



heads *up*

 the
group
of
analysts

Artificial Intelligence

Innovation in image recognition



26
COMPUTER
VISION

Traditional tagging, as it is still used today in media management, will soon cease to be used. Technologies such as the semiautomatic tagging are revolutionising image file management already today.



8

ARTIFICIAL INTELLIGENCE –
EXTRA-POWER FOR ISCM

Artificial Intelligence is a ground-breaking trend and systems using machine learning and similar technologies can have a positive impact on Information Supply Chain Management and therefore on the data flow within the company.



34

OLIVER BAUM
General Manager
at apollon



18

The Group of Analysts

Introduction 4
Dialogue and DNA

Exposé 6
Take care of your image(s)!

Artificial Intelligence – Extra-Power for ISCM

Introduction to ISCM 10
The networking of systems

Classification of software markets in ISCM 12
The three pillars of ISCM

Organisational incorporation 14
Digitalisation and the human factor

A solution for every need 14

The amount of image data will continue to increase and the automatic generation of metadata will soon be the only way to handle this data.

AI-controlled tagging

Thesis/Antithesis 16

Photography through the ages 20
The beginnings

How AI is revolutionising image data management 22
Metadata and tagging

Image recognition – models to suit every demand 28
Prebuilt versus custom models

Summary

Heads

Oliver Baum 34
General Manager at apollon

Temel Kahyaoglu 36
Board member of The Group of Analysts AG

Carmela Melone 38
Analyst at The Group of Analysts AG

Sections

32 Imprint 40

Picture credits 41

The Group of Analysts

The Group of Analysts (TGOA) is a globally active market research and consulting firm that has held a pioneering position since its foundation. The Market Performance Wheels (MPW), a new analysis method for the performance of the software market based on a revolutionary methodology of analysis, create beautifully designed, circular forms of knowledge. They enable a multi-dimensional examination of companies and their products on the market. The independent analysts support companies with this knowledge during the personalisation of the ISCM (Information Supply Chain Management) from its analysis to its introduction.

Whitepapers on red-hot digital topics and revolutionary projects such as the EDEN study are also among their range of services. The latter stands for European Digital Entity Norm and describes an ongoing project to determine the current state of digitalisation in European companies, motivating them to tap into and use their existing digital potential.

In order to provide a platform for this wealth of knowledge along with giving the entire software market the opportunity to communicate and collaborate, TGOA has created a space at tgoa.com where demand and supply can meet and exchange views on the topic of digitisation with our independent analysts.

With its services, the company headquartered in Meersburg demonstrates an almost engineer-like depth of detail which provides a high-quality alternative to American analysts such as Gartner, Forrester and co.

Take care of your image(s)!

In this Whitepaper we have worked together with apollon in order to tackle the growing significance of image files to companies, while taking a look at the opportunities companies have at their disposal in order to master ever-increasing volumes of data. This not only concerns the efficient archiving and storage of image and video material, but also the ability to quickly locate relevant media while connecting it with the appropriate information from Product Information Systems, for instance. There are a multitude of diverse examples of the use of image files within companies, and every scenario contains its own requirements of participating systems and persons.

Manufacturers and retailers naturally manage a large volume of image files for internal archiving and logging purposes on the one hand and for communicating with customers on the other hand, such as through an online shop, a print catalogue or through social media channels and the company's website. It is crucial for the system that stores the image files to be integratively linked with other relevant systems, such as the PIM system or distribution channels. This ensures that the information can be accessed at any point at this location as soon as it is required.

Renowned brands, to name a further example, are often affected by brand piracy, meaning that it is of vital importance for them to continuously monitor the market to ensure that they are able to respond immediately in such cases. Image files again play a particularly vital role in this scenario: Searching for images online is becoming increasingly important and is especially useful when researching the competition. Optical Character Recognition (OCR), for example, quickly and reliably searches the internet for relevant online presences by converting images of text into machine-readable characters.

This technology also provides major advantages when it comes to insurance: The initial process of recording damages becomes much quicker through the use of relevant software which is able to automatically identify specific types of damage, such as hail damage, therefore virtually presorting inbound damage cases.

There are just as many diverse possible uses of image recognition software as there are large varieties of current models and technologies available on the market – be it manual or semiautomatic tagging or computer vision, prebuilt and custom models. Various metadata standards and a variety of solutions and different combinations of technologies make it hard to keep everything in perspective. Every company must decide for itself which concepts meet existing requirements the most, which internal roles require defining as well as how important the use of image files will be to the company in future. This much is clear: image recognition is still in its infancy.

Artificial Intelligence – Extra-Power for ISCM



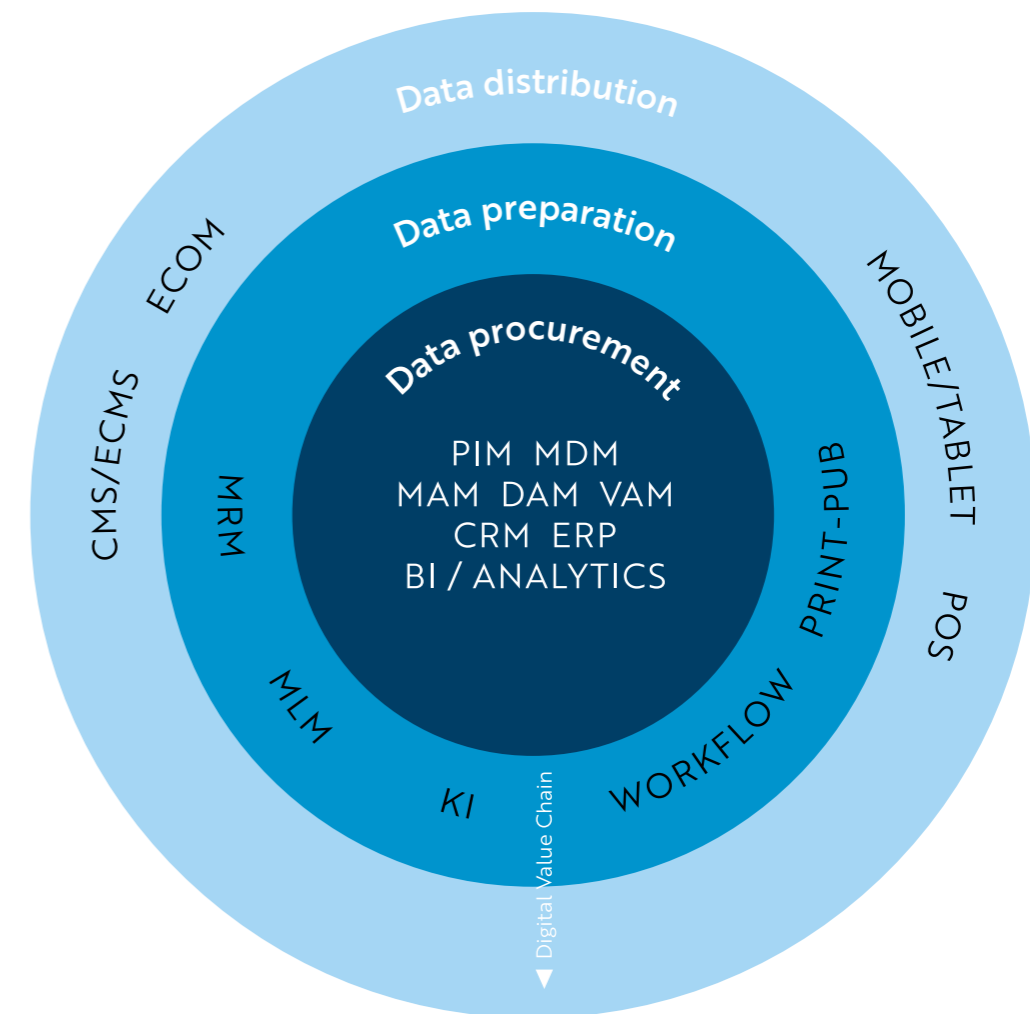
The Information Supply Chain Management

The Information Supply Chain represents a chain of connected software solutions in parallel with the classic Supply Chain. All relevant data on products and their associated services and target group-orientated information is therefore procured, recorded, completed, prepared, processed and distributed.

The term ISCM was coined by the Chief Analyst of The Group of Analysts, Temel Kahyaoglu, in 2008. From this point onwards, the term has been firmly implemented in the digital world, becoming indispensable. This can be traced back to the universal concept behind the ISCM in particular. Existing software markets are subdivided, simultaneously creating flexibility for the necessary adjustments. As depicted in the below illustration, the software markets that are currently crucial for ISCM to the Supply Chain Management's (SCM) are assigned three pillars. These are acquisition, preparation, and distribution of data.

While the digital value chain runs linearly from procurement over processing to distribution, the underlying flow of digital information is connected across all systems.

The categorisation of the software markets listed should not be seen as a rigid construction, however. The basic structure of the ISCM model with its subdivision into three levels remains permanent, to which software markets are flexibly connected. Different software markets have developed at varying paces. In addition, some markets are closer to each other than others. The software markets depicted only represent a current extract. The elements on the various levels can be adapted through alternating requirements and conditions. These adaptations will become necessary within the course of future developments on the market, resulting in new software solutions, processing new types of information, supporting innovative models of cooperation and the operation of new distribution channels. This merges existing structures, diffusing them and causing them to potentially migrate. Software markets can be adapted individually to the predetermined requirements of the respective company. It is therefore only a matter of time until further new fields are incorporated into even fully-elaborated overall ISCM pictures for the respective company.



Classification of software markets in ISCM

Data procurement

Systems classified as data procurement provide special functions for the maintenance, upkeep, onboarding, preparation, export and mutation of data. The following systems in particular are included: Product Information Management (PIM), Media/Digital/Video Asset Management (MAM/DAM/VAM), Customer Relationship Management (CRM), Enterprise Resource Planning (ERP), Business Intelligence (BI) and Analytics. Data on the product is created as early as the beginning of the product lifecycle. When setting up products in the ERP system, the

master data is entered and enriched with content and information relevant to sales in the PIM system.

Through the course of further steps, media such as product photos or videos is added to the MAM system. CRM systems meanwhile contain all customer-relevant information that is also related to products. Valuable data from various sources, including distribution channels, is included through systems for BI and Analytics. This results in a cycle in which data is evaluated by its relation to each other.

Data preparation

This describes the area where content is produced in various languages, complexities and forms. The area also regulates workflows and cooperation. Examples include topics such as Marketing Resource Management (MRM), translation management with multi-language management (MLM) or Translation Memory Systems (TMS), workflow management, print-publishing (Print-Pub) or dynamic publishing systems (DPS) as well as artificial intelli-

gence (AI). The systems available for MRM and Workflow Management are still relatively recent. The supply market is therefore correspondingly fragmented. Multilingual management is more developed when it comes to providers, but is often still operated through external service providers. This results in a cycle in which data is evaluated by its relation to each other.

Data distribution

Data distribution includes the release of all information from the ISCM area of data procurement through the various channels facing the customer. It is important to plan the release of the data procurement system, such as PIM, and to provide the associated processes and options for collaboration before delivering them into the respective channels. The area of data distribution mostly consists of (enterprise) content management systems (ECMS/CMS), e-commerce systems (ECOM) as well as solutions for mobile/tablet and Points of Sale (POS). The fact that the software market is undergoing changes applies here, too. The areas of e-commerce and (E)CMS are the exception, as they are clearly defined.

The decisive factor is the release competence, as customers must be able to access any relevant product information through all channels and touchpoints. The comprehensive supply of product data to the client is also essential for B2B.

Manufacturers of systems upstream to data distribution must therefore be able to perfectly orchestrate the output while working together in harmony with all data distribution system providers. This is guaranteed by established PIM, MAM / DAM and MRM manufacturers.

Organisational incorporation as a significant success factor

Companies have to be aware of the overall ISCM challenge. This also includes the task of securing the smooth interaction of the numerous individual systems. The ability to connect is accordingly an essential factor for success that needs to be guaranteed with the help of ISCM. Connecting the individual software markets to one functioning ISCM total system landscape consists of experience mixed with technical knowhow during the introduction and connection of the required software solutions. It also involves numerous people spread throughout the company, who are expected to work together in an effective manner.

Alongside ensuring the quality, actuality, security, and data protection, the management of the Information Supply Chain also guarantees the consistent digital and central storage of data in suitable systems. Ensuring the smooth exchange of data between all systems involved and an imperatively high process quality are also among the main tasks of Information Supply Chain Management. As a result, accepted forms of cooperation and solutions for controlling respective processes are required. Major changes always go hand in hand with more comprehensive ISCM projects, which is why we speak of digital transformation. The internal effort associated with the topics of digitalisation is often underesti-

mated by companies. Awareness for long-term projects equipped with funds and resources in specialist departments is often insufficient. Digitalisation in the sense of ISCM has to be positioned at C-level due to its holistic view and financial and resource-devouring dimension. Correlative new positions such as Chief Digital Officer are only just being created. It is vital that this settlement takes places in a cross-departmental manner while being integrated and coordinated. The 2017 EDEN study (The Group of Analysts 09/2017) however has shown that this recommended procedure is only being carried out by approximately a quarter of all companies. There is significant demand for action for the remaining three quarters. Furthermore, the lack of connected coordination partially leads to considerable time expenditure for the companies. Which is why a permanent area of responsibility has to be introduced for the management of the Information Supply Chain in the company during the project introduction phase of ISCM technologies. A responsible team therefore has to be set up.

A solution for every need

Companies have a vast amount of opportunities at their disposal in order to target their system landscapes towards specific requirements when it comes to data procurement and data preparation in general, and in particular when it comes to image file management. The type of image file stored also

depends on the type of company, as does the use and purpose of these files. Because the current and the planned future use of image recognition is closely linked to the company's digital strategy, the decision for or against potential

technologies must be made at management level. A dialogue should be pursued with various departments to enable an informed decision to be made. Opting for pioneering image recognition technologies often requires creating new roles and responsibilities within the company. Some models that use artificial intelligence require somebody who trains the system for company purposes, which involves feeding the system with image files and information pertaining to their contents. This ensures that the recognition system's precision meets practical requirements. While manual tagging requires less explanation, it often takes up a lot more time, especially with large volumes of images. In addition to this, there is a risk that the tags are inconsistent, and the image metadata standards are unable to fully and comprehensively describe complex images.

There is no one-size-fits-all solution when it comes to Digital Asset Management. Due to the rapid pace of developments in image recognition through innovation and major investment in artificial intelligence, it is becoming increasingly important to closely observe the market for these systems, while evaluating it on the basis of the company's own requirements. There are several crucial questions which companies must ask themselves in this context.

Which areas do I even require images for? This refers to product catalogues and e-commerce, but also to company communication through websites and social media as well as internal knowledge systems. Requirements can then be

derived from these needs. What could these requirements include? This could include multilingualism when it comes to tagging. For companies that operate on an international level and which have an online store it is crucial that images can also be found using English tags, while key words in the local language may be sufficient for regional companies with a very small range of products. The purpose of course has an impact on the need for the image recognition's level of precision. A large online retailer such as Zalando will require high-precision standardised metadata descriptions in order to enable the customer to filter clothing or shoes according to colour, style or material. The system must therefore have previously been trained with a large number of photographs from various perspectives in order to guarantee a satisfactory level of precision.

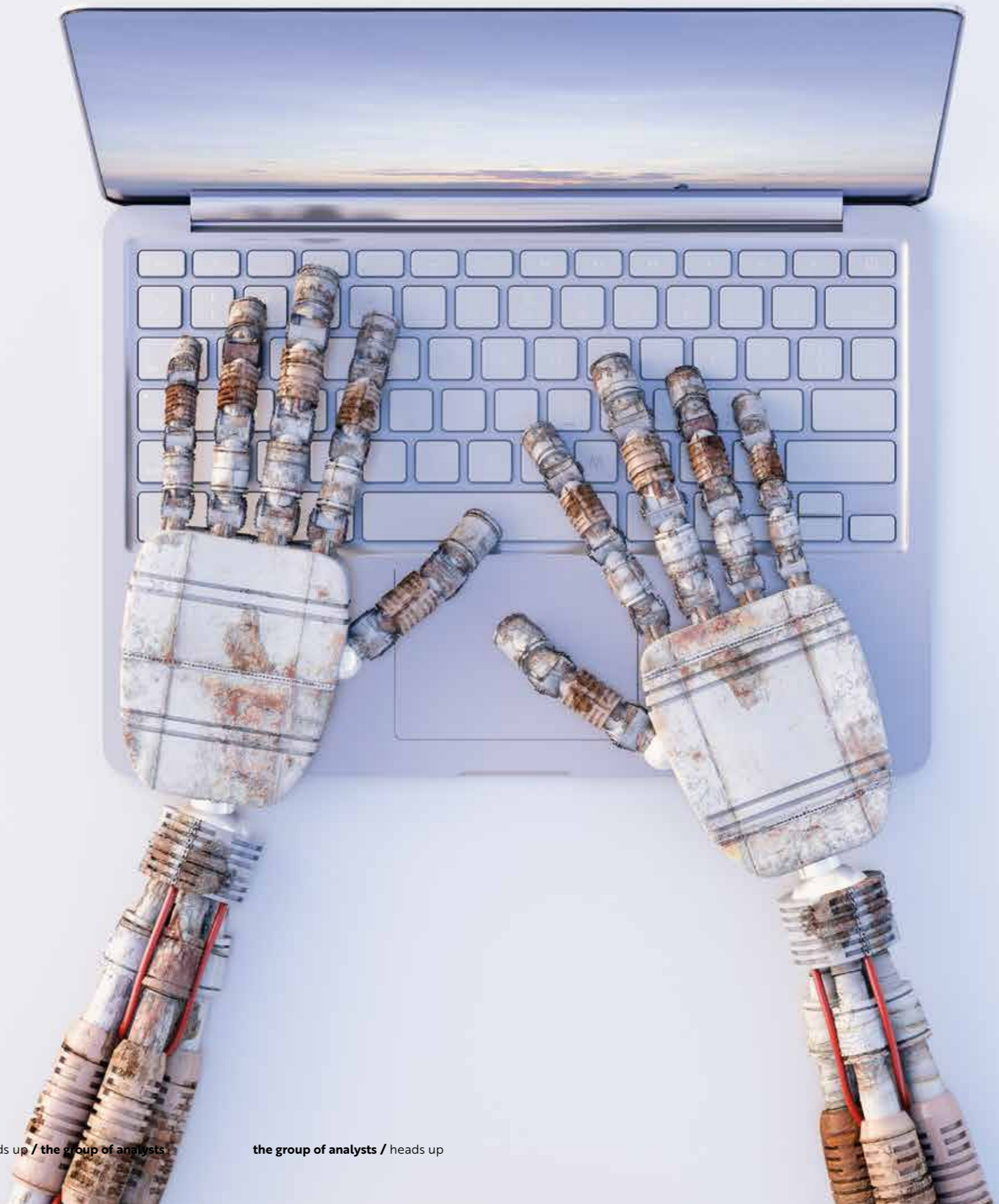
What type of model is suitable? A variety of different models are available that are suitable depending on their purpose. Factors to consider include whether pretrained models or own models are preferred, whether deep learning is currently critical to success or whether it will be in future, whether user-generated content is an issue and if so, whether machine-assisted monitoring of foreign contents in accordance with relevant guidelines will be used or whether cloud services or in-house solutions are preferred.

Using various types of services particularly makes sense in more complex companies who wish to remain flexible and open throughout their digitalisation journey. Promising technologies that are increasingly influencing our everyday lives, such as augmented reality and voice search, require efficient key words, which is why AI-supported tagging plays an important role in this. Companies ultimately benefit the most through the much faster, more precise and consistent editing of image metadata, and therefore through an optimised Information Supply Chain.

**At the moment AI
is only relevant
to companies
such as Google and
Facebook.**

**AI can already
be crucial to the
success of numerous
business processes
of all manner of
companies.**

AI-controlled tagging



Photography through the ages

The beginnings

The Frenchmen Nicéphore Niépce and Louis Daguerre developed the first technical and chemical procedures in order to permanently record a light-resistant image on a physical carrier around 1826 and are therefore known as the inventors of photography. What began as an experiment has since been continuously developed and perfected. Nowadays, photography is fast, high-quality and, above all, available across a large scale. Photography is characterised by constant changes in close connection with technological progress, just as it has been from the start. It has gone through many evolutionary stages: from black and white to colour, from analogue to digital, from digital to networked.

Our approach to images has significantly changed with each technological development. The very first images were so-called ferrotypes. This procedure required long exposure times during which the subject was not allowed to move for fifteen seconds, and the photograph would take a further ten minutes to develop. This waiting period was reduced to mere minutes following the introduction of the first Polaroid cameras in the 50s. It has been a long time since people have had to put up with waiting for their photographs to develop, and they have now become used to creating them in real time, something which is taken for granted. A photograph's creation process was associated with considerable effort and was a demanding craft, not taking into

The next evolutionary stage

Every new stage of photography was and will continue to be associated with enormous changes in various areas of life, be it in our professional or private lives.

Looking back at the past ten years, we can see that the distances between these changes are becoming smaller and smaller, mostly due to shorter technological development cycles and the resulting innovations. It was not long ago that only a few photographers believed that digital photography was the future, or that it could ever reach the same level of quality as analogue photography. Digital photography has since become indispensable, while analogue photography and its associated technologies

account the time required for creating the photo itself. Nowadays anyone can create a high number of photographs in a quick and uncomplicated manner and in the highest quality, while we are also able to share, use and reproduce photos across various channels. This simplicity in the creation and distribution has led to the continuous growth of the number of images created.

A total of 1.2 billion photos were taken across the globe in 2017. To put this into perspective, only half of this amount were taken just five years ago. The volume of image data will continue to grow over the coming years at an average of nine per cent per year. The main reasons for this rapid growth include the fast and continuous expansion of mobile internet with increasingly high bandwidths and speeds as well as the emergence and establishment of social networks and messenger services such as Facebook, Instagram, Snapchat, WhatsApp and many more. The development of smart phones as a primary and ubiquitous camera and therefore the centre of connected digital photography is certainly the most significant factor. Today, approximately 85 per cent of photographs are taken with a smart phone, while just over 10 per cent are taken with a digital camera. It is no wonder then that the saying goes "pics with a smart phone or it didn't happen, pics with a digital camera or it wasn't special enough".

have almost disappeared into irrelevance. Nowadays you can barely tell the difference in quality between a photo that was taken with a smart phone and one that was taken with a professional camera. Today, we are already in an entirely new evolutionary stage, the AI era (Artificial Intelligence). Artificial Intelligence has made its way into the world of photographs where it is already causing far-reaching changes. So-called Machine Learning, a sub-area of Artificial Intelligence, has become a hot topic in many industries, including amongst smart phone manufacturers. The main fields of application of Machine Learning are in image, text and speech recognition as well as intelligent translations. Machine Learning is set to make smart phones even more

intelligent in future, and to autonomously generate knowledge through previous experiences, for example by recognising objects through the camera's lens.

Machine Learning, which is currently mostly cloud-based due to its required computing capacity, will gradually take place on the end devices themselves in future. It is estimated that every fifth smart

phone sold across the globe in the next twelve months will support on-board Machine Learning and will enable these functions even without a network connection. This requires extremely fast local computing capacity and the necessary processors, so-called NPUs (Neural Processing Units), which specialise in autonomous learning and are able to quickly carry out computing operations such as real-time recognition.

The challenge of searching and finding

Photographs can capture moments, elicit feelings and permanently influence purchasing decisions. In fact, images have the power to directly create buying incentives. Buying a product in an online store without seeing a corresponding picture is almost unimaginable. The daily use of smart phones and social media platforms such as Facebook, Instagram and Snapchat is promoting pure communication through images. Photographs and videos have long since become the new universal language, and the saying "a picture is worth a thousand words" has never been truer. If put in the right setting, an image doesn't require any text in order to convey a certain message.

Digitalisation and the associated rapid increase of image material has also led to a challenge in finding or retrieving created photographs and making them useable. Effective image data organisation together with corresponding key words or search terms (also known as tags) is required in order to utilise these digital treasures. This is made more complicated by the fact that images and visual information cannot simply be searched, unlike text. A photograph is a type of black box, comparable to a tin can – the contents of which cannot be identified without an informative label attached to its exterior. In a digital context, this means that conclusions cannot be drawn in relation to the content without its visual examination. The image must therefore be opened in order to identify its actual contents.

At the same time, nowadays there are enormous amounts of visual data, such as on current products,

trends, people and situations which can only be found and understood with great difficulty by marketing professionals. Manually viewing and tagging the entirety of this previously created content is simply not possible and even less feasible from a financial standpoint. This is where Artificial Intelligence is said to help out by automatically recognising contents and tagging photos based on previous experience while even being able to help with searching for the most suitable image.

The extent to which image recognition is already continuously being improved through AI will be mirrored in the search for contents which will develop from simple, text-based search to a more image-based search, including in combination with textual search terms. To illustrate this more clearly, simply imagine the following scenario: You see someone on the street who is wearing the latest design by your favourite sports shoe brand. You have simply got to get hold of a pair of these shoes, so you take a photo on your smart phone. Instead of going through the effort of searching for the model by trying out all possible search terms, you simply upload the photo you took into the search bar of the online shop of your choice, in your favourite shopping app or a search engine before being given opportunity to place an order. This scenario will soon become part of everyday life – Google is already capable of recognising certain consumer products through image recognition using Artificial Intelligence. So-called metadata is required to enable such a search, meaning information pertaining to the content of an image.

How AI is revolutionising image data management

Metadata and tagging

Metadata contains data and descriptions that cannot usually be extracted directly from the primary data – in other words, from the photo. It ultimately serves the purpose of searching, finding, and archiving, as well as supporting the transfer of information. Metadata is often separate from the described physical object when translated into the analogue world, just like an inventory list, as an example. In the digital world however, metadata can be managed alongside the object, such as through a Media Asset Management System, or it can also be stored in the object itself.

Nowadays conventional images normally contain a number of metadata by default, regardless of whether the images were taken with a smart phone or a camera. This tends to be technical information, such as information on the time and location of the photo, or information on the camera with which the photo was taken and its technical settings. But this lacks precisely the type of information which is most important: Information which describes the contents – be it the person, the product or object, the mood, colour, pattern or the text in the image.

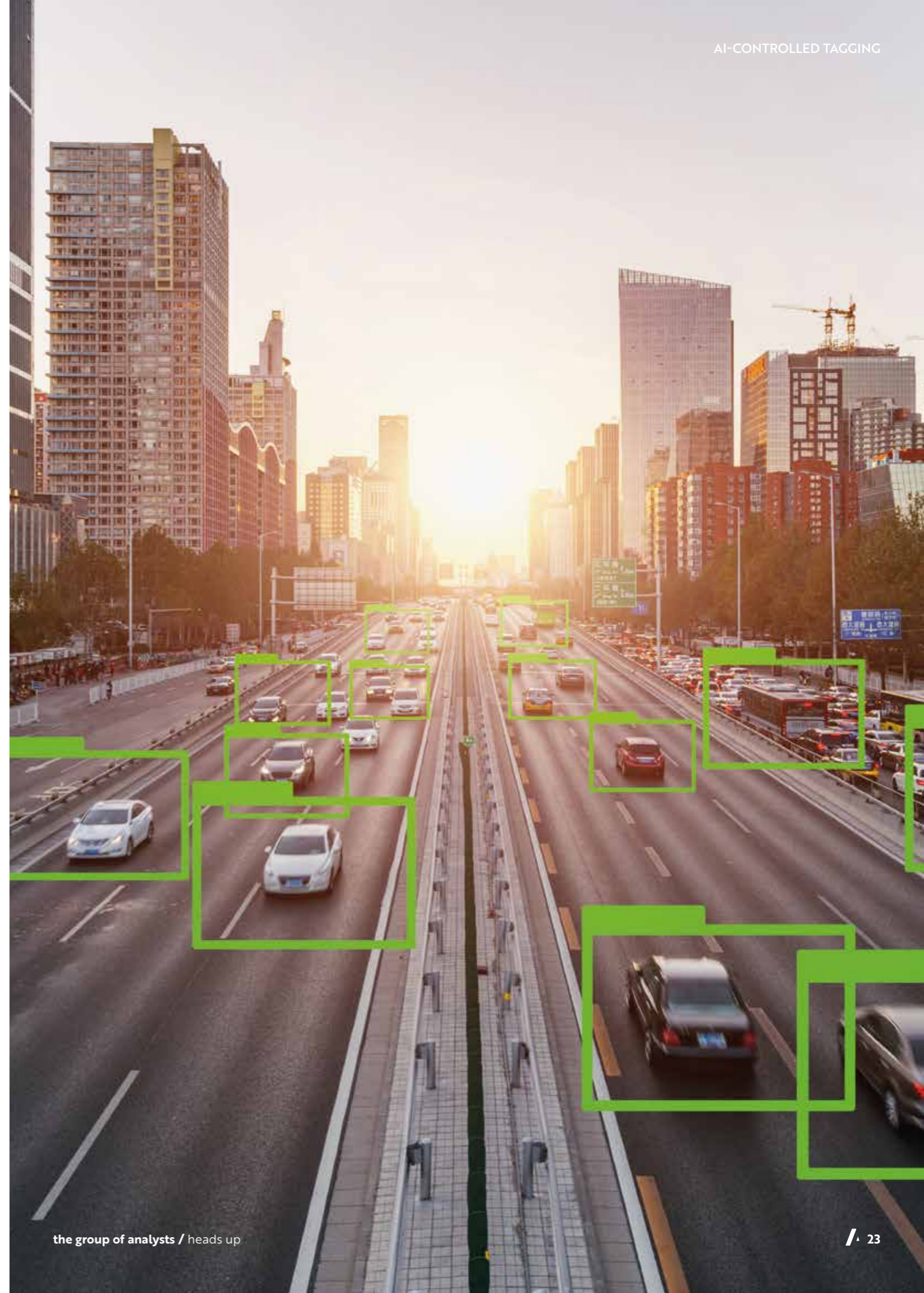
One of the characteristics of metadata is that it consists of structured information which can be read and analysed by machines. So-called metadata standards are supposed to help with the structured storage of important content-related information in an image object or in a database. As well as describing digital objects, these standards support the efficient and useful exchange of information beyond the borders of operating systems.

There is a multitude of metadata standards mostly invisible to the untrained eye which serve the structured capture and storage of (meta)information for images. The EXIF (Exchangeable Image File Format) standard was developed in order to store technical information about a camera shot regardless of the manufacturer, for example. This metadata is automatically generated and barely perceived or changed by the user. Examples of such metadata include the date, time, resolution, orientation and GPS information. Dublin Core or IPTC metadata standards, on the other hand, primarily deal with the description of content-related information, so-called de-

scriptive metadata. As an example, the standard initially developed in 1979 by the International Press Telecommunications Council (IPTC) together with the Newspaper Association of America (NAA) for press photography consists of a list of defined entry fields such as captions, key words (tags) and information relating to the photographer and copyright. Last but not least, XMP (Extensible Metadata Platform) is a further example worth mentioning. The XML-based metadata standard by Adobe is more focused on embedding any number of metadata (such as EXIF or IPTC) into an image or a separate sidecar file in a standardised manner without addressing the content-related standardisation of information.

Anyone who has dealt with image metadata standards in recent years will know that they haven't really changed much or developed further during this period, while the images themselves have meanwhile become increasingly complex. As a result, one single standard is usually insufficient in order to describe a digital object in its entirety.

The IPTC standard is mostly used for describing contents due its widespread use and support by most image editing programmes. The disadvantage is that the fields designed specifically for press photography tend to be unsuitable for specific tagging in companies, so that the applied standards are often used for other purposes (out of necessity), leading to other information being inserted into fields that were predetermined for specific contents. In addition to this, the information is often not inserted in plain text, but rather in a coded version, such as in abbreviated forms, meaning that in general this information is unusable without knowledge of the abbreviation's meaning, thus requiring prior interpretation. A further example includes the storage »





of the name of a person depicted in an image. Although Artificial Intelligence now enables faces to be recognised in an image and to be tagged with a name, there are currently no options of standardising this information and comprehensively displaying it across systems.

The bottom line is that, despite the existence of many schemes for capturing metadata, there is still a lack of a solution that consistently describes and manages an image as a digital cultural asset. As long as this is the case, standards will not be applied at all or will be applied for other purposes, while information will merely be captured in the file name in a typically encrypted form. Yet metadata is indispensable in the practical world for ensuring that digital image objects remain retrievable and comprehensible in future.

Many companies are still a long way off from automatic tagging and therefore still mainly use elaborate file-naming conventions for their images. Their aim is to manually accommodate as much information as possible, such as the item number, the image type or information about the image's content and condition in the file's name, which is often limited to a few characters. It is clear, however, that this information does not disclose much about the actual contents.

Semiautomatic tagging

One method to overcome this issue is by using so-called semiautomatic tagging. The aim of semiautomatic tagging is to automatically recognise previously captured (mostly encrypted) information in the file name or through metadata fields based on rules, to extract this information and to use it for various purposes. One application area where this is especially useful is the automatic linking of an image from the MAM module with an item stored in the PIM module based on one or several item numbers as a result of this extraction. A further example is the transfer of various information from a file name into custom metadata fields, for example for use in attribute-based searches.

The prerequisite for semiautomatic tagging is for the actual metadata to have already been captured. This information is then made accessible to the users on the one hand and to automatisms on the other hand through respective automatic interpretation. Flexible standards for the corresponding storage location in the MAM module guarantee possible applications specific to the image type, and therefore a high degree of flexibility. When it comes to product images, an entirely different variety of information and therefore rules and intervals for the recognition of metadata is required than is needed for background images, to name just one example.





Computer vision

Manual tagging can require a great deal of effort – marketing and PR professionals will be familiar with the task of selecting the best ten per cent out of a vast array of images, such as of an event, before assigning them descriptive key words. While there are many tools available for quickly selecting and tagging images, manual tagging nevertheless tends to require a lot of time and therefore money. One potential risk factor is for the integrity of the tags to be inconsistent. It is often the case that information describing the contents subsequently requires manual completion. Manual tagging is usually an inconceivable option when it comes to large volumes of user-generated content. As a result, the image data often contains too little information to enable it to later be found in a targeted manner.

A new type of technology known as computer vision is guaranteeing support for a maintained image archive with searchable contents. Computer vision is an interdisciplinary field with the aim of achieving a sophisticated understanding of digital images or videos, which are essentially collections of pixels. From an engineering point of view, it is an attempt to automate tasks which the human visual system is capable of, but which are ineffective to carry out by hand due to the considerable amount of time required. Image recognition is not a recent phenomenon, it simply wasn't reliable

enough in the past and industries that require these technologies were unable to sufficiently rely on it. Since Google revolutionised the image recognition scene not long ago with the so-called neural network, AI research has experienced an upswing with the continuous release of new image recognition systems with increasingly sophisticated architecture. Facebook has irrevocably introduced image recognition into the mainstream and has been automatically analysing images in its network for years, for instance in order to identify offensive contents in uploaded pictures. Other tools and applications have followed suit and are integrating image recognition into their products and services. Image recognition can already be found in several consumer sectors today. Some photography apps by major players such as Google or Microsoft are already supporting image organisation through their own image recognition features, for instance. Many of these recognition systems are still in their infancy and are therefore often labelled with the key word 'beta' while undergoing constant further development.

The advantage of this technology is that a large volume of data can be processed in a very short space of time, therefore returning valuable, quantitative and consistent data, even if it has not yet advanced as far as emotion recognition.

Image recognition – models to suit every demand

The functional principle of image recognition is remarkably simple. The starting point is an input in the form of an image or video which is transmitted to an AI service before being analysed. The type of model that is to analyse the photo must also be determined. Models are trained to recognise certain types of objects, such as people, flowers, items of clothing or food. Some AI providers also allow the

combination of various models. This image recognition provides a list of key phrases (also known as tags, key words or concepts) with their corresponding probabilities. The more precise the recognition is, the higher the corresponding probability value will be. This process is also known as prediction.

Prebuilt vs. custom models

There are various types of pretrained models available for image recognition, so-called prebuilt models, depending on the AI service. Even amongst pretrained models there are vast differences in the recognition rates of various manufacturers, meaning that some models are only able to differentiate a few thousand objects, while others are able to tell the difference between over ten thousand general concepts, which can be more than enough for many types of image. The more general the tag requirements are, the higher the added value of pretrained models becomes.

In certain cases it may also be necessary to create individual models with their own concepts. Only a handful of providers have so far enabled the installation and training of so-called custom models directly in the system and without requiring mathematical competence. The advantage of these custom models lies in the fact that they can be tailored toward the specific application, for example if certain key words and classifications are to be used or certain highly specific objects or products

require identification. The training process of a custom model is comparable with the learning process of a toddler. At the start of the learning process, AI is unable to identify which content or which object is in the picture it is asked to analyse. By feeding the AI both positive and negative examples, similar to teaching a toddler the difference between hot and cold, the AI system is able to independently update its own recognition templates (the neural network), therefore increasing its precision in order to successfully identify a certain object. The system learns from itself and becomes more accurate in its recognition in correspondence to the amount of data it is supplied with. Depending on the object or application, fewer than a hundred images may be sufficient for training purposes, while several thousand may be required in more complex cases.

Cognitive services

Companies such as IBM, Microsoft, Google, and Amazon are heavily investing in Artificial Intelligence and Machine Learning and are offering corresponding APIs in order to render these technologies useable. There is also a wide range of specialised AI companies and startups that also offer excellent image recognition solutions. The majority of suppliers provide their AI as cloud services, while a handful offer on-premise solutions, in other words application and licence models. On-premise solutions are useful for the analysis of sensitive legal image data. On the other hand, these solutions require an expert to be present at the company who is able to operate and further develop such a solution themselves. The advantage that cloud-based

services offer is that they can be scaled individually while being highly cost-effective.

The foundations of image recognition and image classification involve returning information concerning the visual content of an image in a certain language. This requires the AI to be capable of recognising the appropriate objects, scenes and activities in the image. There are various types of image classification depending on the suppliers, such as tagging, categorising and captioning. Tagging involves transmitting a list of key words

with their corresponding probability values in a certain language as a result of prediction, which in turn forms the basis of the search. Depending on the AI supplier, a varying amount of tags will be returned, however not all AIs support multilingual tags and many only provide them in English. Returned tags can have a very broad range of meanings (scope), which means that incorrect or nonsensical tags are inevitable, which are also known as false positives. Tags additionally neither have a hierarchical structure, nor do they provide the opportunity to illustrate their correlation.

Categorising sees the AI attempt to categorise the recognised object into a classification system, meaning one or several predefined categories which most fit the image. Categorising is only supported by a few AIs, and each one of these AIs uses its own classification system, as an overarching standard classification system does not yet exist. Captioning involves a caption being returned based on the recognition of a scene or activity. A collection of content tags forms the basis of a caption, which is returned in a full-sentence format that can be read by humans. So-called human-AI hybrid technology involves automatically recognised images being checked by humans who then provide highly precise captions, the provision of which takes ten to twenty times as long as other AIs, however.

There are many other specialised recognition methods alongside the above-mentioned three main cognitive services, such as the recognition of landmarks and points of interest, the recognition of general image attributes, facial recognition, the recognition of unsafe image content (content moderation) as well as text recognition. There are barely any similarities and even fewer standards when it comes to these specialised recognition systems. Every AI service goes its own way when it comes to recognising general image attributes. Several AI engines are capable of recognising certain image attributes, such as dominant co-

lours, black and white, vectors and clip art. Other AI suppliers meanwhile can differentiate between good and bad quality features such as red eyes, good and poor lighting as well as definition and focus. Extracting the colour from images enables the categorisation of images by colour which can in turn be used when searching for images of similar colours.

AI suppliers also take different approaches when it comes to facial recognition and offer specialised models in order to be able to identify human faces in a photograph. The algorithms are being continuously improved in order to also be able to identify faces that are turned away or partially covered. This type of AI tends to provide rectangles for positioning faces in the image. Depending on the AI, it may also provide various facial characteristics such as demographic data, information relating to age or age group, general attributes such as glasses, facial hair and even orientation points, so-called landmarks for certain facial features such as eyelid placement. Some AIs are also integrating the recognition of emotions into their facial recognition technology, which is able to reflect a broad range of emotions for every face depicted in an image, such as anger, contempt, disgust, fear, joy, neutrality, sadness and surprise. These emotions are cross-cultural and are communicated through certain facial expressions across the globe.

Content moderation refers to the machine-aided analysis of unsafe contents, the further use of which is subsequently decided on by a human. This includes the recognition of potentially offensive or undesirable images, contents which may be lewd or inappropriate for minors, as well as for the purpose of verifying individuals. Again, AI services take different approaches when it comes to this type of recognition. Some AIs merely differentiate between safe and unsafe, while others attempt to provide specific key words or classify unsafe contents into predefined categories such as violence, nudity, drugs or racism.

A small number of AIs are also capable of recognising text in an image through so-called Optical Character Recognition (OCR) and to extract these recognised characters into a sequence of letters that can be read by a computer. Fields of application include identifying signs, advertisements and image captions. This form of recognition is also continuously being improved in order to also be able to identify contorted writing and different font styles.

**More than just one solution
is required in the field**

The recognition services of leading AI manufacturers therefore provide very different approaches, while individual company requirements in terms of automated tagging can also differ greatly depending on their objectives. Some image types, such as background images, tend to require as many key words as possible from general models in order to receive a broad scope of search terms, while other image types and company's own product images tend to require more recognition precision provided by a specially-created custom model. The same applies to the location of the recognition. Some images require direct and easy tagging through a cloud service, while sensitive data should be analysed by in-house AI. It could therefore be very useful to combine the most significant AI services with one another, so that the appropriate AI can assume recognition tasks as required. It is also possible to analyse an image through selected or even all recognition services with various models in order to gather as much quantitative metadata as possible.

It is also possible to combine the various tags returned by AI for a comprehensive search and to store them directly in the image in standard metadata fields such as IPTC. This would eliminate the need for time-consuming tagging entirely, leaving more time for what's really important – creating creative assets.



Summary

It is high time that companies consider using the support of image recognition software and AI-controlled tagging, as there are already plenty of opportunities to optimise business processes through AI services, and the number of such specific applications will continue to grow in future. This step appears to be unavoidable on the path towards digitalisation when taking into account the potential this technology has for companies' increased efficiency and innovation.

Augmented Reality, for example, is a type of technology which many different companies are experimenting with. Furniture stores are now offering apps which enable you to project pieces of furniture into a room in order to visualise their size and effect. Augmented Reality can also generate added value within a company, for example for technicians, who can receive fast assistance for operating new machinery through the appropriate equipment. Capturing and identifying objects, structures and products is of vital importance with this technology, and the image or object recognition often initiates a certain function which determines the content to display. Voice search is another area that benefits from more efficient tagging. Thanks to Alexa and Siri, voice searching is becoming more commonplace in our everyday lives, but this is only possible if online contents can be found through suitable key words, possibly in multiple languages. Areas such as medicine are also being revolutionised by deep learning. A certain type of image recognition software was able to make a breakthrough in diagnosing skin cancer and recognised 95 per cent of all melanomas – almost ten per cent more than dermatologists. Netflix uses facial recognition in order to provide a list of the actors displayed in a certain scene. Meanwhile, cosmetics and glasses companies also use facial recognition to allow customer to digitally try out their products through their app.

Retailers with online stores will be those who are primarily unable to provide their customers with contemporary offers without the use of efficient image metadata descriptions. Users are accustomed to precisely filtering results that are relevant to them from an enormous product portfolio, or even directly being able to find and order their desired product through a reverse image search. Efficient image management allows companies to realise enormous potential for themselves, as employees no longer need to spend as much time manually reviewing and tagging copious amounts of images. This provides more room for creativity and quality, which companies can only profit from.

"Media data is the commodity of the future."

Artificial Intelligence (AI) in the context of images and videos is a current trend and key topic alike, one which is being addressed not only by smart phone manufacturers and which will significantly alter many areas of our lives. The way in which we search for something, as an example, will change from classic text search to a visual search. One of the main tasks of an omni-channel marketing system alongside managing media and their metadata includes the targeted search for retrieving content. Classic tagging as it is currently still frequently used will soon be a thing of the past, as it is often inconsistent and incomplete. It will gradually be replaced at a rate consistent to the improvement of image recognition.

Oliver Baum is the General Manager and Director Professional Services of the Pforzheim software company apollon GmbH+Co. KG. The trained media designer, graduate information engineer and certified project manager has had over 20 years of project experience with renowned customers in areas such as Digital Asset Management (DAM), Product Information Management (PIM), Business Process Management, E-Commerce and Digital and Print-Publishing. Through his extensive technical, professional and process knowledge from diverse industries such as mail order business, retail, manufacturing and publishing and his farsighted project vision, the successful implementation of major client projects is always a priority. He also dedicates his time to training new IT professionals as well as continuously developing the Online Media Net product, especially in respect to identifying trends and key technologies early on.

Oliver Baum
General Manager and Director Professional Services of apollon

T +49 7231 941 224
obaum@apollon.de





"The time is now."

Nowadays, nobody has the time to carry out extensive online research when searching for certain products or services. One or two words are typed – or spoken – into a search engine before the most promising result is selected. Results on the second or even third page barely stand a chance of being acknowledged at all. The same applies to image searches, which is why effective tagging should be of the utmost importance to companies.

Temel Kahyaoglu is a board member of The Group of Analysts AG and Chief Analyst for Information Supply Chain Management. He is the publisher and editor-in-chief of the bilingual "Produktkulturmagazin", the publisher of Europe's largest overview of suppliers, "The Book of Analysts" (circulation of 24,000), and has significantly and sustainably helped shape digital transformations in the demand market in over 200 evaluations.

Temel Kahyaoglu
Board member of The Group of Analysts AG

M +49 151 585 526 18
t.kahyaoglu@tgoa.com

"From Artificial Intelligence to Corporate Intelligence."

Deep Learning technologies should be on the agenda of any company that wants to gear up for its digitalisation journey. AI is no longer purely a topic of the future and is instead already revolutionising our everyday lives in a variety of ways. The technology has the potential to take on time-consuming tasks, thereby giving employees better development opportunities. AI therefore contributes to increasing the company's own human capital while also having a positive influence on employee satisfaction through higher quality work.

Carmela Melone has been an analyst at The Group of Analysts since 2018 and is responsible for the project coordination and implementation of the 'heads up' series of the whitepapers. Furthermore, she supports the Produktkulturmagazin editorial team. She has been working in market research with various technological focuses for more than five years now and has written a whole series of studies and specialist articles.

Carmela Melone
Analyst at The Group of Analysts AG

T +49 7532 494 35-13
c.melone@tgoa.com



Imprint

Publisher

Temel Kahyaoglu t.kahyaoglu@tgoa.com

Scientist-in-chief

Mario Bretzler m.bretzler@tgoa.com

Senior Analyst

Jürgen Burger j.burger@tgoa.com

Supply market

Andrea Gorova a.gorova@tgoa.com

Project team

Carmela Melone c.melone@tgoa.com

Layout & Design

Sabine Fuss
Carmela Melone

Proofreading

Hilke Bemm
Johanna Späth

Translation

Steve Britt
Nora Manthey
Christopher Mayhew
Johanna Späth

Our bureaus

Frankfurt Bureau: Anja Fahs, Dietmar Rieger

Berlin Bureau: Dr Claudia Pelzer

Paris Bureau: Patricia Philbin

Amsterdam Bureau: Arne Biesma, Jeroen Junte

Rome Bureau: Dr Stefan Pollak

London Bureau: Nora Manthey

Silicon Valley Bureau: Matthias Hohensee

San Francisco Bureau: Nicole Kidd

Online & Social Media

Katharina Kull

Publishing house

The Group of Analysts AG
Uferpromenade 5
88709 Meersburg, Germany

www.tgoa.com · info@tgoa.com

Picture credits

Cover: Prettysleep1/Pixabay; P. 4-5 Max Langelott/Unsplash; P. 8-9 Alexander Köpke/EyeEm/Getty Images; P. 18-19 Westend61/Getty Images; P. 23 Dong Wenjie/Getty Images; P. 24-25 Michael Zwahlen / EyeEm/Getty Images; P. 26 Saketh Garuda/Unsplash; P. 31 Vstock/Getty Images; P. 35 Oliver Baum ; P. 36 Temel Kahyaoglu; P. 39 Carmela Melone



The Group of Analysts AG
Uferpromenade 5
88709 Meersburg
www.tgoa.com