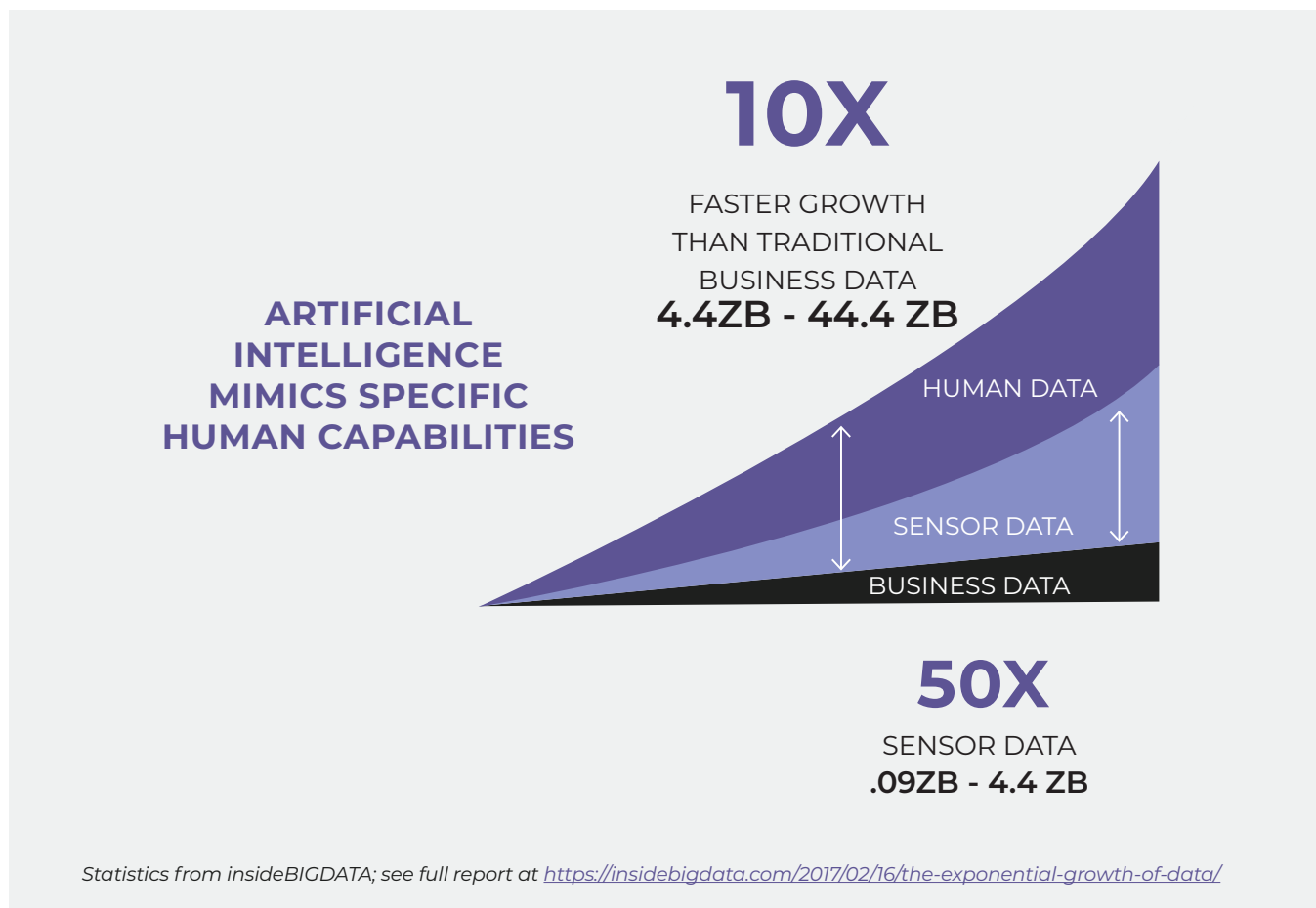# 2021 State of AI and Technology Adoption in eDiscovery

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.

ARTIFICIAL INTELLIGENCE MIMICS SPECIFIC HUMAN CAPABILITIES

10X

FASTER GROWTH THAN TRADITIONAL BUSINESS DATA
4.4ZB - 44.4 ZB

HUMAN DATA

SENSOR DATA

BUSINESS DATA

50X

SENSOR DATA
.09ZB - 4.4 ZB

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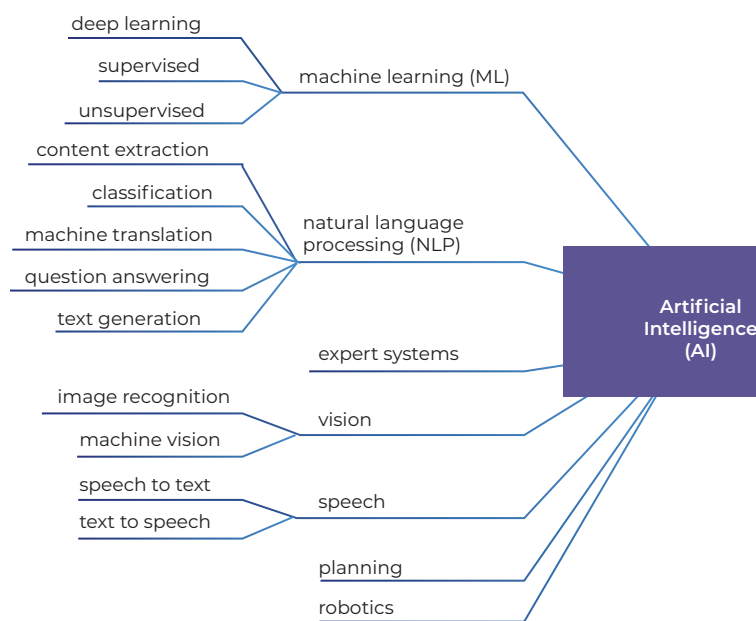
This white paper reports our results and recommends next steps for how eDiscovery leaders can advance the use of AI and other technologies in eDiscovery.

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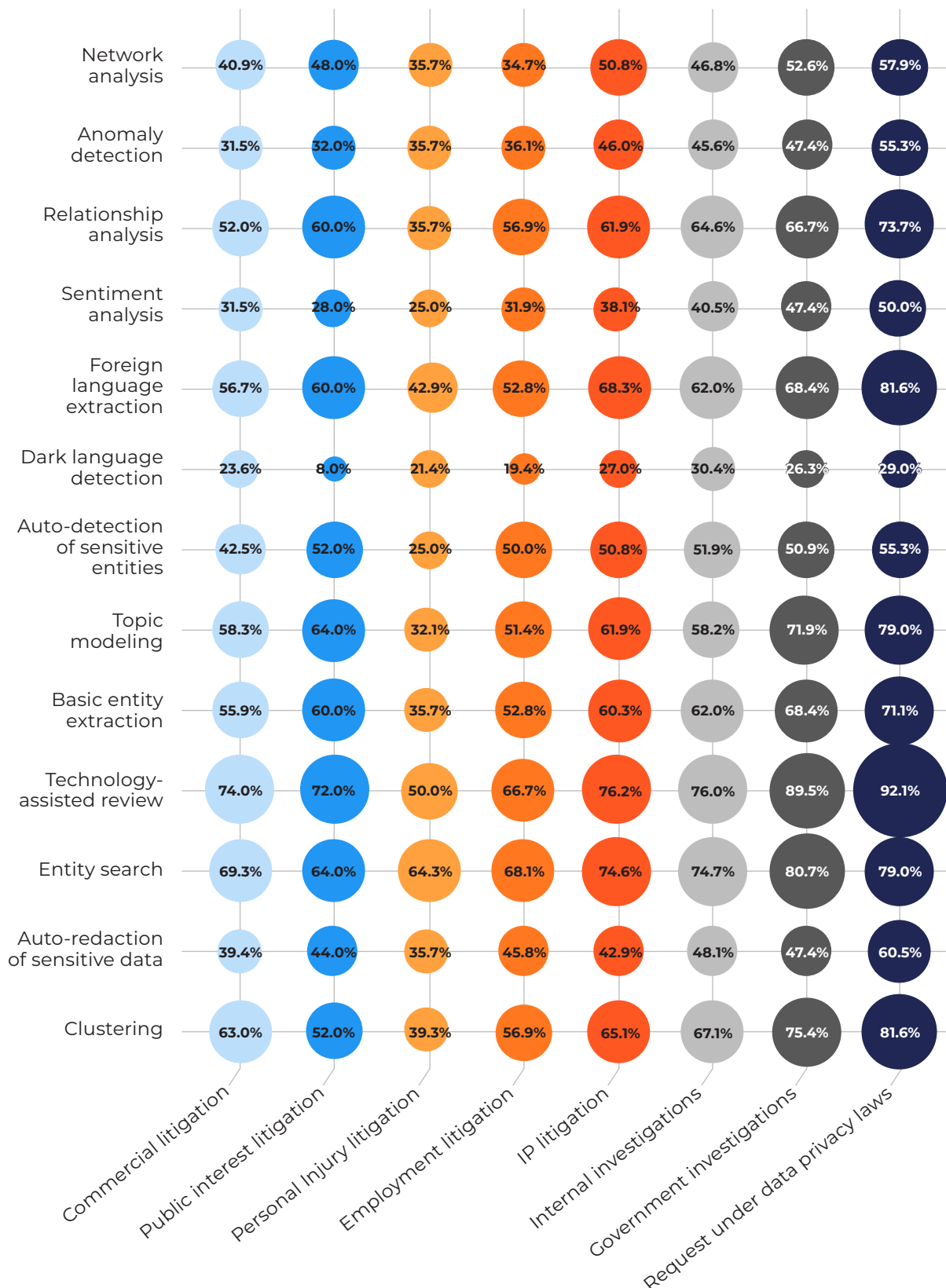
But while AI is a persistent buzzword in the legal technology field, not all helpful technologies involve its application. Simple automation, for example, lacks the panache of machine learning, but it offers tremendous advantages throughout the eDiscovery pipeline.

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.

deep learning
supervised
unsupervised
→ machine learning (ML)

content extraction
classification
machine translation
question answering
text generation
→ natural language processing (NLP)

expert systems

image recognition
machine vision
→ vision

speech to text
text to speech
→ speech

planning
robotics

→ Artificial Intelligence (AI)

AI applications across eDiscovery use cases

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

	Commercial litigation	Public interest litigation	Personal Injury litigation	Employment litigation	IP litigation	Internal investigations	Government investigations	Request under data privacy laws
Network analysis	40.9%	48.0%	35.7%	34.7%	50.8%	46.8%	52.6%	57.9%
Anomaly detection	31.5%	32.0%	35.7%	36.1%	46.0%	45.6%	47.4%	55.3%
Relationship analysis	52.0%	60.0%	35.7%	56.9%	61.9%	64.6%	66.7%	73.7%
Sentiment analysis	31.5%	28.0%	25.0%	31.9%	38.1%	40.5%	47.4%	50.0%
Foreign language extraction	56.7%	60.0%	42.9%	52.8%	68.3%	62.0%	68.4%	81.6%
Dark language detection	23.6%	8.0%	21.4%	19.4%	27.0%	30.4%	26.3%	29.0%
Auto-detection of sensitive entities	42.5%	52.0%	25.0%	50.0%	50.8%	51.9%	50.9%	55.3%
Topic modeling	58.3%	64.0%	32.1%	51.4%	61.9%	58.2%	71.9%	79.0%
Basic entity extraction	55.9%	60.0%	35.7%	52.8%	60.3%	62.0%	68.4%	71.1%
Technology-assisted review	74.0%	72.0%	50.0%	66.7%	76.2%	76.0%	89.5%	92.1%
Entity search	69.3%	64.0%	64.3%	68.1%	74.6%	74.7%	80.7%	79.0%
Auto-redaction of sensitive data	39.4%	44.0%	35.7%	45.8%	42.9%	48.1%	47.4%	60.5%
Clustering	63.0%	52.0%	39.3%	56.9%	65.1%	67.1%	75.4%	81.6%

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.