Deep Analysis

Analyst Brief

A Guide to Cognitive Capture for Buyers and Implementors

By: Alan Pelz-Sharpe

The Essentials

Document capture has been around for decades, but recent advances have breathed new life into this market. The ability to liberate documents from their physical form and move them into the digital realm has long yielded benefits in productivity and security while helping to reduce costs. Recently, advances in artificial intelligence (AI) and machine learning (ML) combined with relatively mature optical character recognition (OCR) technology have made it possible to liberate not only the document but also the embedded business information required to drive business operations. We call this process cognitive capture, and we believe it opens new possibilities to reimagine traditional workflows.

The entry of heavyweight cloud services players such as Google, Microsoft, and Amazon, providing a variety of AI/ML and OCR-based services, has reinvigorated the document capture market (see Figure 1). The widespread availability of these commoditized services is resulting in a renewed interest in not only capturing documents but also understanding and acting on the business information that resides within them. The simultaneous rise of robotic process automation (RPA) and its unquenchable appetite for structured business information has further heightened the interest in exploiting this newer ability in order to satisfy this demand.

Before investing in a cognitive capture system, it's important to understand the basics of how they work and the challenges they present in different sectors. This Brief is your guide to cognitive capture and has been written specifically to give guidance to potential buyers and implementors of cognitive capture systems.

What is Cognitive Capture?

Cognitive capture is, like any other enterprise software category, an umbrella term for a lot of different tools that share some similarities. All cognitive capture tools have two things in common:

- Automated recognition of text characters
- Use of AI (neural networks) for content processing

What really unites cognitive capture tools is their use of AI (neural networks) to process data and information. Neural networks are used to decipher characters, words, and spaces. More importantly, they are also used to detect inherent structures and infer relevance and context. They may also be used

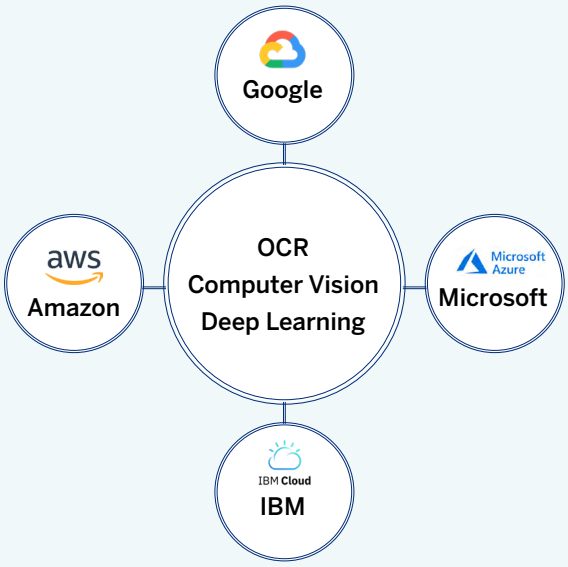
to undertake semantic analysis of content to gain an understanding of the inherent meaning. And, increasingly often, neural networks may be used to translate content from multiple languages so it can be processed and actioned together (see Figure 2).

In many cognitive capture systems, the AI capabilities are optimized to meet specific use cases, such as understanding financial analysis reports in an inbox or reading and understanding property leases. In other cases, they may be optimized to meet more general enterprise purposes such as processing invoices, statements, and purchase orders. Some cognitive capture systems run in the cloud and access vast volumes of data for learning and processing, while others run on premises and leverage small, curated data sets. The differences between tools can be wide, and as such, a great deal of caution should be undertaken when selecting the right cognitive capture product to meet your specific needs. One may be more "sophisticated" technically than another and therefore appealing, but in practice a poor fit for your needs.

The Current Situation

Though traditional technologies continue to dominate the document capture market, things are changing via the use of machine learning (ML) and AI. The unrelenting growth in the volume, diversity, and structural complexity of documents demands new approaches, and cognitive capture can meet these ever-evolving needs. The concepts behind cognitive capture are much wider than traditional document capture; rather than a standalone solution, cognitive capture is instead the starting point for, among other things, generating and managing clean and accurate data in extended ML and intelligent automation business processes.

Figure 1
Heavy-Hitters Providing AI/ML and OCR-Based Services

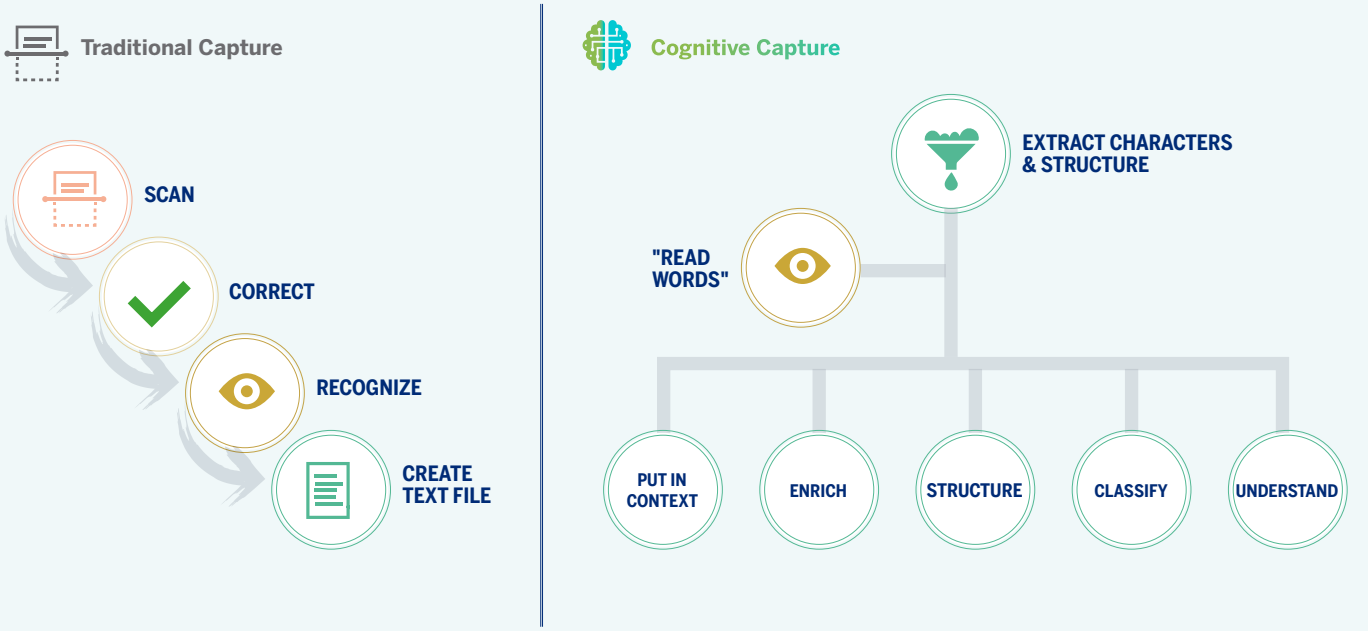


What's in a Name?

Throughout this report we use the terms "traditional capture" and "cognitive capture" to differentiate between older and newer approaches and technologies. While "traditional capture" is universally used for the former, other firms use terms such as capture 2.0, intelligent document processing, and next generation capture to describe the latter. We see these terms as interchangeable as they all address the same topic. We opted for the term "cognitive capture" simply because it sounds cool and ably expresses the intelligence, albeit artificial, that is inherent in these newer approaches. You should feel free to use whatever terminology you prefer.

For background on the evolution of document capture from traditional to cognitive capture, see our Brief titled "The Dawn of Cognitive Capture."

Figure 2
Neural Networks Boost Cognitive Capture Capabilities



An illustrative example is found in the insurance and healthcare sectors, which have been among the earliest to explore the use of cognitive capture as it has the potential to help in these areas:

- Reduce the average handling time of complicated customer and patient inquiries
- Reduce the time and cost to ingest large volumes of disparate and complex content
- Integrate and bring new life to legacy business applications
- Reduce or eliminate repetitive human-based activities
- Reduce the overhead costs of regulatory compliance
- Dramatically increase the percentage of straight-through processing (without human intervention) of claims and inquiries

Although every organization is a little different, such benefits of using cognitive capture are usually focused on reducing manual and repetitive activities (for example, human key entry work). Though that focus is natural enough, there are more ways that cognitive capture can provide benefits through opening new possibilities to reimagine traditional workflows.

Though a focus on reducing manual and repetitive activities is natural enough, there are more ways that cognitive capture can provide benefits through opening new possibilities to reimagine traditional workflows.

To drill a little deeper, here are two real-world examples.

1. Firm A, a manufacturer, has to process 5,000 invoices a month. Those invoices come from more than 800 suppliers, and despite the firm's best efforts they come in a variety of formats with different levels of consistency and accuracy. Today, the firm scans and captures those invoices to digitize them, and it uses techniques such as zonal OCR to automatically extract invoice and supplier numbers, addresses, invoice totals, etc. However, the exception rate – the number of processed invoices that need further checking and manual key entry work to be accepted – is around 45%. Through the use of cognitive capture technology that can more accurately read and extract relevant data using ML & AI, the exception rate drops to under 7% (see Figure 3).

   That improvement alone justifies the cost of the system, and that is only the start of the benefits for the firm. The cognitive capture system can also automatically associate almost all the captured invoices with the relevant purchase orders and delivery notes in the firm's ERP system, eliminating a further swathe of human involvement. In this basic example, cognitive capture technology reduced invoice processing times from approximately 4 days to just a couple of hours and halved the employee costs.

2. Firm B, a healthcare provider, illustrates a more ambitious real-world use of cognitive capture. The use of AI by healthcare providers is not new, but to date much of the focus has been on analyzing clinical data. Recently, attention has begun to shift toward resolving the endemic issues of medical health records (MHR). At one level,

Figure 3

Cognitive Capture Advantages for Invoice Processing

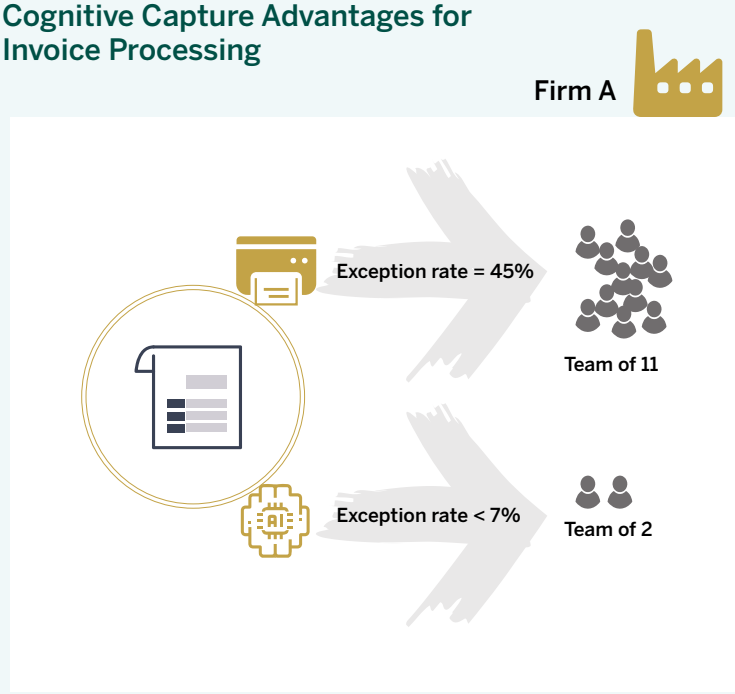
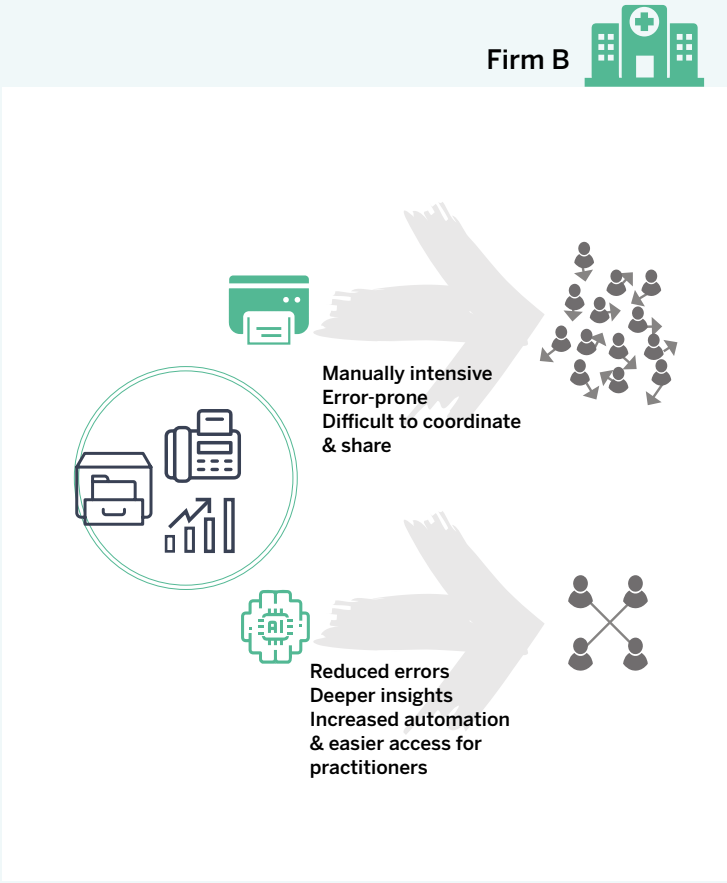


Figure 4

Cognitive Capture Advantages for Healthcare Records



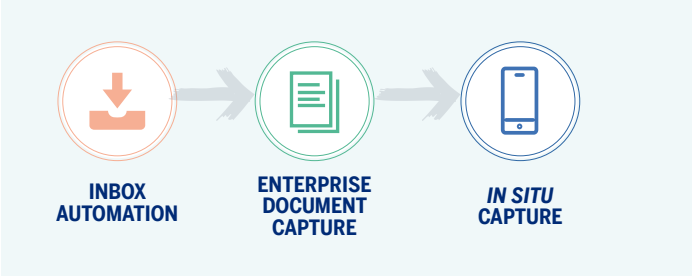
the benefits of using cognitive capture are the same as elsewhere, for example the ability to ingest multiple document and data formats, from scanned documents to faxes and beyond, and to do so accurately and quickly while extracting key data for further processing. MHR work today is hampered by interoperability issues, is manually intensive, and is worryingly prone to error. So, implementing cognitive capture can bring clear benefits, but the ingestion, processing, reduction in errors, and, importantly, a machine-level understanding of the relevance of the information captured can trigger much more profound improvements. The use of ML within cognitive capture systems can provide a firm foundation for a truly holistic and standardized overview of a specific patient and their needs, resulting in improved visibility and coordination of patient care. It can also provide the data foundation needed for broader programmatic patient and practice research, while in the process eliminating duplicate data, identifying missing or incorrect data, and ultimately improving patient care and overall clinical efficiencies (see Figure 4).

On the horizon are even more ambitious use cases such as, for example, integrating cognitive capture software into cameras in drones for industrial maintenance and law enforcement. But even in both the above examples, invoice processing and EHR management, the challenges are not at the technology level but in changing existing workflows, gaining buy-in and cooperation across different operating units, and building a business case to convince those holding the purse strings that this is a priority.

Breaking Down Cognitive Capture

A simple set of Google searches reveals hundreds of technology vendors listed under the various terms that are used for cognitive capture: RPA, Intelligent Automation, Digital Workflow, Intelligent Capture, Intelligent Document Processing, Capture 2.0, etc. On the surface they all may appear to do much the same thing, and they all leverage machine learning and/or artificial intelligence. The differences come in the form of the underlying technology used, the way in which it is architected, and, importantly, in the use cases they have been trained and optimized to meet.

To provide some market context, in this section we compare and contrast three cognitive capture systems that are available in today's marketplace: the first niche, the second broad, and the third mobile/*in situ*.



Inbox automation

Alkymi Data Inbox is a product targeted at financial services firms. It captures all the user's incoming emails, documents, and attachments in one place, usually in the form of text, tables, charts, PDFs, and XML. Data Inbox can automatically identify, mine, and extract from these documents the types of complex, tabular data common in the financial services arena. Cognitive capture systems typically classify high-volume, complex documents

into one of many categories and then search for data elements based on each document's classification. Alkymi takes a different approach. Rather than putting each document into an individual bucket, the system places each detected element into multiple relevant categories. If a document contains an income statement, for example, the income statement may be relevant to both the financial statement and equity research categories, so it would be added to both.

Enterprise document capture

IBM Automation Document Processing is an AI-underpinned, pre-configured, enterprise-grade, cloud software solution. Importantly, the capture activities in ADP provide a foundation for end-to-end document processing. Ultimately, processing documents straight through their lifecycle is the goal of any document capture, but in reality, most document capture products stop at the capture phase and involve clunky and often manual handoffs. ADP, though, has been built to feed into broader lifecycle document activities, particularly when used in conjunction with IBM FileNet Content Manager and Enterprise Records Manager through the available integrations. Indeed, these content repositories greatly benefit from the increased visibility of the information contained within the documents processed and subsequently stored on them, making it possible to automate enterprise search, auditability, security, and governance. Similarly, it is possible to integrate ADP with existing workflows and process management systems and to feed RPA systems automatically.

In situ capture

Anyline is a Europe-based firm that provides OCR functionality that can be embedded into other applications. On the surface it doesn't seem to fit our definition of cognitive capture, as it is more a component than a complete solution, but a deeper dive shows some remarkable innovation far outside of the traditional capture technology market. In essence, what Anyline builds are neural network-based enterprise applications to run inside smartphones and tablets. These applications are designed to meet the most extreme of capture situations, for example reading the TIN number on a car's tire. Imagine black embossed rubber characters, likely dirty, on the same black rubber surface. A device using the Anyline app can capture this accurately and immediately, running a specific series of neural networks (not computer vision), and send the results back to wherever they are needed. Capture in such extreme circumstances opens up a wide range of *in situ* use cases.

These are just three examples of very different products designed to do very different things. Alkymi is a cognitive capture tool designed to meet a critical but niche problem affecting email inboxes in financial services. IBM makes bigger, more ambitious efforts to tackle much broader capture concerns. Anyline is a highly specialized firm that embeds neural networks on devices to capture the previously uncapturable.

Rather than putting each document into an individual bucket, Alkymi's system places each detected element into multiple relevant categories.

Though these are three extremes, the fact is that every one of the hundreds of other cognitive capture systems available will be different, even though they share some generic commonalities. Some are long established in the document capture market and making a transition to more modern ML/AI methods, while others such as Nanonets (low code), the aforementioned Alkymi, Haystack, and Leverton (contract analysis) are start-ups coming to this space afresh.

As cognitive capture systems make use of AI & ML, by default this means that they also make use of pre-built models and algorithms which are typically compiled and managed within a library service. This is where the biggest differences between one cognitive capture system and another are typically found. Though all will likely run OCR in a similar fashion, after the initial character and word capture they will diverge radically. Some systems will be optimized to meet specific healthcare industry, insurance, supply chain, energy, legal, etc. needs. This level of specialization will vary; for example, one invoice, regardless of the industry, will be similar to another. But the specific requirements of analyzing clauses in a New York commercial lease, for example, will differ from those of another state or, for that matter, a residential lease in the same state. The important thing to note here is that, although the underlying technology is important, often more important is the supplier's knowledge and pre-built library of tools for your industry.

Challenges

We have stated that no two cognitive capture systems will ever be exactly the same, but the question arises, what if you are using a common cloud-based system from Amazon, Google, IBM, or Microsoft? The answer is, though you will be using the same cloud-based system that has been trained on some common data, you will be training this pre-configured technology further on your own data. Though generic tasks such as recognizing an invoice or, for that matter, recognizing characters or text will be a shared identical service, you will optimize and further train the system with your own data to meet your specific enterprise requirements.

Supervising cognitive capture

All AI systems need to be supervised; they are not perfect, and they will make mistakes. The problem is that a mistake that goes unchallenged or uncorrected will compound the error. AI is constantly learning, and it does not inherently know the difference between right and wrong. If it makes an error, you need to flag that error and take corrective action as soon as possible. In practice this means that cognitive capture, just like any other AI system, needs to be constantly supervised by human subject matter experts. This process is referred to as "Human in the Loop" (HITL) and requires a capability within the system to observe and correct, along with provisioning people to do this work on a continuous or regular basis. In short, you cannot switch the system on and

Though generic tasks such as recognizing an invoice or, for that matter, recognizing characters or text will be a shared identical service, you will optimize and further train the system with your own data to meet your specific enterprise requirements.

walk away – you need to actively supervise and guide it through its learning process. Any cognitive capture project will need to plan and budget for this work.

Transparency and explainability

AI systems are very complex, to the point that some forms of AI are unexplainable to a human being. It is literally impossible to know what process the system followed to come to a decision. To repeat, it is literally impossible, not just extremely difficult. AI systems that leverage deep learning, for example, and complex arrays of neural networks cannot be explained. Many other forms of ML & AI can be explained and have been engineered to provide transparency and human visibility into the decision-making process. Most cognitive capture systems have been architected to ensure that there is transparency into critical decision-making capabilities, but not all of them, and this aspect needs to be explored carefully before deciding on using any one system within your organization. Take, for example, a cognitive capture system that plays a significant role in approving or denying a loan application. If there is a dispute regarding the loan application, can you go back and explain why this decision was made? If not, you could be in trouble.

Compliance

AI systems feed ravenously and process prodigious amounts of data, but not all data is equal. Some data is highly sensitive and/or personal and needs to be strictly controlled. Depending on your specific use case, you will need to make sure that any cognitive capture system meets your compliance, privacy, security, and regulatory needs.

Again, the companies that build cognitive capture systems typically understand these requirements and build in the necessary controls and means to redact, share securely, or remove sensitive data when required. But not all systems do this the same way, and some do not do it at all.

Language limitations

One of the biggest limitations of most cognitive capture systems is their relative weakness in languages other than English. On the surface, many cognitive capture systems claim to be able to translate and work in a wide range of languages, from Mandarin to French. To some degree they can, but there is a catch: they typically understand English far better than they do other languages. In fact, there are many languages that they do not understand at all. Bear in mind there are over 7,000 languages in the world, and almost 4,000 of them are written. As of today, Chinese (Mandarin/ Cantonese) is the most widely spoken language, followed by Spanish, English, Hindi, and Arabic. Capture systems tend to focus on English, Spanish, French, and German. There is good reason for that, as the characters are largely from the same alphabet and that makes it easier for the underlying OCR; also, these are the markets they sell to. Conversely, it makes sense that an OCR built in China will be optimized very differently than one built in the US. But the challenges are not simply in the character sets used, be they Roman, Greek, Korean, or Hebrew. The challenges run much deeper as different languages are structured and phrased differently, not to mention fluid and ever-changing.

Call to Action

If you want to use cognitive capture systems as part of a digital transformation, only a subset of your team needs to understand the technicalities and underlying architectures. But everyone on your team should have an understanding of the strengths, weaknesses, challenges, and opportunities that a cognitive capture system brings. It's a rare case of a little knowledge going a long way. In other words, your non-technical staff must lead the initial selection of any cognitive capture technologies; they will need to undertake a lot of solid pre-analysis of your existing business situation, tasks, and processes and have a clear understanding in advance of what you want to achieve.

Until you have a solid understanding of your needs and limitations, you cannot start the product selection process. Once you do start that process, you need to narrow down and thoroughly test any possible suitors before even thinking of making a final selection or kicking off an implementation project. This is practical advice that we advise you to follow closely, as skipping ahead never works out well. Regardless, cognitive capture technologies are set to ignite change in the market over the next five years, transforming dull, tedious, and costly activities and providing the accelerant for true digital transformations.

Over the last few years, automating business processes has become critical for many organizations across public and private sectors, and this has been further accelerated by the pandemic's disruption of traditional working environments. But automation takes time to do well, and though RPA tools have proved useful, without an accurate stream of data they can at times be more trouble than they are worth. Cognitive capture and RPA are best buddies; their success feeds off one another. As such, we can expect to see these two fields move closer together and truly complement one another in the future. That's interesting and is an area we will watch closely at Deep Analysis. However, though further down the line, emerging use cases hold immense promise, and the start-up world and its investment community are particularly active here.

At the end of the day, though, for all the analyst firms' efforts to divide and categorize technologies into discrete and fiercely separated silos, at a technology level it's all about enterprise AI moving into the mainstream, and from a business perspective it's all about automation. At Deep Analysis we see RPA, cognitive capture, process analysis, and even blockchain as part and parcel of the same thing. Ultimately, the goal is finding a solution to your existing business problems and finding new approaches to build new business opportunities. No one technology will do that for you; it's going to require a patchwork quilt of technology tools, and cognitive capture will play an increasingly important role.

About Deep Analysis

Deep Analysis is an advisory firm that helps organizations understand and address the challenges of innovative and disruptive technologies in the enterprise software marketplace.

Its work is built on decades of experience in advising and consulting to global technology firms large and small, from SAP, Oracle, and HP to countless start-ups.

Led by Alan Pelz-Sharpe, the firm focuses on Information Management and the business application of Cloud, Artificial Intelligence, and Blockchain. Deep Analysis recently published the book "Practical Artificial Intelligence: An Enterprise Playbook," co-authored by Alan and Kashyap Kompella, outlining strategies for organizations to avoid pitfalls and successfully deploy AI.

Deep Analysis works with technology vendors to improve their understanding and provide actionable guidance on current and future market opportunities.

Yet, unlike traditional analyst firms, Deep Analysis takes a buyer-centric approach to its research and understands real-world buyer and market needs versus the "echo chamber" of the technology industry.

About the Author

Alan Pelz-Sharpe is the founder of Deep Analysis. He has over 25 years of experience in the IT industry, working with a wide variety of end-user organizations like FedEx, The Mayo Clinic, and Allstate, and vendors ranging from Oracle and IBM to start-ups around the world. Alan was formerly a Partner at The Real Story Group, Consulting Director at Indian Services firm Wipro, Research Director at 451, and VP for North America at industry analyst firm Ovum. He is regularly quoted in the press, including the *Wall Street Journal* and *The Guardian*, and has appeared on the BBC, CNBC, and ABC as an expert guest.

Contact us:

info@deep-analysis.net

+1 978 877 7915

