Sponsored by: ///ALEGION

dimensional research

Artificial Intelligence and Machine Learning Projects Are Obstructed by Data Issues

Global Survey of Data Scientists, AI Experts and Stakeholders

May 2019

Executive Summary

This research finds Artificial intelligence (AI) and machine learning (ML) projects are no longer novel, as nearly half of companies surveyed have undertaken 4 or more active projects. But it is still early days, as 7 out 10 have started their project in only the last two years, and just half of those projects have been put into production. Also, half of those surveyed reported their AI/ML teams have 10 or fewer members.

The nascency of enterprise AI has led more than half of the surveyed companies to label their training data internally or build their own data annotation tool. Unfortunately 8 out of 10 companies indicate that training AI/ML algorithms is more challenging than they expected, and nearly as many report problems with projects stalling. 96% of companies surveyed stated they have run into training-related problems with data quality, labeling required to train the AI, and building model confidence.

This leads to 7 out of 10 companies utilizing external services for the AI or ML projects with many of them focusing on data collection, labeling and expertise. With AI/ML talent rare and expensive, this research suggests that enterprises should consider using external solution providers for critical activities like data labeling and model scoring. The data provides evidence that such outsourcing leads to improved outcomes.

This survey reveals that enterprises assign strategic value to their machine learning initiatives. They expect AI and ML to improve all aspects of their businesses, and potentially to be disruptive in their industry sectors.

dimensional research

Key Findings

Enterprise AI/ML is Young But Growing

- 70% report that first AI/ML investment was within last 24 months
- Over half of enterprises report they have undertaken fewer than 4 AI and ML projects

Numerous AI/ML Project Hurdles

- 78% of AI/ML projects stall at some stage before deployment

Training Data Issues Constrain Project Success

- 96% of enterprises encounter data quality and labeling challenges
- 63% have tried to build their own technology solutions
- 71% of teams ultimately outsource ML project activities
- Teams that outsource data labeling get projects into production faster

DETAILED FINDINGS

Over Half of Enterprises Report They Have Undertaken Fewer than 4 AI or ML Projects

How many AI or ML projects has your company started?
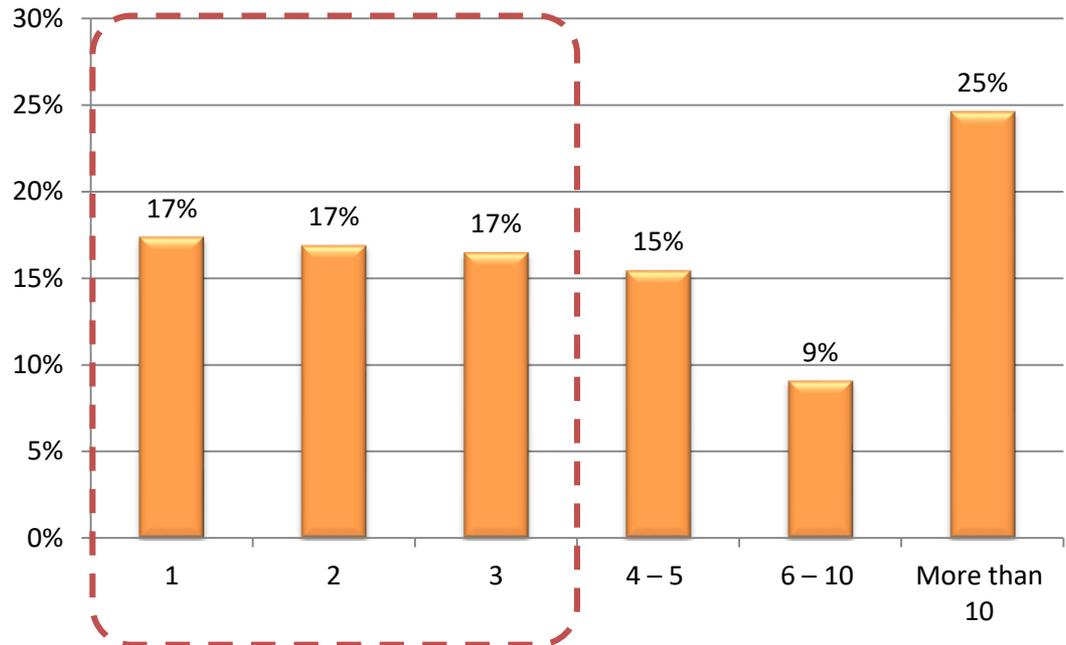
70% Report Their First AI/ML Project Involvement Was Within the Last 24 Months

When did you start your first AI or ML project?
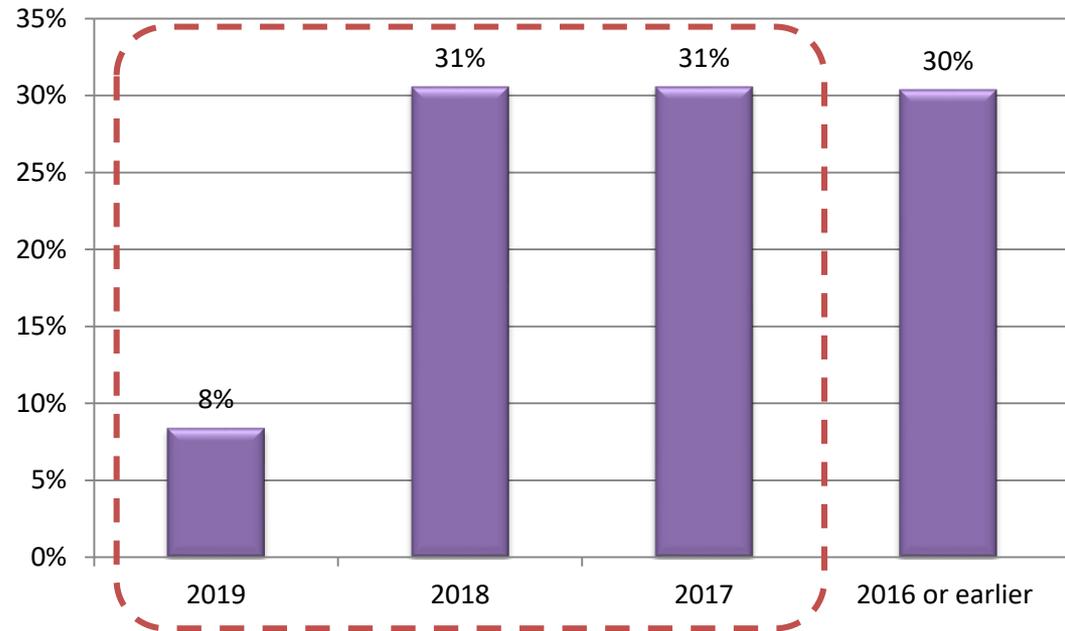
The Majority of Enterprise AI/ML Project Teams Have 10 or Fewer Members

Approximately, how many people are involved with your AI or ML projects?

dimensional research

Sponsored by: ///ALEGION
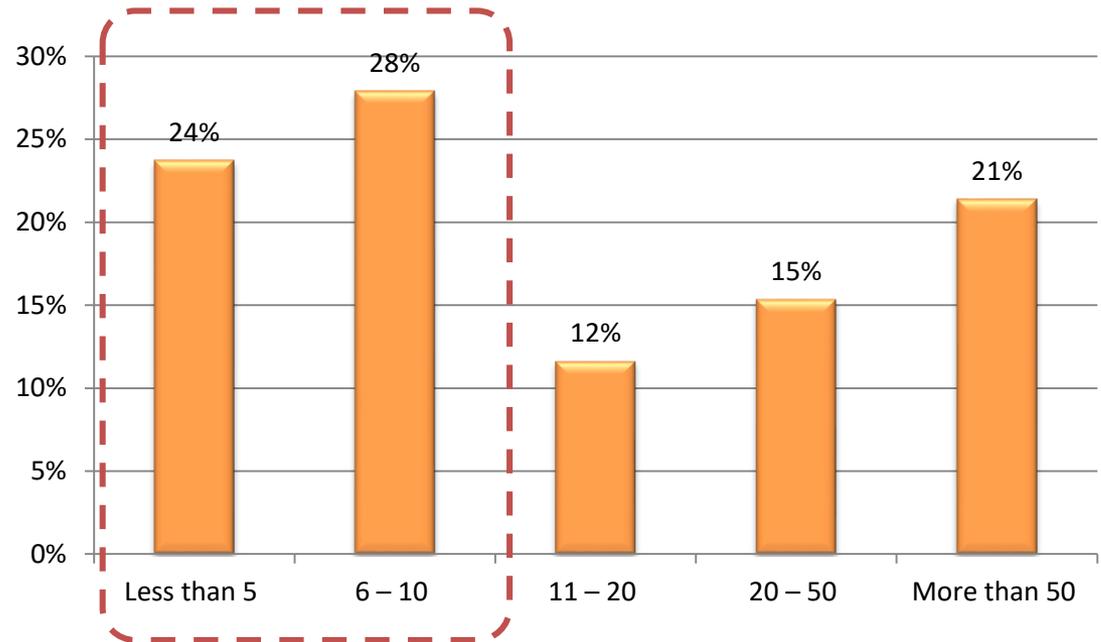
Most AI/ML Projects Involve NLP, CV, or Entity Resolution

What AI approach best describes your most recent AI or ML project?

Other
13%

Natural language processing
45%

Entity resolution
17%

Computer vision
25%

dimensional research

Sponsored by: ///ALEGION
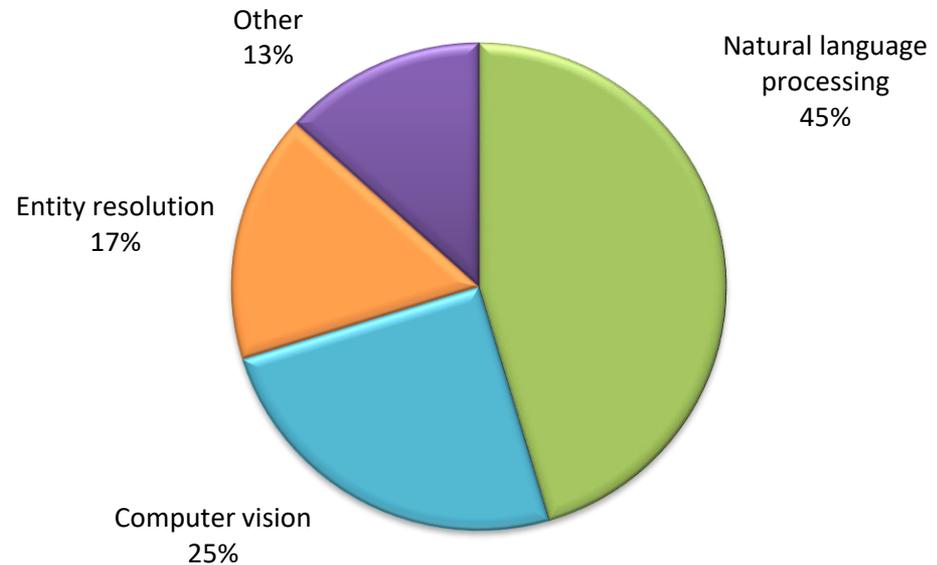
Only Half of Enterprises Have Released An AI/ML Project Into Production

Have any of your company's AI or ML projects moved into production yet?

No
49%

Yes
51%
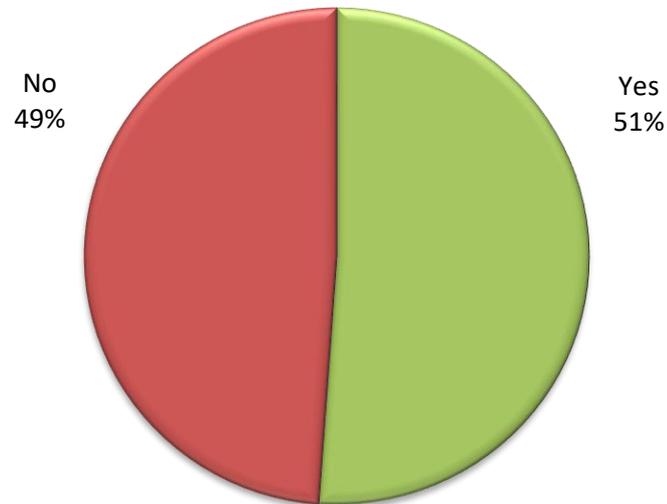
65% of Projects Have Reached the Data Labeling and Algorithm Training Phase

Has your AI or ML project reached the stage where it is using labeled training data to train its algorithm?

No
35%

Yes
65%
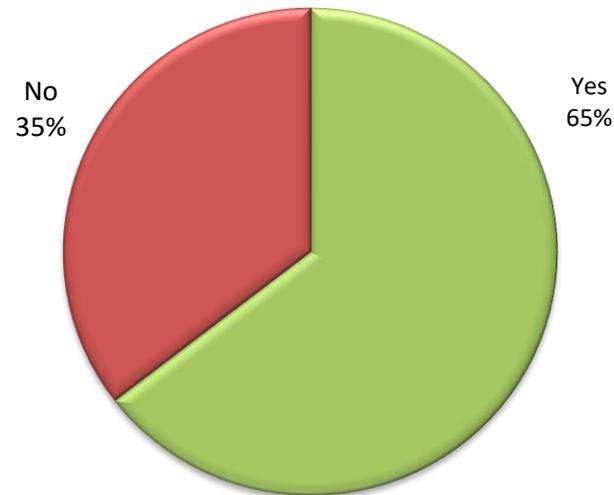
72% Report that Production-Level Model Confidence Will Require More than 100,000 Labeled Data Items

Approximately, how much training data will be required to deploy your model with confidence?

Chart data:
- Fewer than 100,000 data items: 28%
- 100,000 – 1,000,000 data items: 43%
- 1 million – 10 million data items: 19%
- More than 10 million data items: 10%
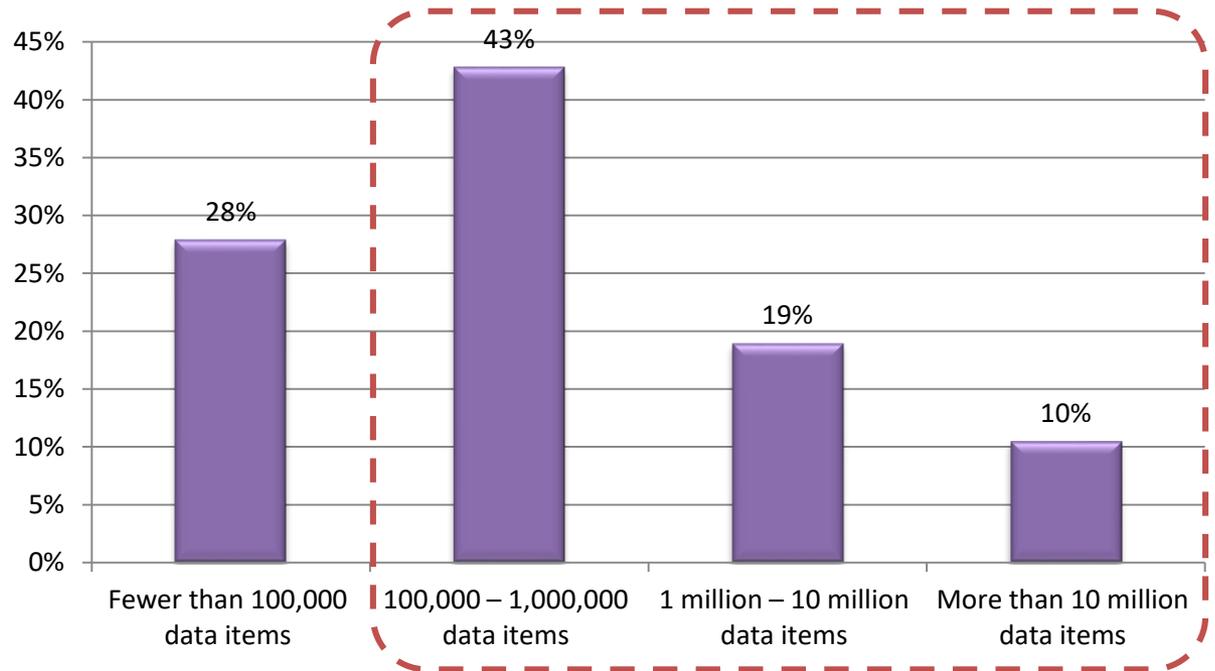
81% Admit Training AI with Data Is More Difficult Than Expected

Has training the AI with data been more challenging than expected?

No
19%

Yes
81%

dimensional research

Sponsored by: ///ALEGION
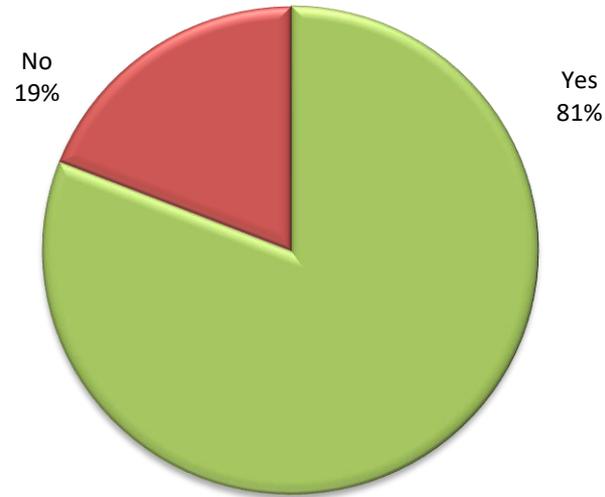
96% Encounter Challenges with Training Data Quality and Quantity

Which problems has your company experienced with AI training data specifically?

Category	Percentage
Bias or errors in the data	66%
Not enough data	51%
Data not in a usable form	50%
Don't have the people needed to label the data	28%
Don't have the tools needed to label the data	27%
We have not experienced any data training problems	4%
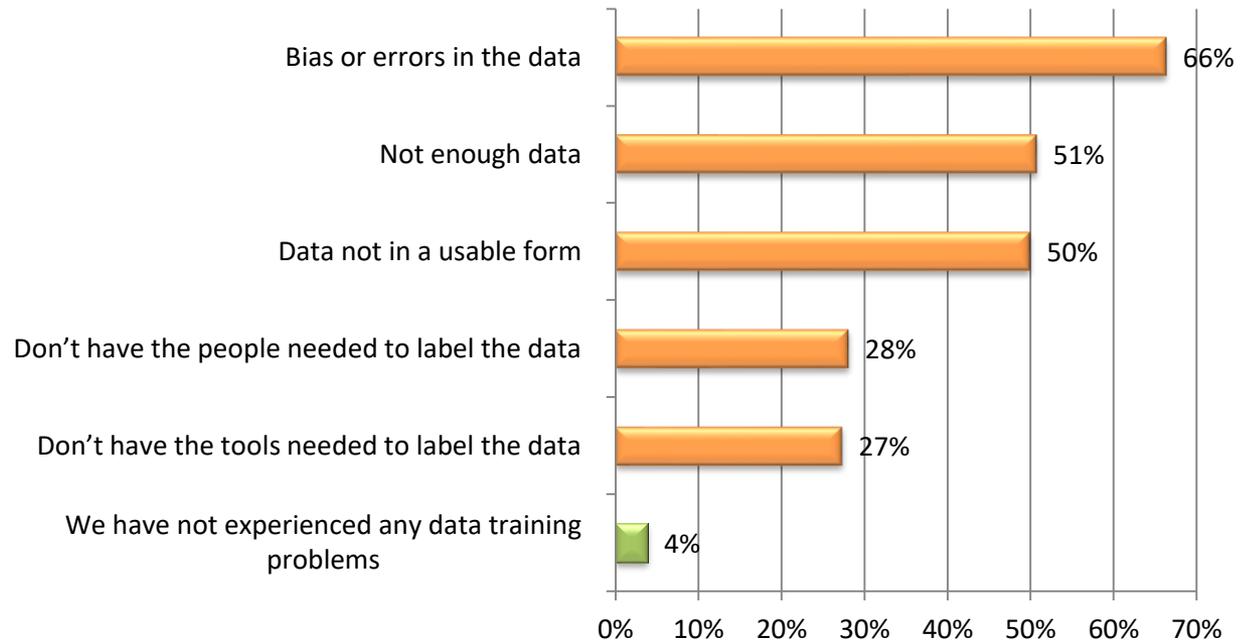
78% of AI or ML Projects Stall at Some Stage Before Deployment

In which phase did your AI or ML project stall?

Phase	Percentage
Proof of concept	33%
Model validation and scoring	13%
Algorithm development	9%
Training data preparation	7%
Model deployment	6%
Algorithm training	6%
Post-deployment enhancement	4%
Our project has not stalled	22%
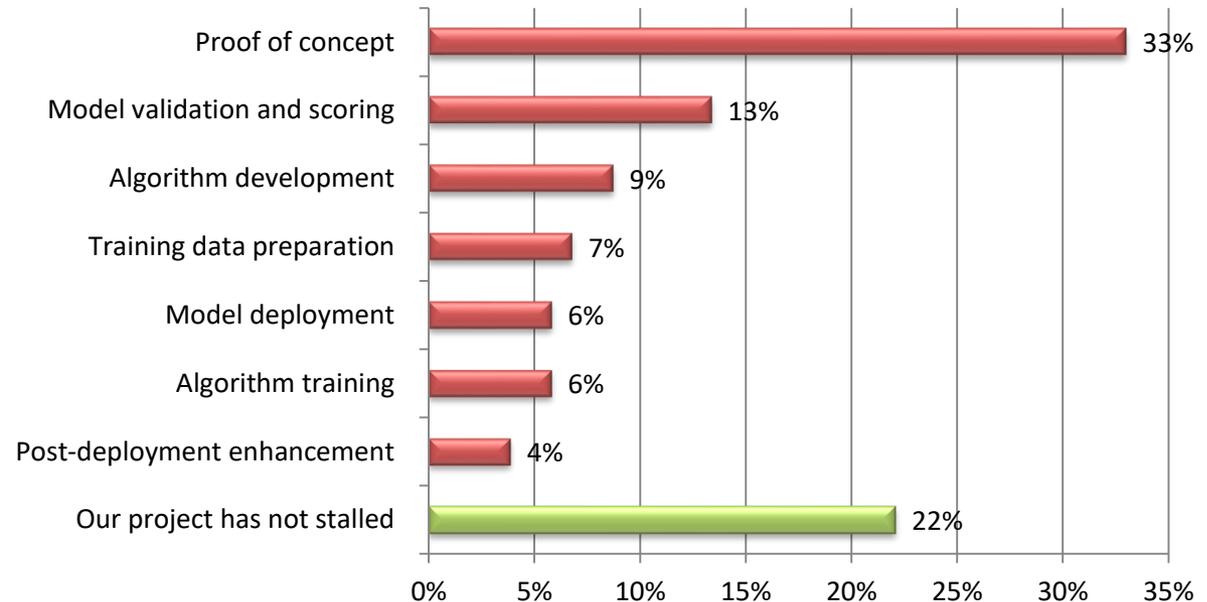
Companies Pursue a Mix of Data Labeling Approaches

What approach is being used to train the initial algorithm?

Approach	Percentage
Labeled and annotated company data internally	76%
Acquired pre-labeled data	45%
Contracted labeling and annotation of company data	30%
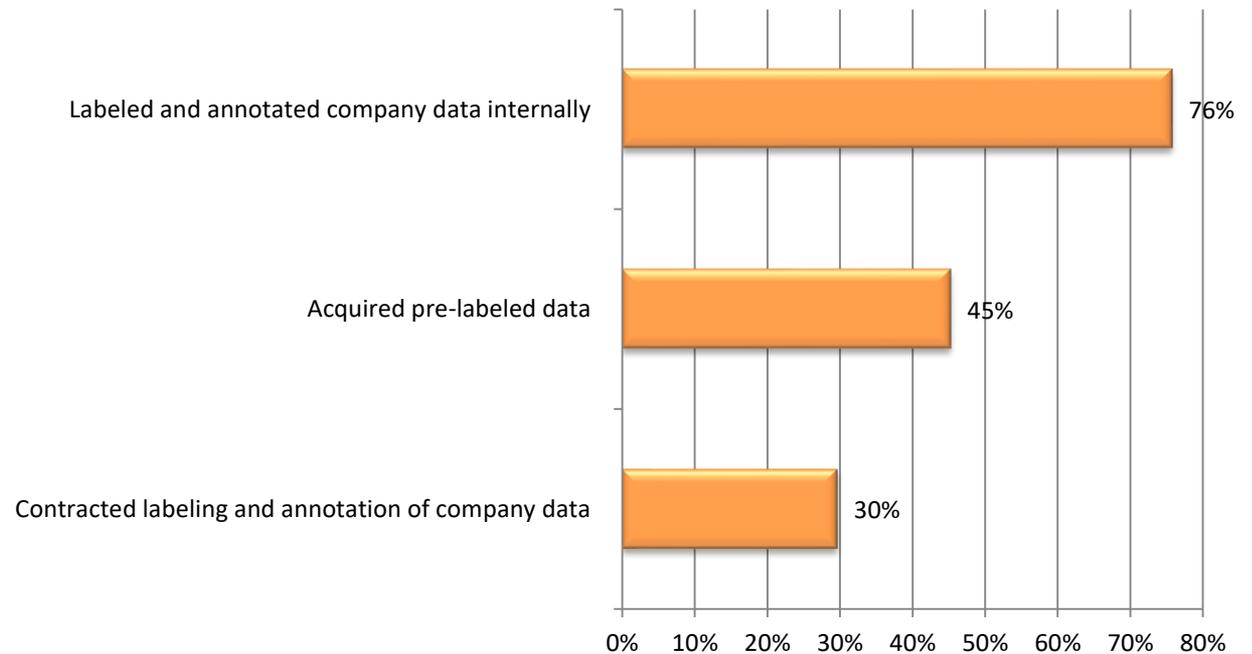
Companies Pursue a Mix of Data Labeling Technology Strategies

What solution was used in your most recent project to label and annotate your training data?

In-house solution	63%
Commercial solution	34%
No solution (manual process)	19%
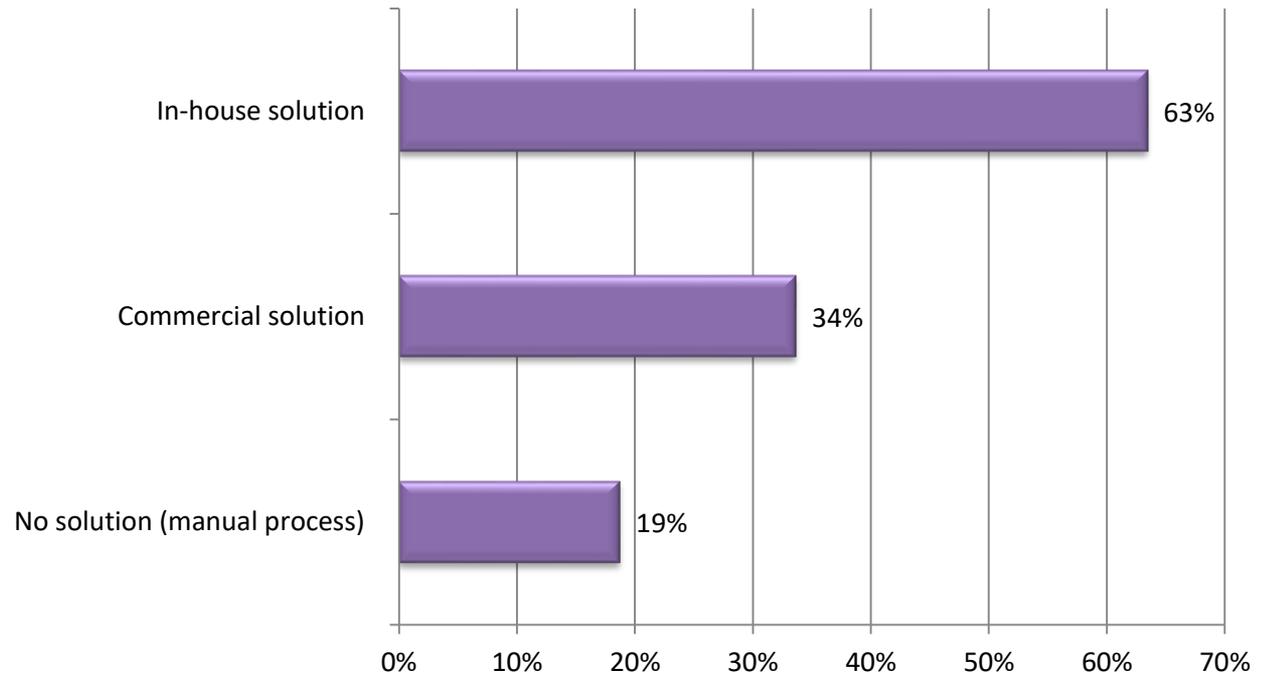
71% of Companies Outsource AI/ML Activities with 3 of the Top 5 Related to Training Data

Which of the following external services has your company used for its AI and ML projects?

| Activity	Percentage
Data collection →	36%
Model development and testing	36%
Data labeling and annotation* →	28%
Overall project strategy and objectives	27%
Data specialist personnel →	26%
Deployment into production	19%
Measurement and maintenance of production…	18%
Governance and management	15%
None, our project was done entirely with…	29%
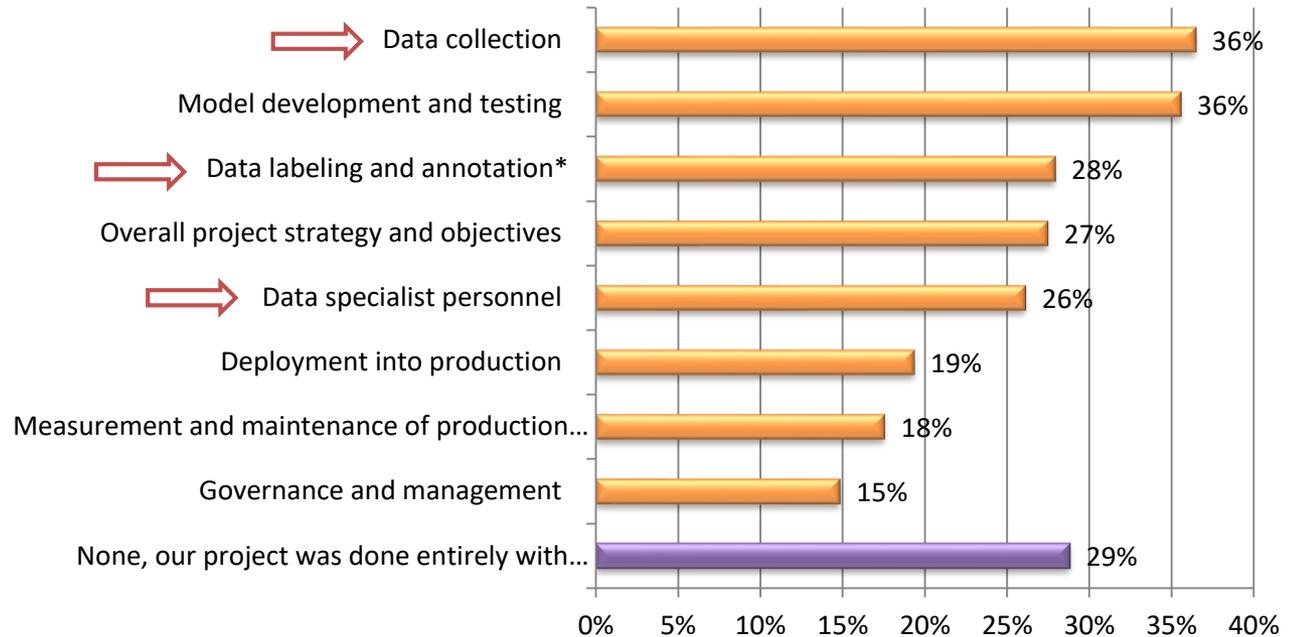
*Survey answer was labeled 'Data Preparation'

dimensional research

17

Sponsored by: //ALEGION

Utilizing Data Labeling Solution Providers Moved Projects Into Production Faster

What approach is being used to train the initial algorithm?

Contracted Labeling and Annotation of Company Data

Bar chart:
- Project in Production: 33%
- Project not in Production: 23%

dimensional research

Sponsored by: ///ALEGION
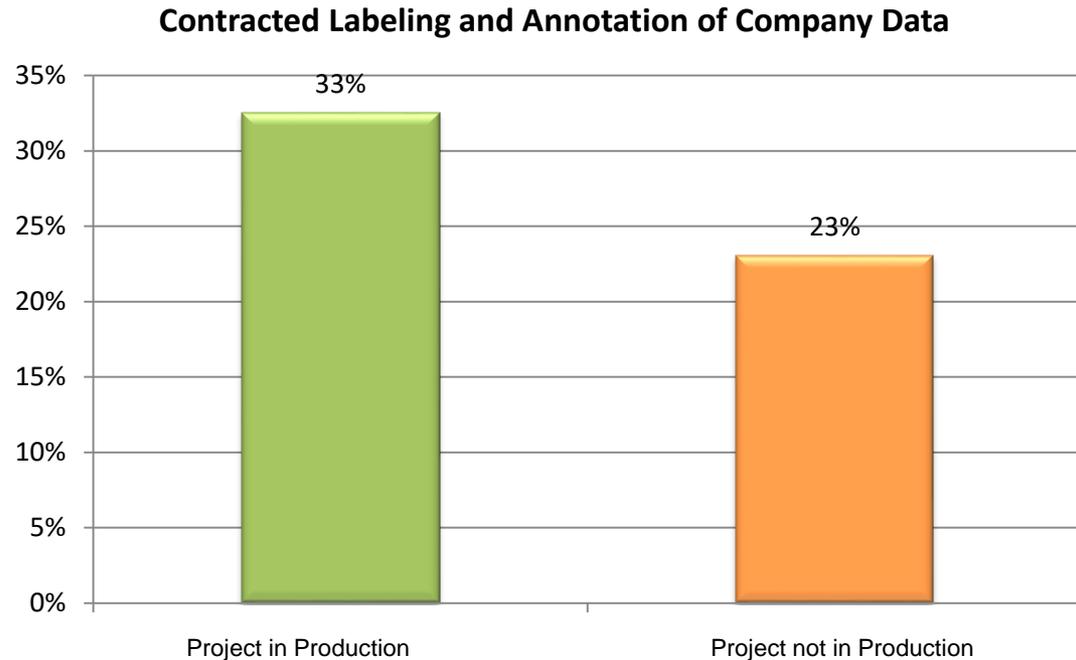
METHODOLOGY AND PARTICIPANTS

dimensional research

Sponsored by: ///ALEGION

Goals and Methodology

Research Goal
The primary research goal was to estimate the maturity of ML in the enterprise. Additionally, the research sought to understand today's ML project challenges, and the tools and resources used in these projects.

Methodology
Data professionals and business stakeholders involved in active AI and ML projects were invited to participate in a survey on their company's use and development of AI and ML projects.

The survey was administered electronically, and participants were offered a token compensation for their participation.

Participants
A total of 227 data scientists, AI experts and business stakeholders completed the survey. Participants were from all 5 continents.

Companies Represented

Size

- 1,000 – 5,000: 37%
- More than 5,000: 63%

Industry

Industry	Percentage
Technology	20%
Healthcare	13%
Financial Services	13%
Education	8%
Manufacturing	8%
Government	6%
Retail	6%
Services	5%
Telecommunications	4%
Food and Beverage	4%
Media and Advertising	3%
Energy and Utilities	3%
Transportation	2%
Pharmaceutical	1%
Non-Profit	1%
Other	3%
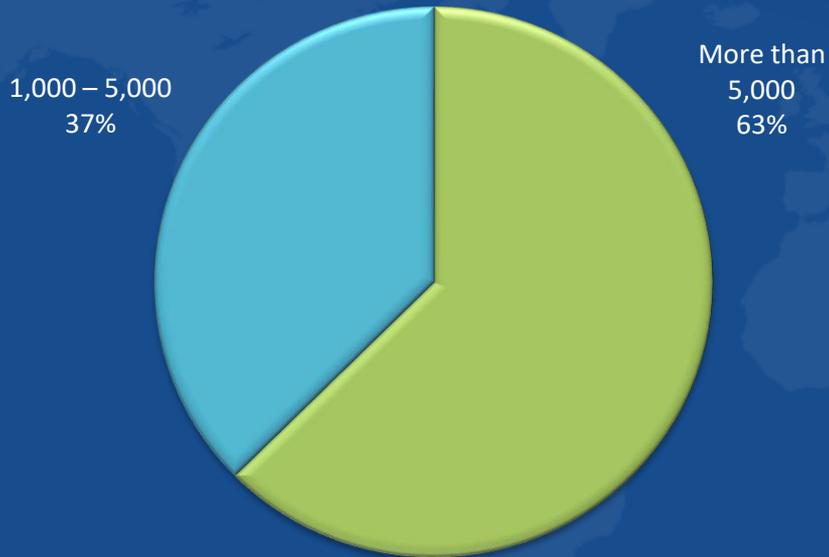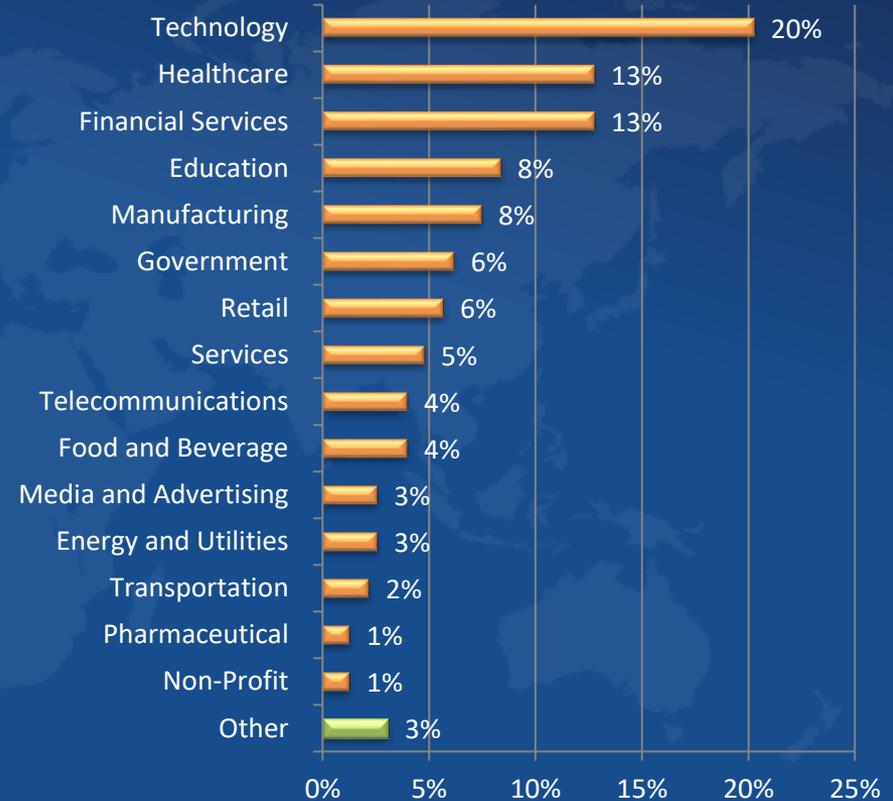
Companies Represented

AI / ML Deployment

Yes, we plan to start our project but more than 6 months from now
18%

Yes, we have already started our AI or ML project
60%

Yes, we will be starting our AI or ML project in the next 6 months
22%

Location

United States or Canada	80%
Europe	11%
Mexico, Central America, or South America	4%
Asia	2%
Australia or New Zealand	2%
Middle East or Africa	1%

0%   20%   40%   60%   80%   100%

dimensional research

Sponsored by: //ALEGION
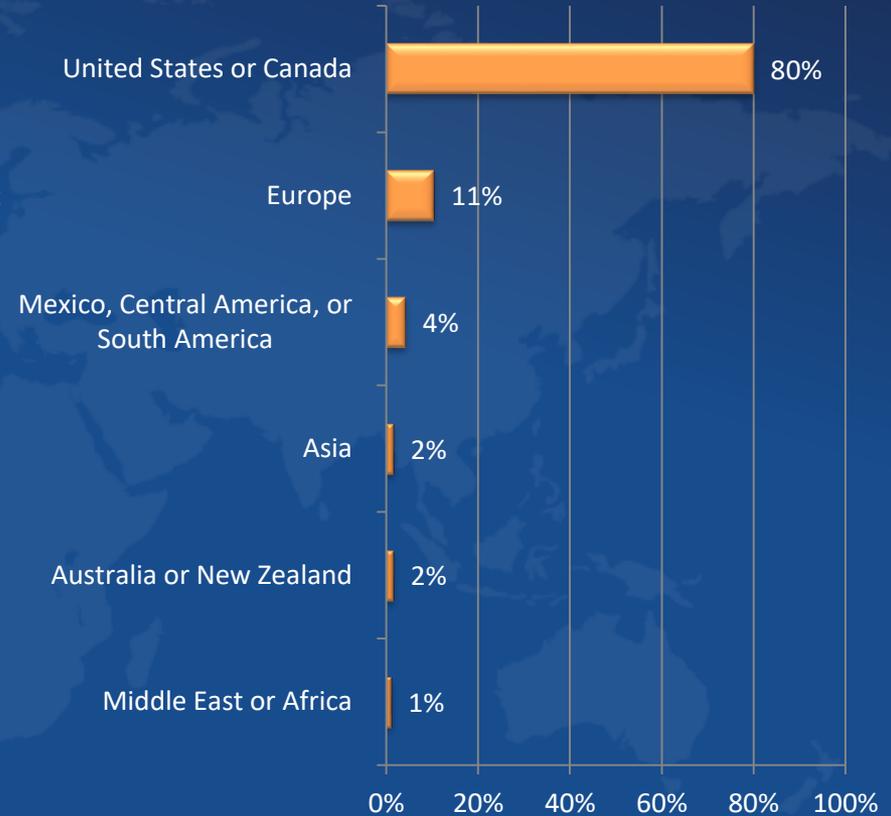
Individuals Represented

Leadership Role

- Team Manager 43%
- Executive 27%
- Frontline or Admin 30%

AI / ML Role

- Business stakeholder for AI or ML projects 7%
- Working with AI or ML is my entire job 18%
- Working with AI or ML is a substantial part of my job 51%
- Working with AI or ML is a minor part of my job 24%
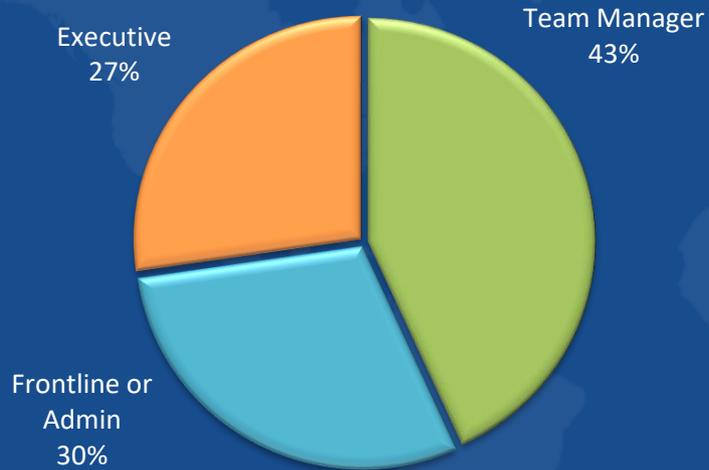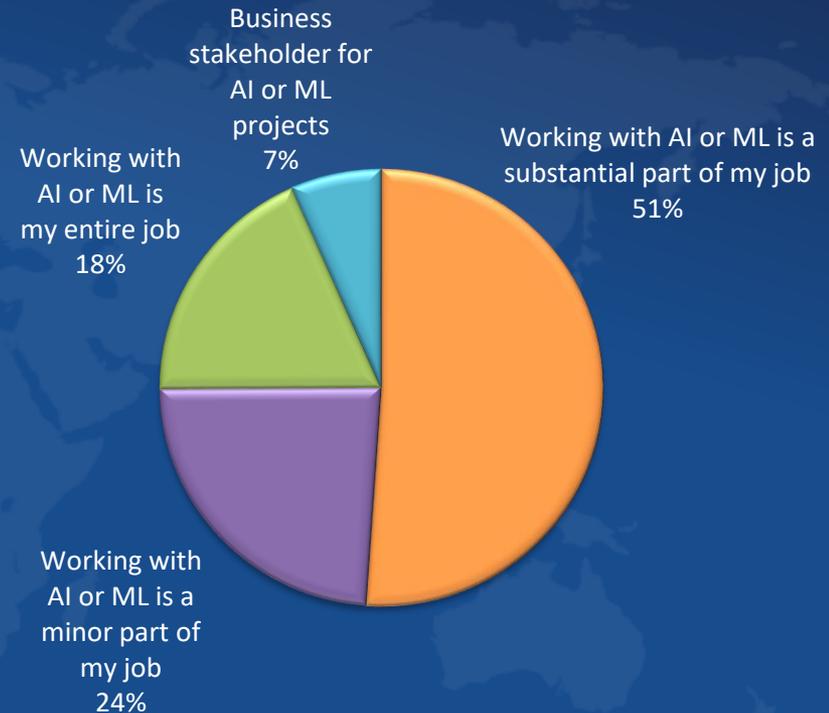
Individuals Represented

Work Role

- Deep learning specialist 4%
- Natural language processing specialist 7%
- Computer vision specialist 7%
- Machine learning specialist 11%
- Product owner 23%
- Data scientist 48%

dimensional research

Sponsored by: ALEGION

For more information…

About Dimensional Research

Dimensional Research provides practical marketing research to help technology companies make smarter business decisions. Our researchers are experts in technology and understand how corporate IT organizations operate. Our qualitative research services deliver a clear understanding of customer and market dynamics.

For more information, visit www.dimensionalresearch.com.

About Alegion

Alegion is the ML training data partner of the Fortune 1000. We offload the entire burden of training data from enterprise data science teams, from delivering custom training datasets to providing machine- and human-scored model testing and offering post-production exception handling. Our solution combines a purpose-built technology platform with a global pool of on-demand data specialists, driven by our own managed service team. We support machine learning projects broadly , with particular emphasis on Computer Vision, Natural Language Processing and Entity Resolution, in financial services, retail, defense, technology and manufacturing.

For more information, visit www.alegion.com.

dimensional research

Sponsored by: //ALEGION

APPENDIX

74% Are Using AI and ML Projects to Disrupt the Market Place

How is your company measuring AI or ML project success?

| Category	Percentage
Improved customer experience	66%
Reduced costs	54%
Competitive differentiation	51%
Increased sales	45%
Develop internal skills	32%
Industry disruption	23%
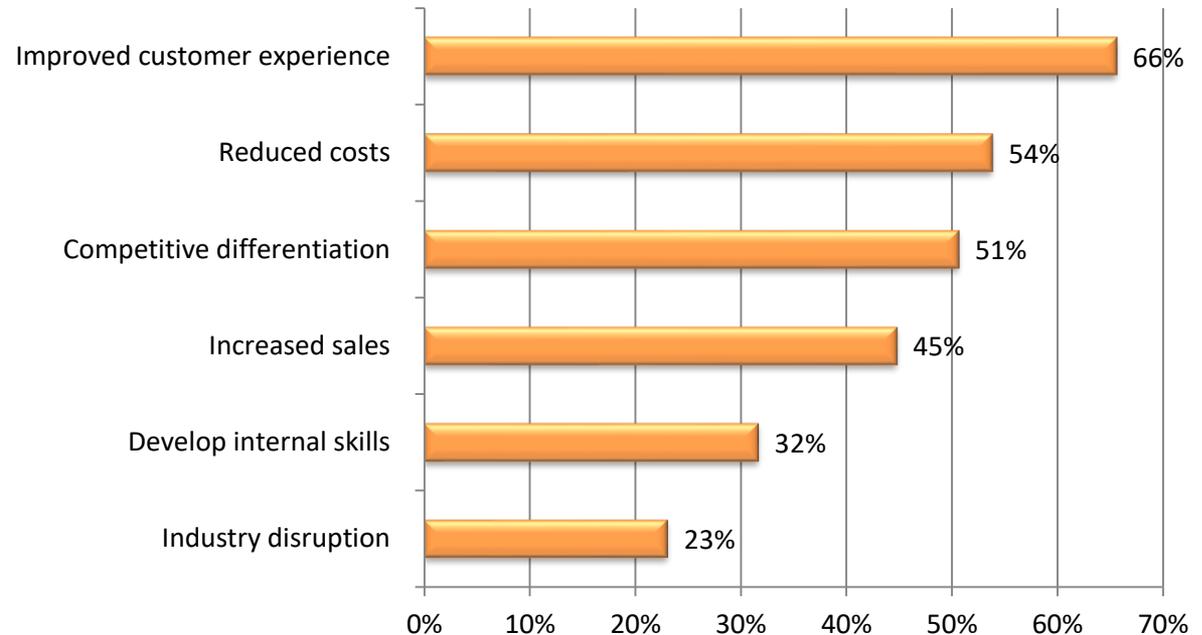
AI and ML Projects Require Continual Commitment

Approximately, what percent of the resources used to develop and deploy the AI/ML solution are still working on that project?

dimensional research

Sponsored by: ///ALEGION
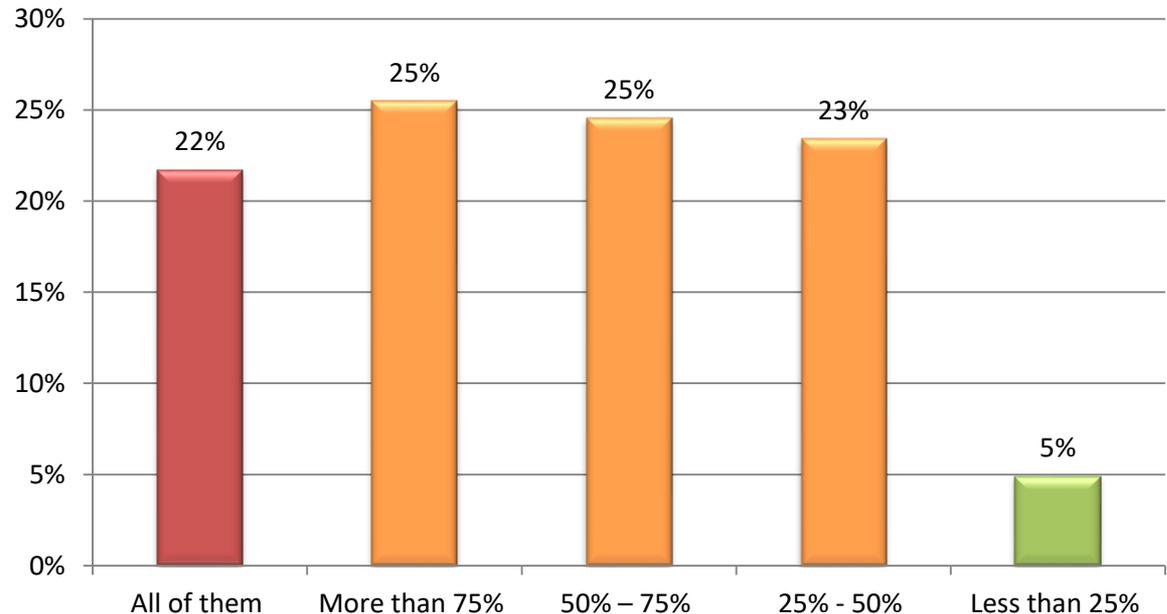
82% Are Satisfied with Project Progress

In general, has your company's most recent AI or ML project been successful thus far?

No
18%

Yes
82%
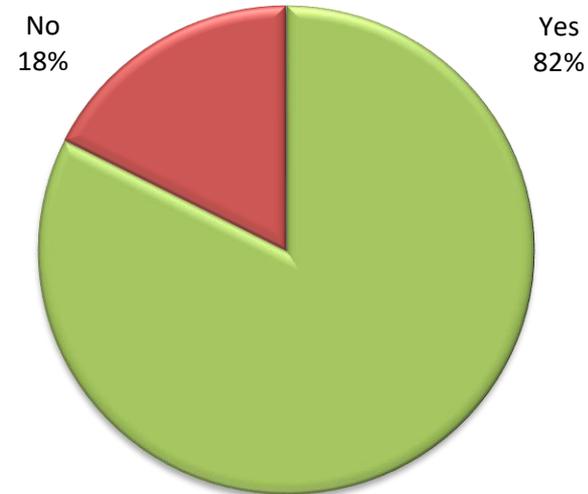
Lack of Experience Impeded Most AI/ML Projects

In your opinion why wasn't the AI or ML project a success?

Reason	Percentage
Project was more complicated than expected	53%
Lack of expertise	53%
Problems with the data	33%
Model confidence wasn't high enough	30%
Budget	27%
Shortage of personnel	23%
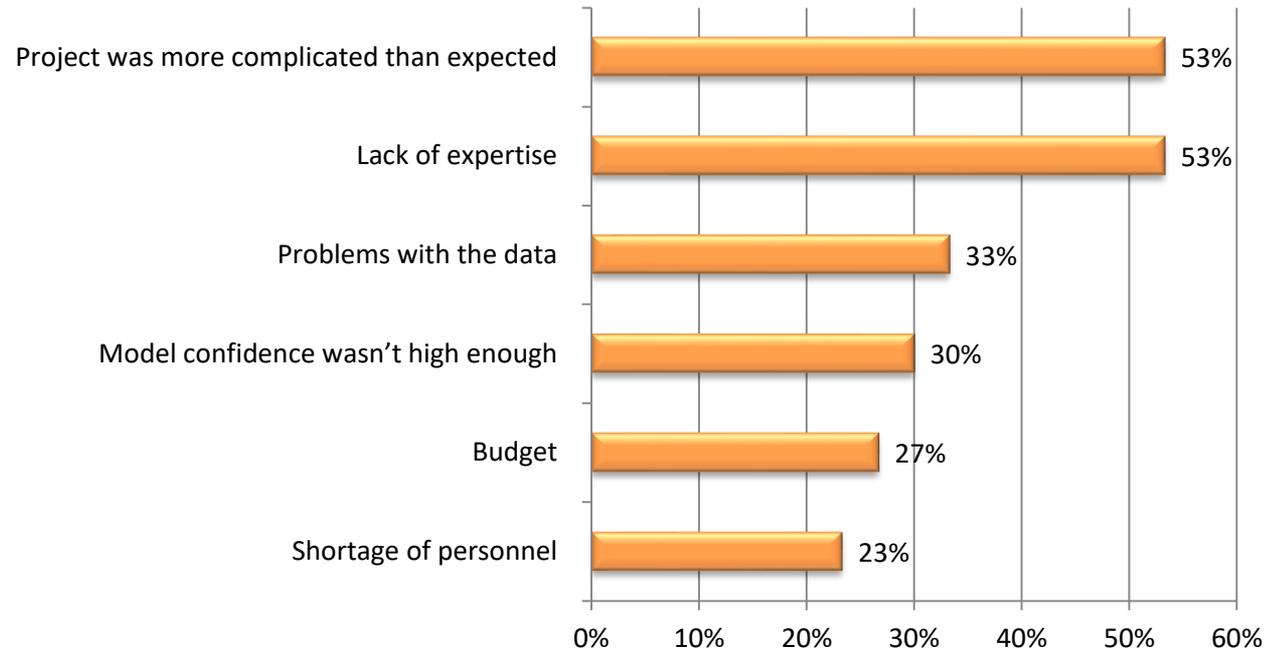
Majority of ML Projects Trained on Text Data

What types of data are being collected in your most recent project?

Data Type	Percentage
Text	77%
Images	42%
Voice/Audio	25%
Video	22%
Other	11%
We haven't start collecting data yet	4%

dimensional research

Sponsored by: ALEGION