Infinia ML℠

ML For Enterprise:

Experts Describe How To Unlock the Power of AI For Your Organization

contact@infiniaml.com (984) 227-8480 www.infiniaml.com

Introduction

"Data is not useful until it becomes information."

— Seth Godin

The revolution in machine learning (ML) is happening all around us. It's just one dimension of the larger field of artificial intelligence, which promises to transform the way we live and work.

But ML isn't just the domain of tech titans. It's the backbone of everyday business practices in companies of every size.

According to McKinsey's The State of AI in 2021 survey, over half of all polled companies have adopted AI into at least one function at their organization.

It's more accessible than ever for organizations new to ML to harness the power of their own data to maximize efficiency, gain new insight — and grow.

But data itself can't do that. Data doesn't equal information. It must first be processed before it can be understood.

ML can suggest connections, questions and avenues for innovation humans could never perceive — there's just too much data to sift through.

ML isn't a magic bullet. It's a process. And it can be difficult to get it right. The first step is to achieve a state of "data liquidity" by transforming text, tables, and spatial information into organized, usable data from which machines can learn. That's a complicated task, but essential for making ML work for your business.

As more and more companies adopt ML, the possibilities expand exponentially — and so do the choices. In this white paper, data scientists, engineers and thought leaders who deeply understand machine learning weigh in on what it takes to deploy ML for small to medium-sized businesses.

We ask some very human questions, like whether to build or buy your ML solution, why offshoring isn't always cheaper, and what the future holds for AI-enabled technologies.

Table of Contents

Build or buy? That's just one question

Machines aren't the only ones that need to learn

"It is imperative to start educating our leaders, our peers and our business partners on AI and the importance of AI. The question of whether a company should develop AI products internally or seek contractors is best made on a case-by-case basis."

— Himanshu Kalkar of Fidelity Investments

Machine learning isn't a one-time purchase or a one-size-fits-all solution. Whether you develop ML solutions in house or outsource them to a specialized firm, it's an ongoing investment in technology.

To help identify the best choice for your organization, Himanshu suggests asking:

- What timeframe has leadership established?
- How much investment can you expect?
- Is the product you're trying to build core to the business?

The timeframe is particularly crucial, including the estimated time it takes to create the project and its overall shelf life, which will help determine how to allocate resources.

Understand the implications

"No matter what you're in the market for, you should almost always choose the best value instead of the lowest cost. Experts provide more than additional value. You want to ensure you're using automation responsibly. There's a balance between what is capable, what is possible and what we actually want to happen right now."

— Wayne Butterfield of ISG Automation

“If your business needs aren't yet met by today's technology, begin to develop small, manageable experiments to find out what you can build yourself. But beware, AI DIY-ers should remain flexible. Ideally, avoid building something so custom that it's impossible to maintain — or impossible to outsource to a market solution when one becomes available.”

— Paul Tepper of WorkFusion

“Like all new promising technologies, people view it as a panacea. But you can’t just buy DataRobot, or whatever machine learning tool you want, and invest millions of dollars in a quote-unquote AI initiative if you don’t deliver a business context to it. The application of AI in a business context is the game-changer.”

— Chris Lynch of AtScale and DataRobot

Usually, tapping the pros for their expertise results in exponentially more value than your internal teams can deliver through trial and error. You may be able to do it cheaper, but chances are you won't be able to do it as well as an ML specialist.

There's a notable exception, however:

If the AI product of your dreams is something you see using and expanding upon for many years to come, it might make sense to create the internal capacity to build and grow with it. But if you're considering a project that will only serve its purpose for a few years, outsourcing makes more sense.

Offshoring, onshoring and collaborating (even with the competition)

Outsourcing doesn’t always require offshoring

“When it comes to choosing between offshoring physical labor and 'digital outsourcing,' the cost-arbitrage play is becoming less and less effective.”

— Rob Delaney of Infinia ML

As labor costs grow globally, companies are looking for alternative options to offshoring, including AI, which Wayne Butterfield calls "digital outsourcing."

But working with data presents its own set of issues. Privacy concerns are more pertinent than ever, especially in Europe, where GDPR compliance is an industry unto itself. Regulation, however, doesn’t have to stifle innovation.

Onshoring and offshoring have their benefits and drawbacks, but there's another option.

Paul Tepper, who directs AI projects at WorkFusion, explains that some of his company's signature strategies — federated learning and differential privacy — are helping highly regulated industries reinvent themselves even when they have to color within the lines.

Weighing the options

"It's one thing to have a data scientist build a model that can make some kind of prediction. We're quite good at doing those kinds of things now, but to actually deploy that at scale, connected to all the different systems you need in an enterprise to actually make it something useful — that's where we come in. We can simplify that process. We've really thought about that full end-to-end pipeline.“

— Paul Tepper of WorkFusion

> *"Cost might be the number one reason a business wants to do offshore outsourcing. But cultural differences and regulatory requirements can present significant barriers to many companies. The advantage of offshore or onshore is actually not that clear to me."*
>
> **— Ya Xue of Infinia ML**

Choosing ML over human labor isn't without its pitfalls. There are disadvantages to both offshoring and onshoring.

For example, the implementation of more AI/automation may actually catalyze reshoring or an increase in onshore labor demand. The solution may in fact be both and.

In the age of AI, it's necessary to do things differently. The more data you can collect, the more you can share in what ML can offer.

"You piece all these things together into one workflow or one process that is optimized," Rob adds.

The present and future of NLP and document processing

Where is that thing?

"In my day-to-day, I'm asked a lot of questions. All of these different bits of information come to me and I don't always remember exactly the medium from which they came. Sometimes I'm sort of 'hunting and pecking' through our Slack channel, or Google Drive or my email. But with Onna, I have the opportunity to search across all of those different applications to more easily pinpoint what I was looking for."

— Kelly Griswold of Onna

Machine learning has already paved the way for a number of miraculous technological possibilities. The ability to search for information across multiple platforms and apply AI to voice technology are just two of the countless examples of how ML is changing business capabilities.

And when our understanding of the definition of documents expands, the possibilities seem endless. Contracts, emails and Slack messages have the potential to reveal new wisdom when aggregated and mined at scale. One day — in a future we can truly foresee — NLP and document processing will be able to ask questions, not just provide answers.

What is a document, anyway?

"It's a philosophical question. And for us, it's an existential question about how we drive our business."

— David Parmenter of Adobe

There are a number of ways to define a document, David explains. The most straightforward, perhaps, is that it's "a piece of information in a container that represents something that you wanted to say at a given time. In that sense, a PDF is the very encapsulation of a document. Certainly, there are trillions of them in the world that back up that statement empirically."

But there are other ways of thinking about it. Take the invoices your landlord sends over the course of a year: Are they 12 documents or one document?

"My accountant would say they were one document, because that was what I needed for my taxes for that year," says David.

Here's another scenario: If you write something on a blog, edit it, revise that version and then edit it again, are you dealing with four documents or one document?

"I think the answer to all of those questions is yes," he adds. "Clearly it goes beyond the bounds of what we think of as a PDF. It can be a collection of documents. It can be a living document. All of these things are important."

A new (illiquid) dawn

"If your data is illiquid, if it's stuck in systems or in [unstructured] documents, it is impossible to actually use machine learning. For companies that want to implement machine learning, data liquidity — meaning that the data is structured in a way that machines can actually use it and access it — is mandatory."

— Rob Delaney of Infinia ML

"One of the holy grails [of document processing] is some kind of universal form for the content that could then easily be transmogrified into other forms. And this is a solvable problem. There could be a canonical form."

— David Parmenter of Adobe

"I can now take a single podcast and convert it programmatically into 38 different languages, but in a hyper-realistic synthetic Voice as a Service (VaaS) solution that allows anyone to create, manage, share and monetize their voices. Everybody's got to start to think of their digital twin in the metaverse and a lot of this is going to be driven by AI and machine learning."

— Ryan Steelberg of Veritone

Transmogrification and data liquidity are two exciting applications of ML-enabled technology.

These two forthcoming technological advancements will allow users to create multiple versions of one piece of content for a variety of audiences and purposes and give businesses data-supported insight from company information previously inaccessible to machines.

Understanding (and navigating) structured and unstructured data

What's your type?

From a data-processing perspective, there are three different types of documents:

- **Structured:** An Excel spreadsheet is a classic example of a structured document, the type that is most easily applied to machine learning. Information is organized in fixed, pre-defined fields, making it easy for people to categorize chunks of data and for algorithms to find patterns.

- **Unstructured:** PDFs and other text-based files are considered unstructured data. Until recently, AI developers didn't see them as valuable for generating data. Companies have historically discarded the information for data-generating purposes. However, AI is rapidly changing this practice.

- **Semi-Structured:** Contracts are a good example of semi-structured data. The boilerplate language of terms and conditions is unstructured, but contracts also include templatized fields like dates and parties.

Big data, big challenges

"Obviously, those functioning at a scale like Google, Amazon and Adobe — we try [to leverage data]. But down at the business level, you have a smaller universe of things you have to look at. Simply the documents on your website, or the documents in your internal data store or every email that everybody sent to everybody else. That's a smaller world. But it's still largely unstructured. We have research in-house that demonstrates our belief that on average, about 80% of all the documents in the world are unstructured or semi-structured. That's just a goldmine of information."

— David Parmenter of Adobe

Let's say there are 2 trillion documents in the world. Even 80% of 2 trillion is a mind-bending amount of data. Few companies have the time to sift through such volumes of information and look for patterns to guide decision-making.

By understanding how documents are structured, businesses are better able to identify what technology can best draw valuable insights.

Structured or unstructured? That’s the question

“Unstructured data may not be in the format you want, may not be readily available and may require data cleanup. It’s a great time commitment.”

— Himanshu Kalkar of Fidelity

"Our goal is to convert free text because it's a structural format that can be easily used by human beings. The distinction is useful to explain what we are doing and why it’s meaningful. We can then extract the useful information from the free text and convert it into structured data. This means the data is in a format with a clearly defined data type and with a pattern that can make it easily searchable.“

— Ya Xue of Infinia ML

“Truly cutting edge technology is pushing the boundaries and going beyond structured and semi-structured documents to be able to find the needle in a haystack in large documents.”

— Wayne Butterfield of ISG Automation

For decades, financial institutions have applied ML to produce structured financial data in documents like spreadsheets, but ML applications are now able to tackle unstructured data like emails, texts and chats.

Structured data is still the best choice for companies just beginning to use the technology because it is concrete, specific and organized, offering more ways to solve problems easily and quickly.

However, business leaders should keep an eye on ML advancements. Tapping semi-structured and unstructured documents for insight is an emerging technological capability that is becoming increasingly affordable for small to mid-sized businesses.

What it takes to successfully apply ML

You can beat the odds

According to a study by Gartner, only 20% of ML projects actually make it to the implementation stage. Those that do are profitable only about 60% of the time.

But Gartner also predicts that by 2024, 75% of organizations will "shift from piloting to operationalizing AI," says VentureBeat. That's because the tech is evolving fast. There are more ways to collect data and better ways to use it. How can your organization stay ahead of the curve?

Big data, big challenges

"Our research study revealed an important aspect of AI that is often overlooked: Just because AI is applied doesn't mean that the result will be the same. Humans are still filtering AI inputs. The human factor can't be discounted. There are some personal characteristics that play a role in determining how humans actually perceive AI."

— Christoph Keding of ESCP Business School

Christoph and co-author Philip Meissner identified three types of decision-makers in their study, "The Human Factor in AI-Based Decision-Making," published in the MIT Sloan Management Review:

- Skeptics: Individuals who resist information provided by AI technologies and are likely to ignore the AI-generated data when making decisions.
- Interactors: Individuals who are open to AI-generated data but filter decision-making with their knowledge and personal experiences.
- Delegators: Individuals who fully trust AI technology and default to AI-generated insight over their own opinions and ideas.

What can business owners do with these preliminary findings?

- Use larger teams to make decisions, a strategy that balances out individual biases.
- Approach findings similarly to how the Meyers-Briggs test is used — know who the skeptics, interactors and delegators are on a team.

Speak your stakeholders' language

“Communication between engineers and high-level business leaders is a two-way street.”

— Amberly Rowles-Lawson of Infinia ML

Her advice:

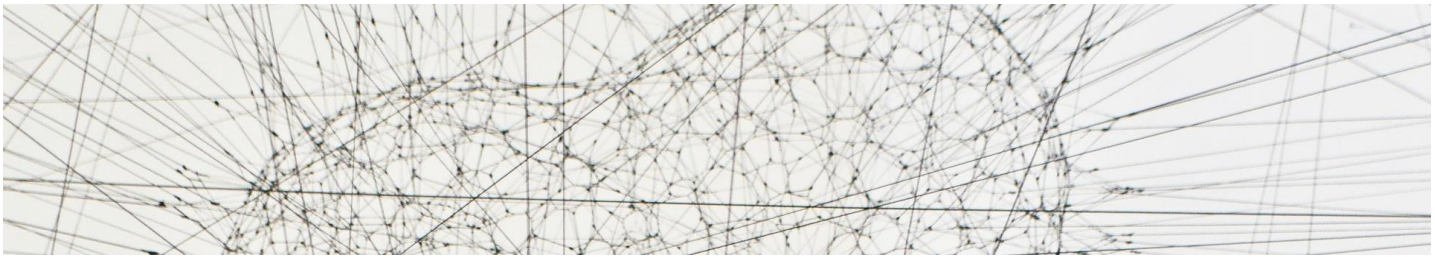
- Business leaders should take time to dive deeper into the everyday tasks of engineers to better understand clients' challenges.
- Engineers need to adjust the way they present information so it makes sense to business leaders, which encourages leaders to act on recommendations.
- Understand exactly where your stakeholders are coming from in order to give them the information they need.

Human experience is still paramount

“We’ve designed work to be completed by the greatest computer on the planet — the human brain."

— Wayne Butterfield of ISG Automation

Wayne categorizes jobs into two categories: "hand work" and "head work." Historically, intelligent automation shined in its abilities to automate tasks that typically fall in the hand work category. Developing intelligent automation tools to replace head work presents a more complex challenge for data scientists.



Strategize to support your goals

"Identify your goals and how data can inform decisions that support your goals when combined with AI."

— Kelly Griswold of Onna

For machine learning to be used well, business leaders need to think carefully and strategically about the data they already have and the data they collect moving forward. Then explore these important questions:

- What data do we already have that can inform our decision making?
- Is the data organized to take advantage of the AI at our disposal?
- Can our data assets be better organized?
- What additional data would give greater insight into our decision making?
- How do we collect that data?
- How do we best organize that data?

Conclusion: Is AI smarter than a baby?

"AI is not about making computers intelligent, as much as it is about making them less dumb."

— Paul Tepper, WorkFusion (quoting a professor he once studied under)

"When you think about it, a lot of the things we do anyway, these simple rules and algorithms, are not particularly smart in and of themselves," Paul adds.

"And even the best AI today doesn't even compare to a one-year-old in terms of its ability to learn and its intelligence. So for me, making automation intelligent really makes it more flexible."

To extend the analogy, we can safely say that machine learning (and AI overall) is in its infancy. So there's much to learn, room to grow — and infinite possibilities. Now's the time to discover what this technological revolution can do for you.

Meet the experts

Featured guests from Season 3 of Machine Meets World

Wayne Butterfield, Global Head of Intelligent Automation Solutions, ISG
A pioneer in the world of automation, Wayne was an early adopter of RPA (robotic process automation) for both business processes and customer-centric operations, such as virtual assistants and chatbots. In his role at ISG, a global technology research and advisory firm with a specialized banking and financial services division, he develops cognitive reasoning, digital customer service and intelligent automation for clients across the US and Europe.

Rob Delaney, CEO, Infinia ML
Rob Delaney is a Director at Carrick Capital Partners and Infinia ML's CEO. He received his MBA from Harvard Business School. Rob previously worked for Accenture as a consultant for venture capital and technology clients and as a corporate mergers and acquisitions specialist.

Kelly Griswold, Chief Operating Officer, Onna
Kelly Griswold has served as Onna's COO since October 2020, having originally joined the company as Chief Strategy Officer. Kelly previously led the go-to-market function at Knowable (which was spun out of Axiom). Prior to Axiom, Kelly worked in finance and corporate development roles in renewable energy.

Himanshu Kalkar, SVP, Head of Intelligent Automation & Data Products, Fidelity
Himanshu's experience includes managing enterprise-scale global software, technology and advanced analytics solutions; implementing emerging technologies to support data-driven operating models; and aligning business and AI strategies. He has an MBA from Duke and as well as an MS and BS in Engineering.

Christoph Keding, ESCP Business School Berlin
Christoph is a research associate at ESCP Business School and a visiting scholar at the University of California, Berkeley. With Philip Meissner, he is the co-author of "The Human Factor in AI-Based Decision-Making," published in the MIT Sloan Management Review.

Chris Lynch, Data Robot, AtScale
Chris is the CEO & Executive Chairman of AtScale. Prior to AtScale, Chris co-founded and was a General Partner at Accomplice, a venture capital firm investing in early-stage tech companies. Chris has held leadership roles at tech startups like Vertica, Acopia Networks, and Arrowpoint Communications.

Philip Meissner, ESCP Business School Berlin
Philip is the Chair of Strategic Management and Decision Making at ESCP Business School in Berlin. He also serves as Academic Director of ESCP Europe's Master in Strategy and Digital Business. With Christoph Keding, he is the co-author of "The Human Factor in AI-Based Decision-Making," published in the MIT Sloan Management Review.

David Parmenter, Director, Data and Engineering, Adobe Document Cloud
David leads data, analytics and machine learning for Adobe's Document Cloud. He has worked extensively as a lead, architect, manager, data scientist and consultant. In February 2017, David co-chaired Adobe's Technical Summit conference.

Amberly Rowles-Lawson, Vice President of Software Operations, Infinia ML
Amberly previously served in roles at Cisco Systems, where she led a team developing test automation and DevOps solutions for the company's virtual and cloud-native firewall offerings. Her background is in architecting product development and deployment workflows. Amberly earned an MS in Computer Science from William & Mary and a BS in Astronomy and Astrophysics from Florida Institute of Technology.

Paul Tepper, Head of AI, WorkFusion
Paul leads AI and ML product strategy for WorkFusion's intelligent automation platform. He has worked in computational linguistics and NLP Research, product management and ML/software Engineering.

Ryan Steelberg, President, Veritone
Ryan is a serial entrepreneur and investor who co-founded AdForce (IPO), 2CAN Media (Sold to CMGi), and dMarc Broadcasting (Sold to Google). He co-founded Veritone with his brother, Chad Steelberg, and serves on its Board of Directors.

Ya Xue, Vice President, Data Science Operations, Infinia ML
Prior to Infinia ML, Ya was a senior machine learning scientist for Align Technologies. She has also worked for Siemens Medical Solution, GE Global Research, and two startups. Ya earned BS, MS and Ph.D. degrees in electrical engineering from Tsinghua University, Arizona State University and Duke University respectively. She is the co-author of seven patents.