



2021 State of AI and Technology Adoption in eDiscovery

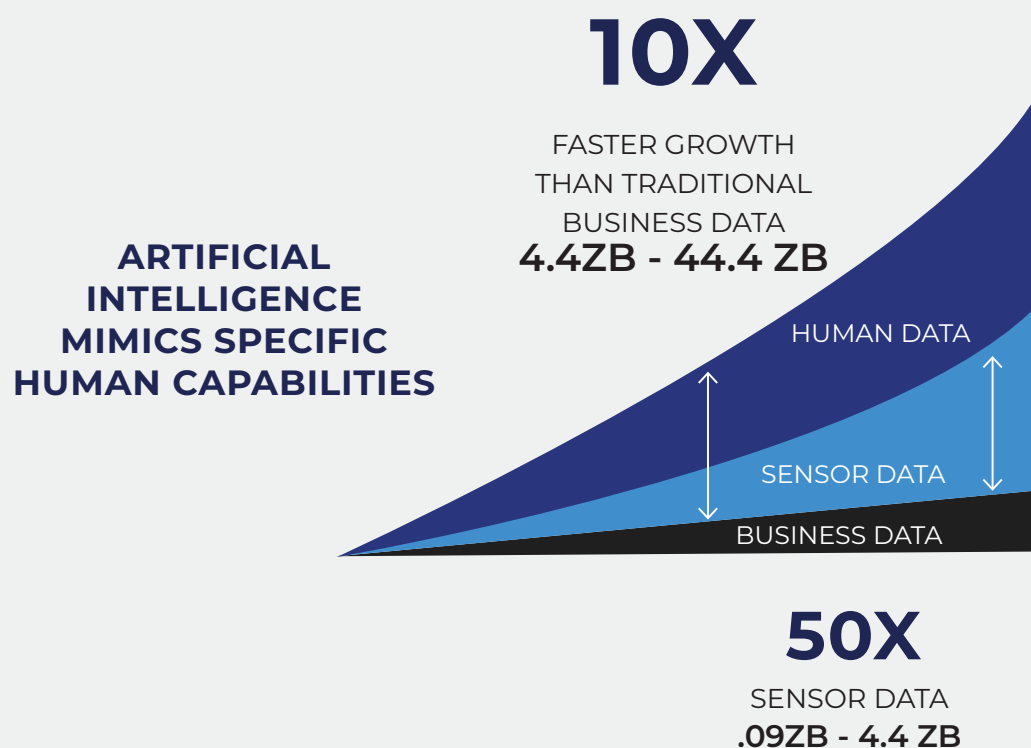
Contents

- 4 | What are we including under the definition of AI and other technologies?
- 5 | How do we measure the adoption of AI?
- 6 | Results: The adoption of AI and other technologies in eDiscovery
- 11 | Overall adoption and summary of results
- 12 | The eDiscovery hype cycle
- 13 | Perceptions of AI among eDiscovery practitioners
- 14 | AI applications across eDiscovery use cases
- 16 | The future of the legal profession and the value of AI for eDiscovery
- 16 | Moving the focus of eDiscovery upstream
- 17 | About the authors
- 18 | About IPRO, ZyLAB, and ACEDS



The State of AI and Technology Adoption in eDiscovery

Every year, businesses generate a massive amount of data—and organizations must manage that ever-growing data store through the eDiscovery process. As the COVID-19 pandemic has further driven the use of digital collaboration tools such as Slack and Microsoft Teams, the growth of data volumes has accelerated and the complexity of eDiscovery data has increased. eDiscovery professionals today must organize and sift through an overwhelming amount of unstructured data in a variety of different formats such as emails, chats, documents, logs, audio, video, images, emojis, and text, looking for the evidentiary “gold” hidden within.



Statistics from insideBIGDATA; see full report at <https://insidebigdata.com/2017/02/16/the-exponential-growth-of-data/>

But eDiscovery professionals don't have to go it alone, thanks to a host of new tools and technologies that can help streamline and simplify each stage of eDiscovery. The question is, how well are those potentially helpful tools being adopted? How are legal professionals dealing with the growing complexity of data today?

To better understand the state of artificial intelligence (AI) and technology adoption in eDiscovery, ZyLAB—an IPRO company—and the Association

of E-Discovery Specialists (ACEDS) recently surveyed eDiscovery practitioners about their use of technology. The 2021 State of AI in eDiscovery survey asked 21 questions on the use and perception of AI and other technologies and collected responses from 184 experienced eDiscovery practitioners.

As a baseline, our respondents overwhelmingly agreed (95%) that both data volumes and the complexity of eDiscovery data are growing every year. Further, 89%

of respondents agreed that this trend leads to an increase in the amount of eDiscovery work to be done. Just over three quarters (77%) stated that they believe the risks posed by eDiscovery are also on the rise.

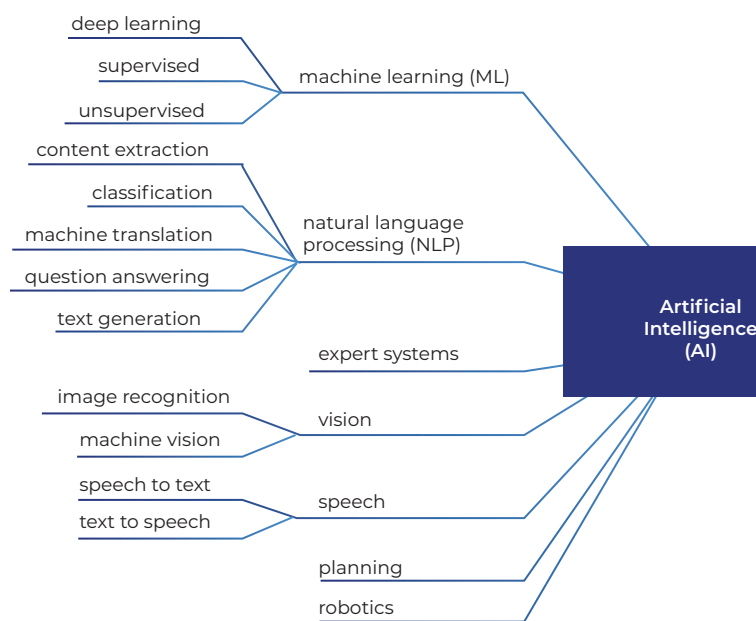
This white paper reports our results and recommends next steps for how eDiscovery leaders can advance the use of AI and other technologies in eDiscovery.

What are we including under the definition of AI and other technologies?

AI is implicated in processes whereby machines learn from experience, adjust to new inputs, and perform tasks that require human-like intelligence or “thought.” AI applications are currently learning how to process visual inputs, speech, language, and more.

But while AI is a persistent buzzword in the legal technology field, not all helpful technologies involve its application. Simple automation, for example, lacks the panache of machine learning, but it offers tremendous advantages throughout the eDiscovery pipeline.

Clearly, then, not all technology used in eDiscovery is AI. Further, not all forms of AI (see image) are used in eDiscovery or generally in the practice of law.



How do we measure the adoption of AI?

While we'd love to poke around in our respondents' tech stacks to see what tools they have and which are most used, for the purposes of this survey we simply asked respondents to self-report their estimated use of a variety of technologies and advanced data analytics. Note that we did not ask about every tactic or approach that a respondent had ever used; rather, we focused on those tools that practitioners most commonly use, so that we could measure the penetration of various technologies into everyday use.

We sorted the technologies into four categories based on the value that they provide, and then measured how many respondents indicated that they commonly use a technology in that category. Across the four categories, we have characterized their overall level of adoption into four levels: early adoption, somewhat adopted, adopted, and widely adopted.

TO BETTER UNDERSTAND THE ADOPTION OF AI AND OTHER TECHNOLOGIES IN EDISCOVERY, WE'VE GROUPED THOSE TOOLS THAT MAY BE HELPFUL TO EDISCOVERY INTO FOUR DISTINCT CATEGORIES:

- 1.** Automation tools that simplify and streamline repetitive, often low-value work, including deNISTing, deduplication, data processing, optical character recognition (OCR), and persistent highlighting.
- 2.** Context tools that parse the content and context of eDiscovery data, such as email threading, entity search, basic entity extraction, foreign language extraction, and translation.
- 3.** Proactive intelligence applications that support legal work and produce consistent quality, including technology-assisted review (TAR) and concept clustering.
- 4.** Emerging intelligence approaches that reveal the story behind data, unlocking new ways of working and allowing legal professionals to focus on high-value work, such as sentiment analysis, pattern analysis, and auto-redaction of sensitive data.

Results: The adoption of AI and other technologies in eDiscovery

AUTOMATION TOOLS: WIDELY ADOPTED (FOR SOME APPLICATIONS) Tools that simplify and streamline repetitive, often low-value work

In our first category are relatively simple technologies that use automation, rather than AI, to increase efficiency. Several of these tools were widely adopted, including deNISTing and deduplication (87% adoption) and data processing (74%). Others, however, were surprisingly low, such as direct cloud collection, which only 17% of respondents stated they commonly used.

One caveat may explain this result: most of our respondents reported that they work primarily in the review and production stages of eDiscovery, rather than in the early phases such as collection and data processing. Therefore, our results are likely skewed toward the right side of the Electronic Discovery Reference Model (EDRM), which would artificially suppress responses for left-side tasks that are commonly aided by automation tools.



MAIN CAPABILITIES THAT HAVE BEEN ADOPTED:

DeNISTing, deduplication, and hashing:

Identifying duplicates and system files and removing them from data.

Data processing: Converting unstructured data into reviewable data, such as by unpacking embedded objects.

Image recognition through optical character recognition (OCR): Recognizing text within images and extracting it into a searchable text form. (Note that some eDiscovery solutions are now conducting OCR through the use of AI models that “see” text and translate it using language models to improve results.)

OPPORTUNITIES FOR FURTHER IMPROVEMENT:

Audio and video transcription: Automating the transcription of spoken language through speech-to-text processing. While audio transcription is commonly used in interviews or meetings when a court reporter is not present, it is not yet regularly used to transcribe text from video calls or webinars, despite the huge uptick in Zoom calls during the pandemic.

Recommendation: Save valuable review hours by adopting transcription tools that return searchable written text for any discoverable audio or video content.

Direct cloud collection: Connecting to different data sources, such as Outlook, Office 365, Microsoft Teams, Google Suite, or Slack to automatically collect potentially relevant data.

Recommendation: Implement a connector within your eDiscovery platform that will allow you to automatically collect data from all of the various applications used within your organization. If you have in-place search functions that allow keyword searching prior to collection or processing, you can reduce processing and hosting fees as well as review costs.

Image classification: Automating the classification of images to group images by type or topic. This helps uncover more relevant images in search and can save valuable time in review.

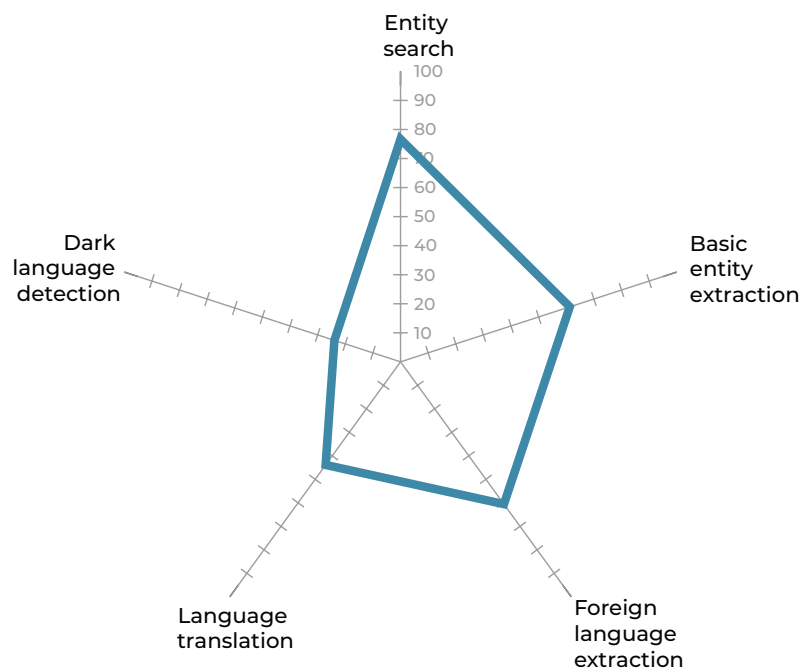
Recommendation: If your dataset contains many images, use an image classifier to save valuable time.

2 CONTEXT TOOLS: SOMEWHAT ADOPTED

Tools that parse the content and context of ediscovery data

The second category includes basic data analytics capabilities that are standard in most eDiscovery platforms. As a result, these technologies are somewhat well adopted, with entity search (77%), basic entity extraction (61%), and foreign language extraction (61%) most commonly used. Most language tools in this category, however, are considerably underused.

These results illustrate that eDiscovery professionals have invested in capabilities to help them understand the content of their legal data, but still lack confidence with techniques that could help them detect language and context clues.



MAIN CAPABILITIES THAT HAVE BEEN ADOPTED:

Entity search: Searching for concepts such as persons, organizations, or localities rather than relying on specific identified keywords. Search engines may accomplish entity searching through latent similarity indexing and analysis or through more sophisticated natural language processing (NLP) approaches that rely on unsupervised machine learning.

Basic entity extraction: Extracting concepts or entities from data to better navigate a dataset by grouping all like concepts together. For example, named entity recognition can identify key concepts such as names, dates, locations, organizations, and more, aiding in early case assessment and matter scoping. Basic entity extraction can also encompass sensitive information such as personally identifiable information or personal health information. Basic entity extraction models may be trained using deep machine learning, artificial neural networks, or supervised machine learning.

Foreign language extraction: Detecting and identifying foreign languages within a document.

OPPORTUNITIES FOR FURTHER IMPROVEMENT:

Language translation: Using NLP to interpret and translate foreign languages wherever they may appear, whether in portions of single documents or entire large datasets.

Recommendation: Expand the scope of your searches and reduce the risk of missing important information by translating foreign language text within your eDiscovery corpus.

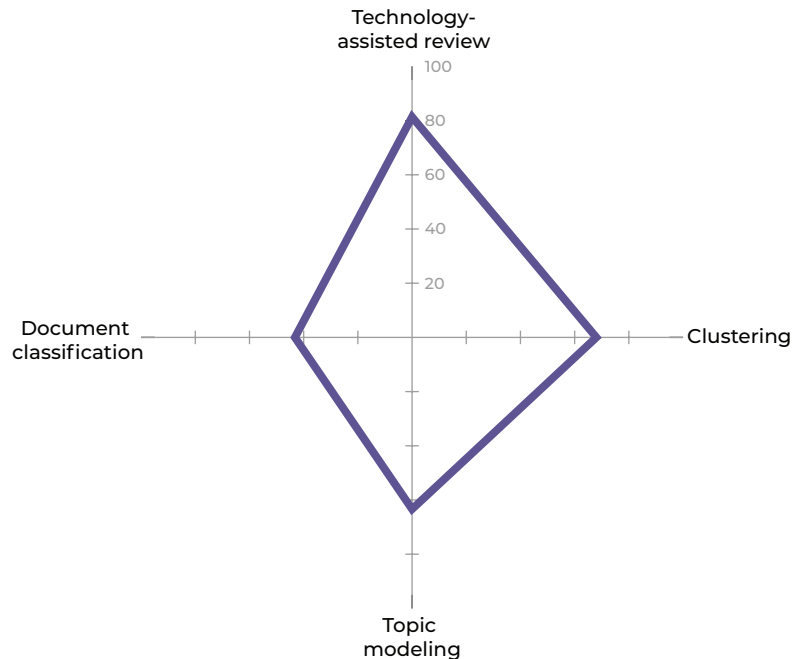
Dark language detection: Identifying key terms that might indicate profanity, code words, harassment or discrimination, and legal concepts such as liability or force majeure.

Recommendation: Use dark language detection tools to identify those custodians and contexts associated with dark language so that you can quickly and accurately detect additional areas of concern and identify further avenues for investigation.

3 PROACTIVE INTELLIGENCE: ADOPTED

Applications that support legal work and produce consistent quality

The techniques we grouped under the heading of proactive intelligence are undoubtedly complex, requiring considerable time and effort to master. However, because these tools improve results dramatically while saving significant time, effort, and cost, they have been widely adopted. Fully four fifths of respondents (81%) reported regularly using technology-assisted review (TAR) to manage their matters. About two-thirds of respondents stated that they used the related techniques of topic modeling (63%) and clustering (68%) to group related concepts together.



MAIN CAPABILITIES THAT HAVE BEEN ADOPTED:

Technology-assisted review (TAR) 1.0:

Accelerating the review process by teaching an algorithm which documents are likely to be tagged as relevant, privileged, or otherwise important. TAR approaches—which may also be known as predictive coding and recently active learning—use a supervised machine learning protocol. Although there are multiple versions of TAR, it must be validated by human quality control measures to ensure that the learning process was free of gaps that could cause the algorithm to miss critical documents.

Topic modeling and clustering:

Grouping topics together within the corpus of documents based on unsupervised machine learning that identifies related concepts within or across documents.

OPPORTUNITIES FOR FURTHER IMPROVEMENT:

Document classification:

Identifying and classifying documents by type, responsiveness, privilege, or protected information content—among other traits—without opening each individual document. Document classification uses NLP classification and supervised machine learning to sort agreements, invoices, correspondence, or other document types into virtual piles.

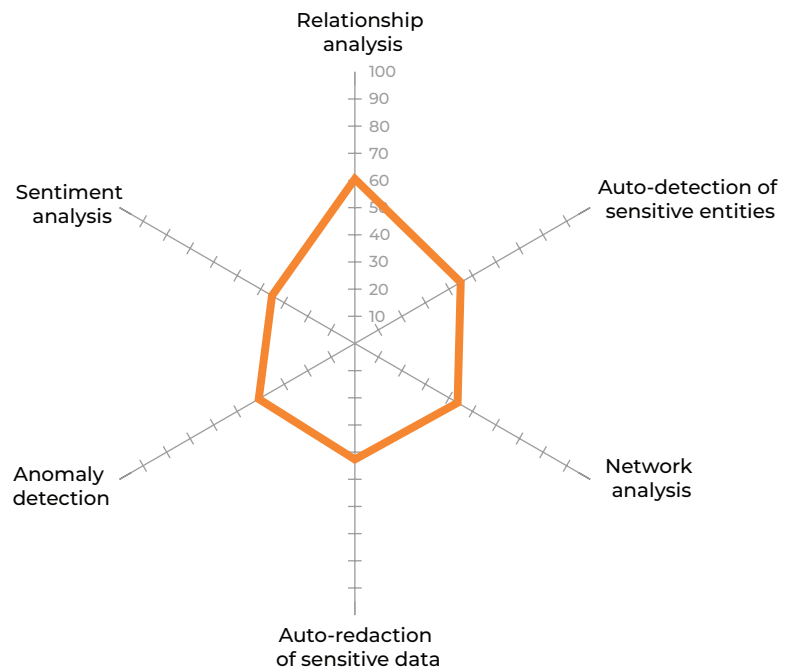
Recommendation: Rather than manually reviewing documents to sort and classify them into types, use automatic document classification to get a jump-start on review. For example, classification tools can detect privileged documents with a high degree of accuracy based on whether they were sent from a privileged party or domain. This pre-sorting not only saves time in review but also improves accuracy by allowing review teams to consider related documents in a coherent set instead of on a piecemeal basis.

4 EMERGING INTELLIGENCE: EARLY ADOPTION

Approaches that reveal the story behind data, unlocking new ways of working and allowing legal professionals to focus on high-value work

The advanced data analysis capabilities in the fourth category represent the next frontier of AI capabilities. While these technologies are not yet in widespread use, early adopters are driving their growth—and unlocking previously hidden insights within their data. Relationship analysis is the best adopted of these approaches, with 61% of respondents reporting its use. Nearly half (49%) of respondents noted that they have adopted pattern analysis tools to identify sensitive data.

Recognizing the possibilities in this category of tools, eDiscovery software vendors are increasingly investing in new data analysis capabilities.



MAIN CAPABILITIES THAT HAVE BEEN ADOPTED:

Relationship analysis: Mapping connections within a community and identifying relationships between individuals based on the pattern of communications between them. This capability relies on both named entity recognition and metadata from email or chat headers to group communications and then calculate the strength and nature of relationships based on their frequency, times, topics, language, and context.

Auto-detection of sensitive entities: Detecting protected or sensitive information—such as personally identifying information, personal health information, or personal commercial information—through entity extraction and pattern recognition.

OPPORTUNITIES FOR FURTHER IMPROVEMENT:

Network analysis and sentiment analysis:

Identifying linked activities and behaviors across networks or groups of constituents and determining whether the general tone of emotions within text is negative, neutral, or positive. These capabilities rely on NLP-based content extraction and supervised machine learning based on multiple algorithms.

Recommendation: Implement network and sentiment analysis tools to gain rapid insight into who said what to whom—and what they meant when they said it.

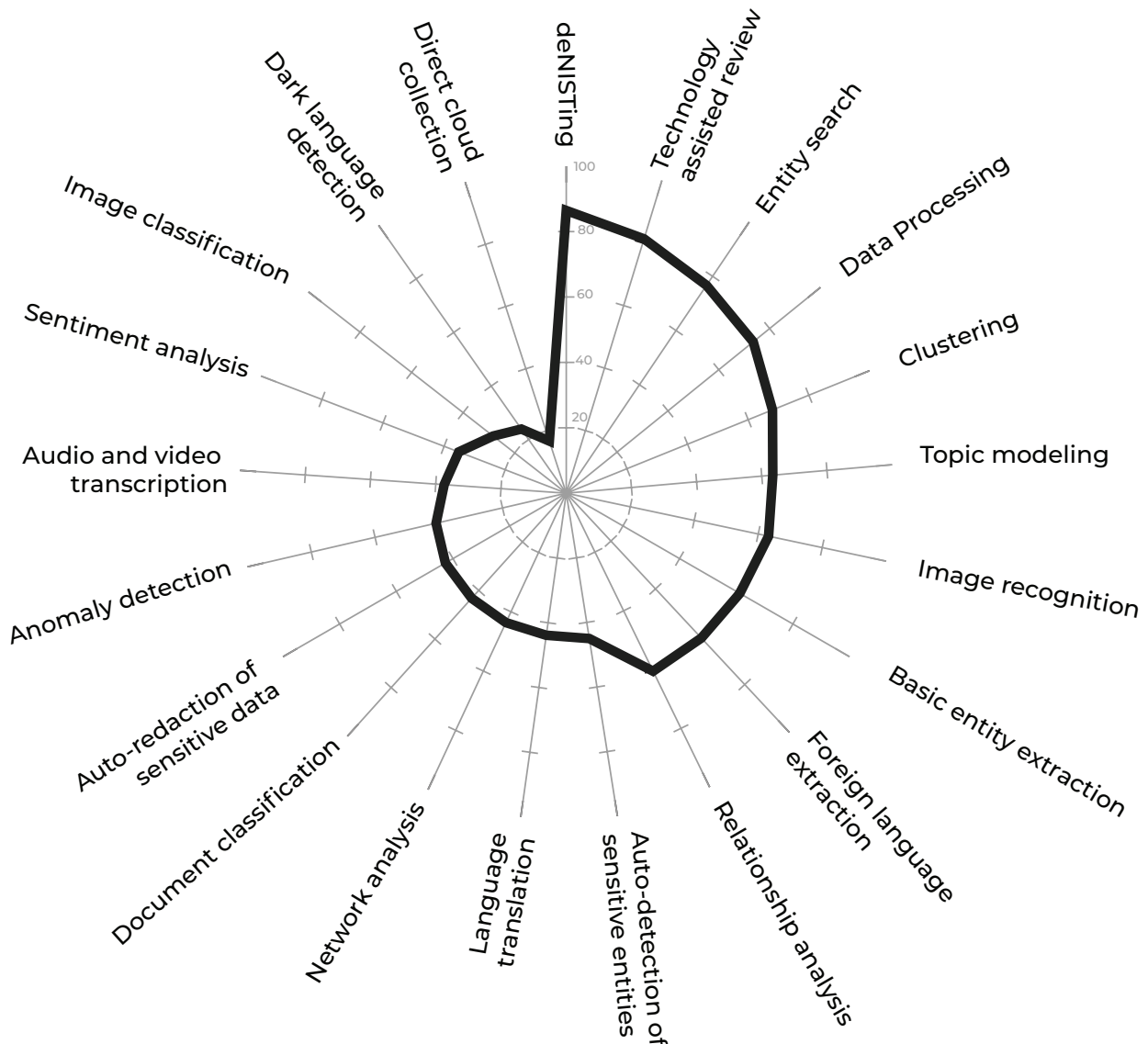
Auto-redaction of sensitive information: Identifying sensitive information based on named entity recognition and automatically redacting that information to prevent its disclosure.

Recommendation: Combine data analytics with in-place search capabilities to pro-actively identify legal risks lurking within the organization's data. This early detection allows for rapid remediation and avoids the specter of litigation or public scandal.

Overall adoption and summary of results

By and large, eDiscovery professionals have adopted at least some of the advanced capabilities we asked about in this survey. Only 27 of 177 respondents (15%) stated that they do not use any form of AI to improve the quality of their work, improve accuracy, or ensure quality control. The remaining 85% reported that they use one or more tools or capabilities. Based on these results, it's safe to conclude that the large majority of the eDiscovery community has adopted at least some AI capabilities that can help them manage the tremendous glut of organizational data as well as automation to improve the speed and accuracy of their work. However, there is substantial room for improvement to continue to tackle the ever-growing volume and complexity of data.

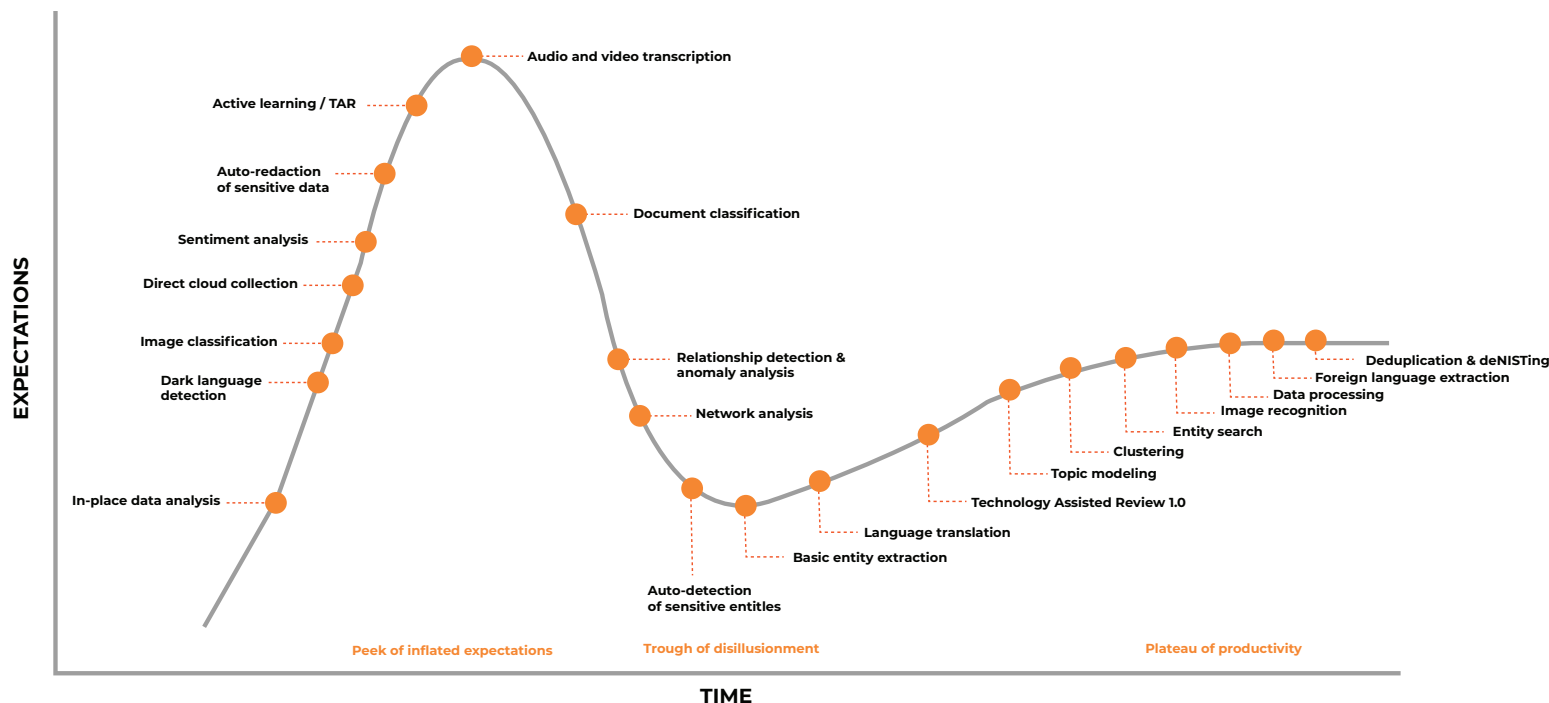
Considering the results in a single view, from most to least adopted, it's interesting to note that TAR is more used by our respondents than early stage data processing. As noted previously, however, this result fits with the profile of our respondents, who primarily work on the right side of the EDRM. This is also to be expected given that nearly half (49%) of respondents work at law firms and only 19% work for external service providers that might handle the earlier data processing and culling stages. Historically, law firms have prioritized enhanced review functionality over data processing.



The eDiscovery hype cycle

Up to this point, we've discussed adoption as if the acceptance of a new technology travels along a straight path from an unknown novelty to a fully adopted, black-letter law tool. Yet as with any other development, AI and other technologies aren't simply adopted or rejected; rather, their popularity waxes and wanes along a fairly predictable path known as the hype cycle.

To better understand our results in the context of the real, non-linear world, we've plotted the technologies and capabilities assessed in our survey onto an eDiscovery hype cycle as follows.



1. This is the age of data analytics.

Data automation techniques such as deduplication and data processing are fully adopted, having reached the plateau of productivity. However, with data volumes continuing to expand, these approaches are no longer sufficient. We have the tools needed to efficiently manage ever-greater data volumes and surface the highly relevant, valuable data that eDiscovery depends on, but those tools—such as sentiment analysis and dark language detection—have yet to reach peak adoption.

2. To unlock the value of AI in eDiscovery, practitioners need better tools for data processing and collection.

Data volume isn't the only challenge in eDiscovery; as businesses generate data in a variety of new formats from collaboration tools like Microsoft Teams, Zoom, and Slack, eDiscovery practitioners need tools that can accommodate those various data formats so that

AI functions can actually work with the data. Data analytics downstream in the review stage of the EDRM can only be valuable if practitioners input clean data in the early stages. Additionally, we haven't seen the end of remote work and virtual meetings, so eDiscovery needs to encompass audio and video content.

3. Practitioners should focus on upstream eDiscovery.

Legal professionals shouldn't have to collect data to then cull and search through it or determine its relevance. By deploying in-place search functions and in-place data analytics, organizations can avoid the costs—and the security risks—associated with unnecessary data collection, excessive storage, and review of unresponsive data. In addition, in-place analytics have the potential to unlock new value adds for legal professionals, by proactively reducing legal risks in data as it sits in place.

Perceptions of AI among eDiscovery practitioners

To improve the acceptance and adoption of these underused capabilities, we sought to understand the barriers to their implementation. What's holding eDiscovery professionals back from adopting tools like dark language detection? Many of these capabilities are baked in to broader eDiscovery platforms, indicating that cost and availability aren't the only factors at play.

Other barriers delaying adoption could be related to ethical concerns. Attorneys ultimately bear responsibility for every aspect of a legal representation. That includes the choices that an AI system makes during eDiscovery. Attorneys may be uncomfortable with their lack of awareness or understanding of how an AI system reaches those decisions. Without a strong sense of the benefits of AI, attorneys might continue to choose older methodologies rather than trusting AI in eDiscovery.

We asked several questions to explore the general perceptions around AI and attempt to tease apart the beliefs and attitudes that might be delaying the adoption of some tools. After all, if an approach isn't commonly thought to yield high-quality results, save time, or reduce risks, the legal industry is unlikely to investigate it further.

93%
AI can help automate eDiscovery work

87%
AI is useful for generating better insights from data

Nearly all respondents (93%) agreed that AI can help to automate eDiscovery work. Almost as many (87%) believe that AI is useful for generating better insights from data. But about half of respondents stated that AI cannot match human competencies across every stage of eDiscovery, from data identification and collection to review and production.

Responses were more optimistic regarding the eventual capacity of AI to reach parity with human competencies for data processing (74%) and review (68%), but by and large, our respondents

seemed to expect that AI would not outperform humans in most eDiscovery applications.

More specifically, just over half (53%) of respondents didn't believe that AI could reduce risk by improving the quality of their work. This result may reflect the underlying notion that while AI is capable of automating aspects of data management and providing some insights into data patterns, it isn't "smart" enough to make contextual decisions. In this regard, AI is seen as an augments of human intelligence rather than a replacement for it.

There are certainly aspects of human intelligence that AI does not, and cannot, match. AI-based machines can assess certain data characteristics quickly and accurately, and they are consistently rational, but they aren't intuitive, emotional, or culturally sensitive. These "soft" abilities are what make humans unique and effective, even with the rules-based legal industry.

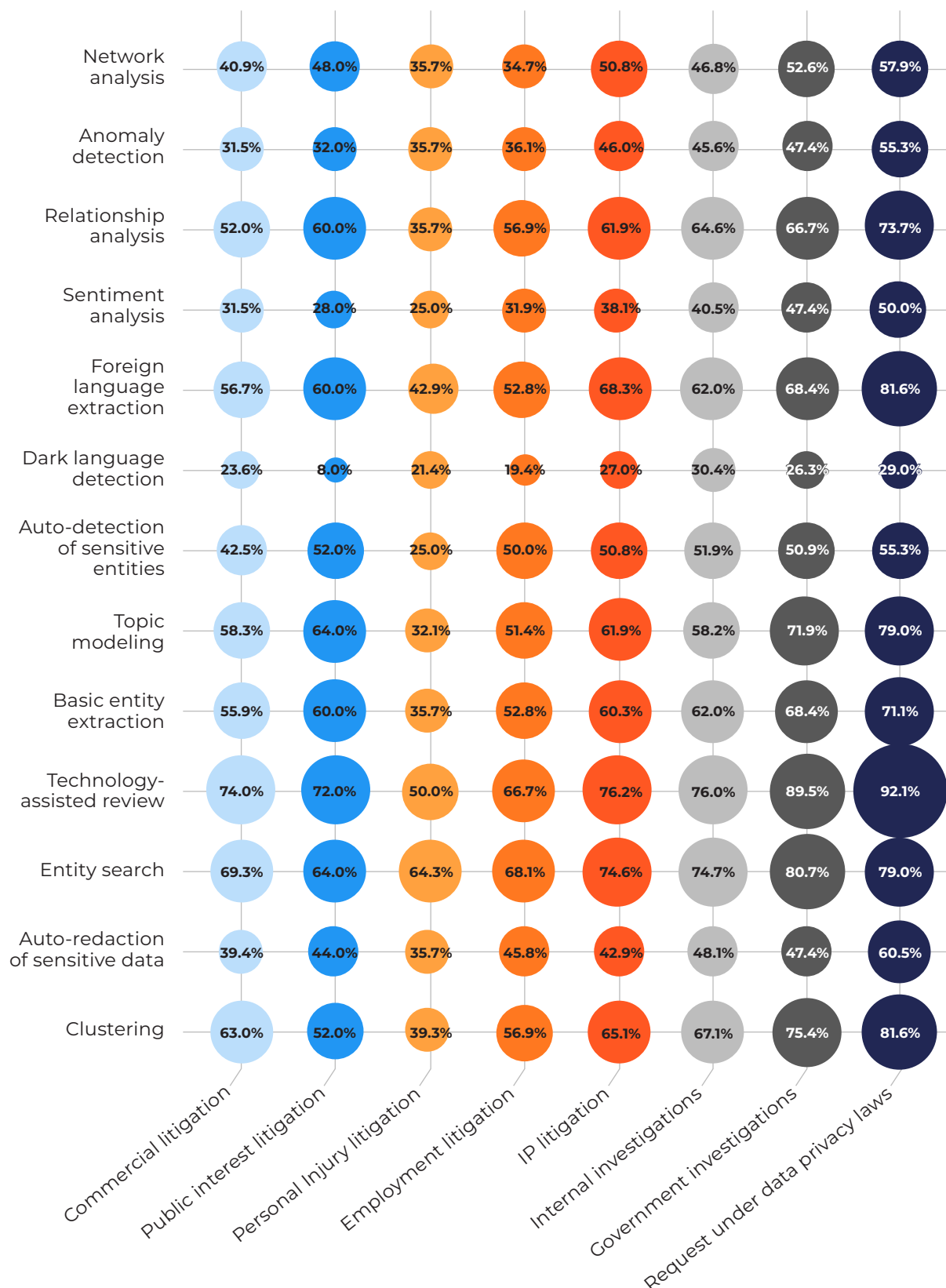
We fully anticipate that AI tools will find their best applications in areas where they can empower, rather than replace, human beings. We envision solutions that provide rapid answers and rich decision support for legal professionals, giving humans the data-based insights they need to draw solid conclusions and take effective action.

Where do we go from here? We asked our respondents whether they expected that more advanced forms of AI could play a role in eDiscovery project resource planning, case strategy planning, or settlement negotiations, among others. More than half of respondents (55%) answered that AI had a part to play in all of the above.

Finally, we wanted to evaluate the degree to which adoption of AI and other technologies varied across different areas of legal practice. We therefore plotted the use of different analytics capabilities in a heatmap display according to our respondents' areas of practice. This representation makes it immediately clear, for example, that dark language detection is poorly adopted across different practice areas, while TAR is heavily used nearly everywhere.

AI applications across eDiscovery use cases

In our survey, this is how tool use mapped across use cases.



As a practical guide, we suggest that eDiscovery professionals could transform the value of their legal work by adopting specific capabilities based on their practice area.

For example:

- Internal investigations and employment disputes could benefit from greater use of network analysis, topic modeling, and sentiment analysis tools. Topic modeling in particular could help investigators identify secret code words or other “unknown unknowns” that might trigger a closer look at particular conversations or individuals, while network and sentiment analysis could help practitioners quickly find clues to guide further investigation.
- Commercial or civil litigation practitioners could improve results, particularly in early case assessment, by implementing document classification to save time, relationship analysis to identify important data faster, and entity searching technologies to reduce the noise in search results and reduce the number of documents advancing to review.

- Government investigations or regulatory requests involving data privacy or compliance could be expedited through the use of relationship analysis, named entity recognition, and auto-redaction to protect sensitive information from inadvertent disclosure.

Across all practice areas, the least-used data analytics capabilities we considered were dark language detection (24%), sentiment analysis (35%), and anomaly detection (41%). These tools are ripe for broader adoption in the eDiscovery community.

LEAST USED

24%
dark
language
detection

35%
sentiment
analysis

41%
anomaly
detection

The future of the legal profession and the value of AI for eDiscovery

Legal professionals are under unrelenting pressure to add more value for their clients without raising their prices. With increased competition in the legal market, adaptation and innovation are critical for continued success. The practice of law is moving away from a reactionary stance in favor of proactively identifying and mitigating potential risks. Those who find ways to leverage the enormous volume of organizational data to gain actionable insights will be well positioned to outperform their competitors.

The new capabilities of AI and other technologies present a remarkable opportunity to rapidly sift through vast quantities of data to identify critical insights fast, thereby expanding value while ensuring legal defensibility. AI and other technologies can help legal professionals through:

- Simple automation of routine legal tasks;
- Data analytics that generate insights from piles of unstructured data; and
- Analytics capabilities that augment human intelligence, improving the accuracy and quality of the work done by legal professionals.

Naturally, there is still room for further improvement. Advanced automation features like document classification, automatic image classification, and auto-redaction may improve the quality of results while saving a significant amount of time. Additionally, as new media formats such as video calls become more commonplace in eDiscovery, transcription of audio and video recordings will greatly reduce the burden of eDiscovery work.

Moving the focus of eDiscovery upstream

One overlooked solution to the rise in data volumes is what we call “upstream eDiscovery.” In this approach, practitioners address problems with data volumes and complexity in the upstream stages of eDiscovery such as collection rather than waiting to manage them in review.

In a sense, eDiscovery is like prospecting for gold. Practitioners often focus on how they can improve their work downstream—on the right side of the EDRM—to sift through data more quickly and accurately identify the “gold” hidden within. As a result, organizations spend somewhere around 70% of their eDiscovery time and budget in the document review phase, endlessly panning for gold.

But there’s another way. Instead of focusing on improving review downstream—and continually expanding the capacity of the input funnel to accommodate more data—we can use AI-powered data analytics to improve how we select the “dirt” that we feed into our eDiscovery stream in the first place. By being more intentional about how we select data in the early stages of the EDRM, we can reduce the overall volume of data we must manage in the review phase and focus on more relevant, high-quality information.

How can eDiscovery professionals become better data prospectors? By deploying many of the capabilities we’ve discussed here, from in-place data search and document classification to relationship analysis and topic modeling.

We can’t wait to see how the field continues to evolve as these tools move along the pathway to full adoption.

About the Authors



Johannes Scholtes

Chief Data Scientist – IPRO

Scholtes oversees research, development and legally defensible application of data science and artificial intelligence in all of IPRO's and ZyLAB's eDiscovery and Information Governance software. He's been involved with ZyLAB since 1988, and is full-professor of text-mining at the Data Science and AI Department of Maastricht University. In addition, he teaches various courses on Legal Technologies at Leiden University.



Sarai Schubert

Chief Operations Officer – IPRO

Schubert oversees all technical and operational aspects of IPRO, including product strategy, cloud solutions and customer service operations. Schubert previously managed the eDiscovery global operations at Latham & Watkins, LLP and brings over 17 years of industry experience.



Reyndert Coppelmans

VP Marketing ZyLAB / Director Demand Generation - IPRO

Coppelmans oversees US and EMEA Marketing at ZyLAB. In the broader IPRO organization he's responsible for demand generation in the US market, and oversee Marketing activities in EMEA. Before Legal Tech, Coppelmans has worked in similar capacities in HR - tech and management consulting.

ABOUT IPRO

IPRO is reinventing the way organizations interact with their data. By transforming the EDRM and thinking “upstream,” we bring insight far earlier in the eDiscovery process, enabling teams inside and outside your organization to collaborate transparently at any stage, before data is even collected. By thinking about and interacting with your data differently, you’ll eliminate waste and reduce risk.

We combine best-in-class governance and eDiscovery software with predictive analytics to give corporations, government agencies, law firms, and legal partners the most open and effective solution to data challenges. We bring you this power with an obsession on time-to-value and your success.

ABOUT ZYLAB

ZyLAB—part of the IPRO family—helps legal professionals of today turn big data challenges into legal successes. Our user-friendly eDiscovery platform, powered by AI, helps legal professionals deliver fast and complete investigations. ZyLAB is trusted by Fortune 1000 companies, government agencies, courts, regulatory agencies, and law firms worldwide and has nearly 10,000 installations and 2 million users.

ABOUT ACEDS

The Association of Certified E-Discovery Specialists provides its members with the tools that are essential for developing or maintaining competency in e-discovery. We seek to provide the highest quality information, training, networking, and certification for our growing membership of e-discovery professionals across the globe. ACEDS has global chapters that assist our diverse community with ongoing education, networking, events, and careers.