can recognize the content of images, find images based on emotion or colors and identify text, brands and locations.

This has important implications for managing campaigns and choosing the right images to suit an audience with particular characteristics or demands.

AI can also help in the creation of written content, analyzing style and sentence length against criteria defined by the user, helping the organization produce text that is easily understandable and appealing to audiences.

censhare Universal Content Management uses these techniques, together with a flexible, semantic database, to generate highly personalized content for individual customers and manage full marketing campaigns.

The power of AI and ML turns the challenge of managing very high volumes of content into a huge opportunity to find, develop and use the exact content item to achieve a precise effect.

about censhare

Our pioneering, universal content management platform lets you connect with your audiences on any channel, in any language, locally or globally. Clients like Allianz, Dyson, Christie's, Lufthansa and many more have already discovered new freedom to create and deliver consistent quality content with exceptional efficiency.

Contact us at **censhare.com**