accusoft

The Future of Software Development and Machine Learning

Seventy-six percent of companies surveyed plan to prioritize machine learning (ML) and artificial intelligence (AI) deployments in 2021. Machine learning is the use of statistics-driven algorithms to help software tools "learn" by training them to recognize and respond to patterns in data. While the concept has been around for decades, improved compute power and on-demand connectivity has significantly improved its scale and scope.

Robust ML tools pave the way for larger-scale AI integration, making it possible for companies to offload key tasks — such as data collection and process automation — to "intelligent" machines with the capacity to reduce data collection error rates, take appropriate action based on data inputs, and predict potential market trends using tested and validated methodologies.

Specifically, machine learning offers a way to streamline the software development process by empowering better connections between development and operations teams. The effective application of ML tools makes it possible to bridge the gap between legacy and modern applications by removing friction and abstracting key processes.

76%

of companies surveyed plan to prioritize ML and AI deployments in 2021.

THE TYPES OF MACHINE LEARNING

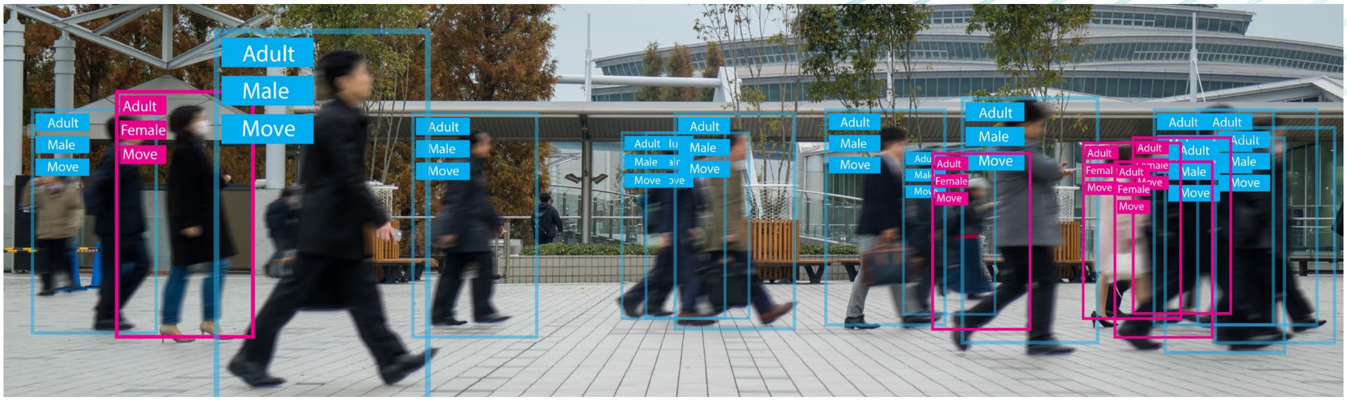
Machine learning can be broken down into several different subcategories, each of which can be deployed in different situations.

Supervised Learning

Supervised learning models use classified and labelled data to guide algorithm learning. By classifying and labelling data, developers are able to control both ML inputs and outputs to ensure specific patterns are recognized and outputs align with expectations. Supervised learning requires the most time and effort from data scientists and developers to control, curate, and capture data sources.

Unsupervised Learning

Unsupervised learning, meanwhile, provides ML algorithms with unlabeled and unclassified data and allows the algorithm to identify patterns based on unique data characteristics. While this type of learning is monitored to observe the results, developers don't interfere with the learning and pattern recognition process itself, instead waiting to see what outputs the algorithm will produce, then evaluating them for accuracy and modifying code as needed.



Semi-Supervised Learning

Semi-supervised learning frameworks combine aspects of both supervised and unsupervised learning to underpin specific pattern recognition methodologies including transfer learning, active learning, and reinforcement learning.

Transfer learning refers to the use of knowledge gained by ML algorithms in one task to help complete related tasks that have different data sets, in effect "transferring" the pattern recognition output from one problem to another.

Active learning allows ML algorithms to improve output accuracy by querying additional sources for information rather than relying on unmarked datasets alone. In some cases, algorithms are able to use connected applications such as search engines to refine their knowledge, while in others they actively engage with users for input. Perhaps the most familiar example of user-led active learning are chatbots, which are now commonplace on eCommerce product and service websites. These chatbots are in fact machine learning algorithms that leverage user input to guide their response, such as providing price or service information, directing users to a specific webpage, or connecting visitors with a live agent for further assistance.

Reinforcement learning uses a behavior-driven approach to "reward" desired behaviors and "punish" undesired behaviors to improve output accuracy. In this approach, ML algorithms are allowed to take unsupervised actions and discover patterns through trial and error. Desired outcomes are assigned positive values and undesired outcomes are assigned negative values; machine learning tools are then designed to prefer positive over negative values, allowing the algorithm to learn over time in response to choice-value outcomes.

Deep Learning

Deep learning is a subset of machine learning that mimics the basic structure of human neural networks to improve pattern detection results and data outputs. This outcome is achieved by creating connected layers of simply computational nodes that work in unison to analyze data and return predictive results. Work on deep learning neural networks is currently in its infancy — and these networks remain easy to derail — but the use of deep learning offers a way to increase the impact of machine learning without increasing complexity.

THE 5 YEAR OUTLOOK

As ML and AI technologies evolve over the next five years, enterprises should anticipate advancements in a number of key areas.

Increasing Use of RPA

Robotic process automation (RPA) refers to the use of software "robots" to automate key functions, improve overall output speed and significantly reduce error rates. The challenge? These robots aren't particularly intelligent — ensuring they're lightweight enough to minimize performance disruptions means limiting their ability to do anything except the specific task they've been programmed to complete.

Machine learning offers a way to connect the dots on RPA by allowing companies to effectively infuse software robots with specific intelligence that pertains to their primary process. Digitalist Magazine describes a straightforward example in the form of accounts payable. First, an RPA function opens relevant corporate email accounts and identifies messages that pertain to invoicing or purchase orders.

Then, these attachments are sent to ML tools that have been trained to recognize multiple invoice types and the specific responses for each one. Once relevant data has been extracted, invoices are sent back to RPA tools for filing and storage. The use of RPA/ML tools for this process also makes it easy for companies to monitor and manage the use of financial data at scale, which is now critical to ensure regulatory compliance. Forms processing integrations and optical character recognition (OCR) data capture tools can extract financial information from structured forms (such as loan applications or tax forms) and pass them along to RPA bots or ML algorithms for additional processing.

While advancements around RPA are ongoing, expect this technology to gain significant ground over the next five years as the price of cloud computing continues to fall and the performance of always-on infrastructure continues to improve.

Expanding Impact on Security

Improved IT security is now a top priority for businesses worldwide — from phishing attacks to ransomware threats to targeted business email compromise campaigns that rely on current news events to deceive users, companies can't afford to ignore the potential impact of attacks on their revenue, reputation, and IT reliability.

The result? A kind of arms race between attackers and defenders as cybercriminals look for new ways to infect corporate networks and IT teams search for ways to streamline the security process. But with malicious actors only concerned with finding new weak points while IT teams are tasked with defending systems, enterprises are often at a disadvantage.

Machine learning can help level the playing field. By training ML tools using internal, external and emerging threat data, it's possible for companies to create more robust and response security postures capable of proactively identifying and responding to threats before they compromise key systems.

As noted by CRN, these next-generation cybersecurity tools can collect key data from internal transaction systems, communication networks, online user actions, and external public sources to facilitate pattern recognition and threat detection that's dynamic rather than static. Over the next few years, enterprises should anticipate the adoption of AI and ML cybersecurity solutions to transition from early adopter initiatives to security table stakes.

Providing good data for these solutions to work with requires versatile content processing tools. For instance, OCR integrations could be deployed to extract content from scanned images and convert it into searchable text. For hand-printed documents, intelligent character recognition (ICR) can be used to quickly generate content that can be easily digested by ML applications to expand data sets. Optical mark recognition (OMR) may also be needed to quickly review the results of testing and survey form data.

Democratizing Digital Intelligence

While initial AI and ML developments were largely focused on proof-of-concept rather than practical use, IT World Canada notes that increasing democratization of analytics initiatives is now paving the way for more effective ML deployments. Data is often described as the "new oil" — the resource no company can do without, but demands effective collection, storage, and analysis to deliver actionable results.

In much the same way that ML algorithms can be taught to recognize key cybersecurity patterns or improve RPA processes, they can also be used to shorten the distance between users and digital intelligence. In effect, machine learning solutions can now handle the heavy lifting of data analytics and pattern recognition, in turn closing the gap between disparate data sources and the ability of users to extract actionable information and apply it at scale.

Building these solutions can be challenging due to the sheer volume of data required. When Accusoft launched an ML initiative in 2019, the engineering team quickly realized that the available data wasn't robust enough to make the system as good as it could be. Rather than bringing in outside contractors to perform the thousands of hours of work needed to "teach" the ML system, the team instead asked other departments within the company for help. Over the course of four days, dozens of people volunteered to help create data examples that would teach the algorithm to make better decisions.

Consider the implication for C-Suite executives tasked with creating corporate strategy. Historically, C-Suite members have relied on a combination of personal experience and available market data to make critical decisions. As data sources have rapidly expanded and market forces have become more complex, however, the efficacy of this method has been significantly reduced. But since most C-Suite members are not IT experts, the divide between data and decision making has been difficult to bridge.

Automated processing and data capture integrations are incredibly effective at collecting information and diverting it to the proper destination, but raw data alone often isn't enough to make meaningful decisions. By combining these tools with ML applications capable of analyzing and drawing insights from the data, it becomes possible to generate more accurate and actionable datasets that C-Suites can use to inform strategic planning. When it comes to data democratization, companies should expect slow and steady adoption as ML-integrated tools prove their worth over the next few years.

Assessing Ethical Implications

Unconscious bias remains a critical issue in the development of machine learning algorithms. This is especially problematic in areas such as human resources (HR), where ML and AI tools may be used to inform or initiate hiring, promotion, or termination decisions. If it comes to light that the algorithms used were biased based on characteristics such as age, race, gender, or other factors that don't relate to employee performance, the results could be disastrous — companies could find themselves facing legal and regulatory challenges that are both costly and time-consuming.

As a result, organizations should expect increasing focus on the ethical implications of AI and ML. This includes ongoing assessment of the data they're using to make decisions, how these decisions are made and if they represent a pattern of action that is potentially problematic. This ethical effort relates back to both the validation and testing stages of ML training. Organizations will need to either hire machine learning experts or work with reputable third party providers to ensure that the data and variable weightings used create reliable results that are absent any implicit or explicit bias.

THE SOFTWARE SOLUTION

It's also important to contextualize the role of machine learning as a subset of the larger software development framework. Machine learning solutions offer a few key advantages on this front.

Continual Application Assessment

Applications don't exist in isolation. Instead, they function simultaneously across multiple frameworks, from user-facing mobile devices to internal storage databases and on-demand cloud services. As a result, even small changes in operational conditions could lead to significant problems for software — functions may suddenly stop working when new code is added, or zero-day threats may be detected when additional features are rolled out.

For software development teams, machine learning algorithms offer the opportunity to continuously assess and evaluate applications both in the development pipeline and once they're actively in use. By combining high-quality data with robust training methods, enterprises can create pattern priorities capable of detecting the hallmarks of minor issues before they become significant problems. In practice, this allows businesses to address these issues immediately and with minimal disruption to current operations.

When contrasted with the time and effort required to remove software from live service, reevaluate critical code, and conduct complete regression testing, investments in ML are significantly more cost-effective over time.

Improved Data Integration

Gone are the days of "waterfall" modeling and development — now, agile application design, testing, and deployment is the expectation. But as data volume and variety increases, the ability of teams to keep pace with agile development expectations begins to suffer, especially as the ratio of structured to unstructured data begins to shift.

While structured data — information stored in familiar, regimented formats such as spreadsheets or tables — has historically comprised the bulk of enterprise information, industry predictions now suggest that by 2025, just 20 percent of data will fall into this category. Unstructured data will account for the remaining 80 percent — data that includes everything from sentiment analysis to customer surveys to natural language processing (NLP) — in effect, any data source that doesn't conform to standard structure but instead relies on patterns that aren't immediately apparent.

80% **of your data will be unstructured in five years**

To make best use of this expanding resource for effective application development, machine learning is critical. Where traditional tools are capable of culling and collecting data from structured data sets, ML algorithms can be trained to detect patterns in unstructured resource pools, in turn empowering more effective integration with applications during the development process. In practice, this improved integration can help shorten the distance between application design and deployment, and increase agile production speed without sacrificing structure.

Take, for instance, the way convolutional neural networks (CNNs) are used to aid in image processing. These programs learn how to identify large images over time by repeatedly analyzing images pixel by pixel and taking note of the relationships between them. This ability to learn patterns over time and draw conclusions is one of the most powerful applications of ML and can be applied to a variety of use cases in a mature data collection application.

Service Integration at Scale

The ultimate goal of machine learning is to facilitate the development of digital algorithms and networks capable of mimicking key aspects of human cognition and decision-making. But achieving this goal depends on more than improved use of data sets to deliver reliable results. It also requires the ability of ML tools to integrate and interoperate with other solutions at scale. Just as human brains are capable of considering and connecting disparate information frameworks to produce a cohesive whole, effective deployment of ML and AI solutions relies on additional software tools that perform specialized tasks.

This is reflected in applications such as the [integration of PrizmDoc Editor] into the larger LegalSifter solution, which empowers automated contract creation. LegalSifter's powerful search algorithm needed tools that could extract text from Microsoft Word documents programmatically and then determine whether or not the contract contained the necessary clauses.

From there, the LegalSifter solution also needed a way to insert comments or missing clauses into the document at the appropriate locations along with any recommended text. That content came courtesy of the application's powerful AI engine, which was able to review contracts and immediately determine whether additional clauses needed to be included or if language needed to be changed. In cases where new text had to be inserted, the LegalSifter solution could easily utilize PrizmDoc Editor's assembly tools to seamlessly drop content into the appropriate places.

After assembling the new document, the application had to be able to track changes and comments during the review process before the contract was finalized. PrizmDoc Editor's document assembly and programmatic editing features integrated smoothly with LegalSifter's application, empowering its innovative technology to deliver truly effective contract automation.

Enhancing On-Demand Response

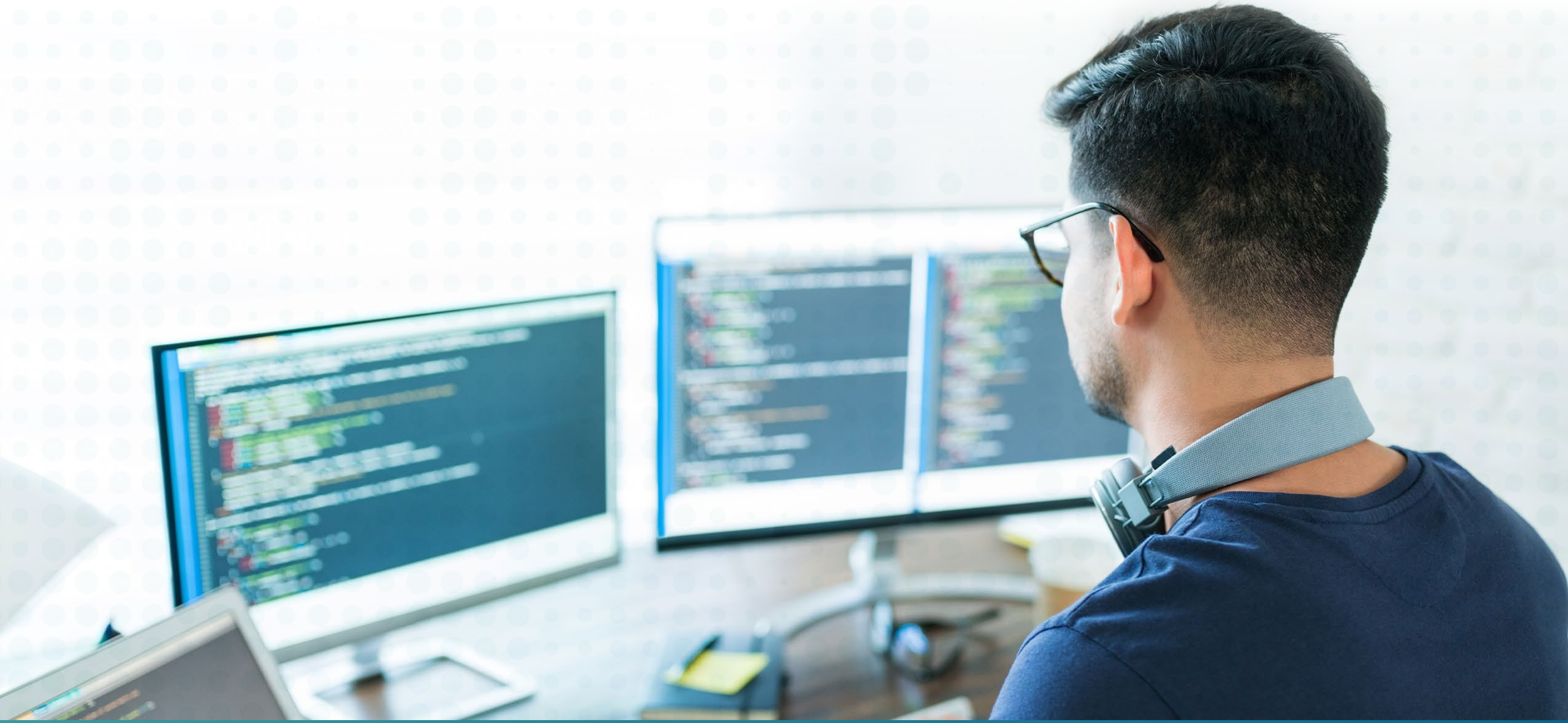
Prediction underlies the promise of ML algorithms. With pattern recognition comes the potential to create software tools capable of predicting future trends using both current and historic data sets. In practice, this could take the form of everything from small-scale predictions about emerging supply and demand curves that may impact inventory management and purchasing decisions to evolving needs around staffing levels and specific expertise.

At a larger scale, predictive tools can help empower on-demand response to worldwide events such as economic upheavals or potential pandemics. By equipping development teams with the tools and technologies they need to create and curate new machine learning frameworks, companies can better respond to — and withstand — rapid market changes.

Open-Source Opportunity

There's also an expanding opportunity for companies to create and curate their machine learning initiatives using open-source solutions. Much as the shift from on-premises computing to cloud-based resource use has helped highlight the need for increased interoperability across operational platforms, the increasing ubiquity of machine learning deployments has paved the way for the ongoing development of open source libraries and frameworks that software developers can use to streamline the process of algorithm creation, testing, and deployment.

Developers can also benefit from deploying proven SDKs and APIs backed by decades of experience and dedicated support. Unlike open source solutions, these integration frameworks are being continuously updated to minimize security risks and add innovative new features that could unlock entirely new ML use cases.



WHAT'S NEXT FOR MACHINE LEARNING?

Machine learning is branching out. Powered by more robust cloud resources and the rise of interoperable, open-source initiatives, there's no shortage of opportunity for ML tools to help companies deliver on the promise of RPA, democratize data analytics, or deploy integrated AI solutions that help drive critical processes such as document automation and creation.

Discover how Accusoft is partnering with our customers to help them leverage the potential of their machine learning tools with our data capture and image processing capabilities. **Get in touch today.**

accusoft®

www.accusoft.com