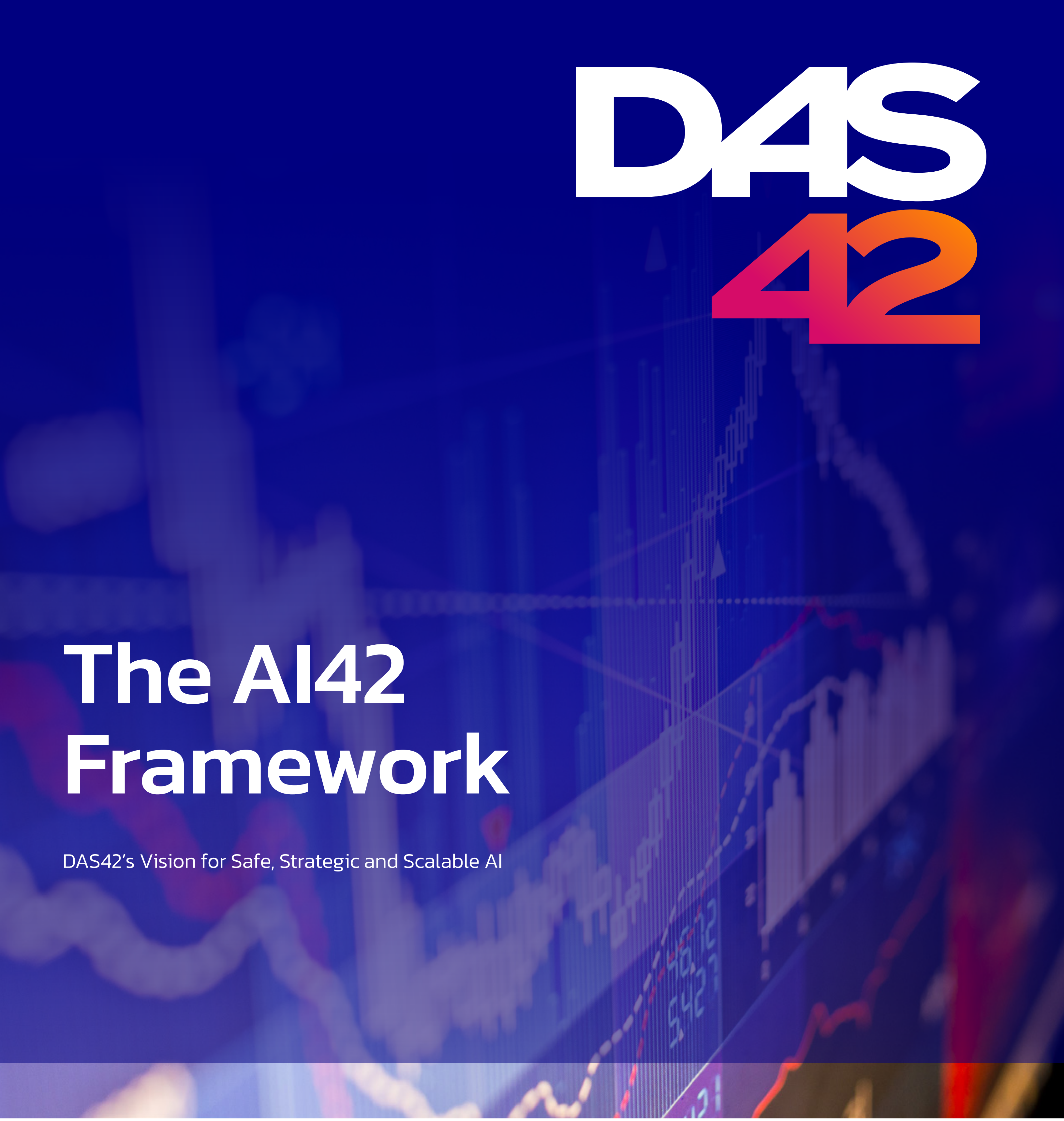


The AI42 Framework

DAS42's Vision for Safe, Strategic and Scalable AI

About the AI42 Framework

AI42 is both a philosophical and technical framework that is intended to guide organizations embarking on new artificial intelligence or machine learning initiatives. At its core are five pillars; Ethics, Strategy, Architecture, Governance, and Impact. These pillars encompass all of the key elements that DAS42 has identified as being crucial to consider and plan for when implementing AI or ML that affects stakeholders.

Table of Contents

SECTION 1

THE DAS42 ETHICS CORE

Automate in the interests of your people

At the heart of modern anxieties about artificial intelligence is what seems to many of us like a science fiction premise: the fear of "the robots taking over". Since the beginning of the Industrial Revolution, we have faced this sentiment repeatedly, in waves. No one who has developed a lifetime of workplace skills and experiences likes to think their specialized expertise could be replaced by a machine. Some worry about losing "the human touch", preferring or valuing interpersonal contact in ordinary settings like the grocery store checkout.

And yet, in sometimes dangerous industries like manufacturing and mining, we have become increasingly reliant on machines to alleviate the stresses and dangers these positions place on humans. How can we apply our learnings from the past century-plus of innovation to our current situation?

AI can do the things that are tedious or dangerous to humans: repetitive tasks; tasks that sap our energy. AI cannot be bored; AI cannot be tired; AI does not experience trauma. And, in some cases, AI can do the things that humans aren't good at: predicting churn risk for a large customer base using many differently-weighted factors; sorting through massive volumes of information for a single answer; identifying minute variances or patterns in CT scans to detect solid cancers in their earliest stages.

Automate thoughtfully

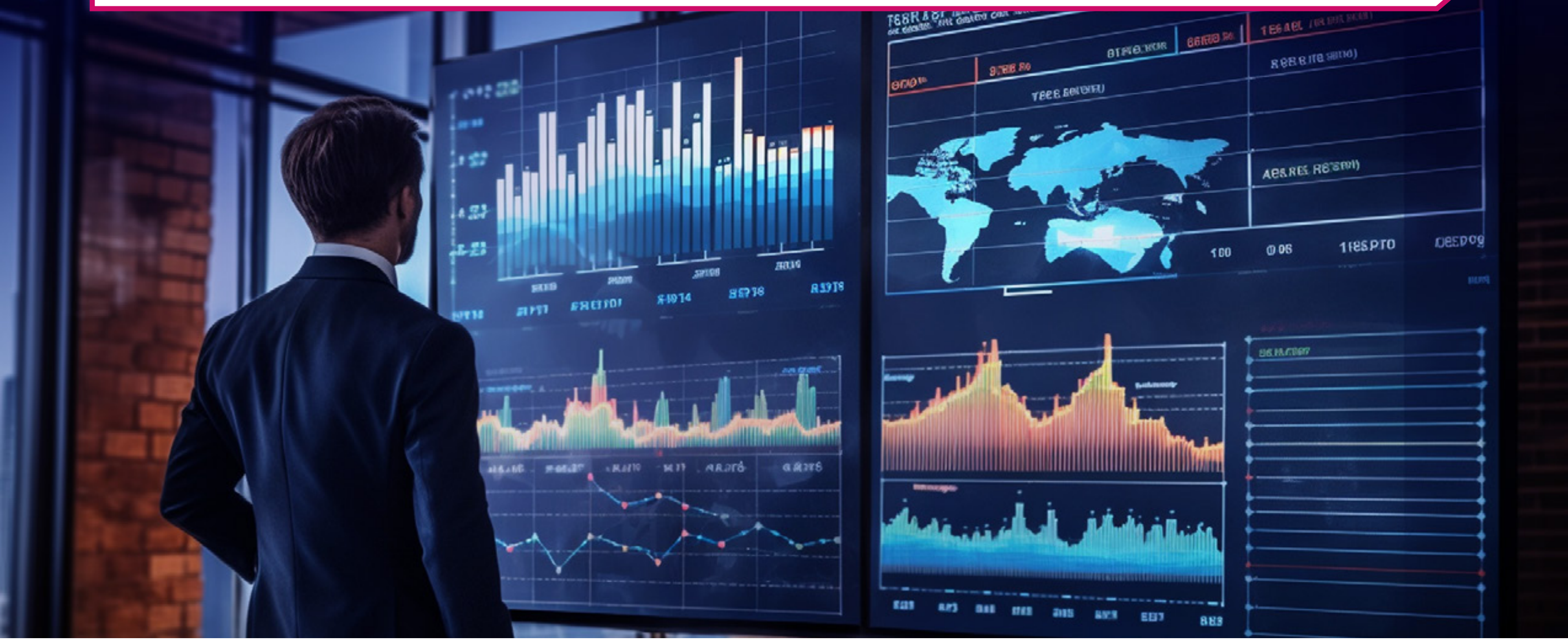
DAS42 does *not* endorse replacing human labor with machine labor indiscriminately. Not all chatbots are created equal, and not all support use cases should be met with the same approach. Some proponents of AI frame this as a zero-sum game: use chatbots in place of human agents, or use human agents in place of chatbots. But in reality the choices can be much more nuanced. We can have human agents and chatbots, each with different use-cases. We can give the human agents access to chatbots to help them improve their efficiency. We can augment human agents' work by using AI to improve grammar or clarity.

When it comes down to it, there's a question at the core of every technology project: *Why are we doing this?* With artificial intelligence and machine learning projects, we find the answer to this question is sometimes seen to be self-evident: "Because it's the technology of the future," or "So that we don't fall behind our competitors," or "It looks good to the shareholders," or "It may be more cost effective to use AI." But we posit that your answer should be, "We want to free up our people's time so that they can focus on high-value tasks that require creative thinking and unquantifiable experience."

A few questions to ask ourselves are:

- Is this a task that **humans are good at?**
- What is the **risk of an incorrect decision?**
- Is this a task that **humans *enjoy?***
- What level of **oversight is needed?**

Automate thoughtfully. Automate in the interests of your people.



True intelligence is transparent

What makes an algorithm "ethical" or "unethical"? Presumably we have all heard stories of machine-learning models that inherit the biases of the humans whose decisions are represented in their training data: a hiring model that uses historical hiring data to predict the most qualified candidates — which rests on the assumption that past decisions were always objective and that future decisions should use the same factors. An algorithm that discriminates based on protected characteristics is an easy example of an unethical one.

To assess bias, AI must be transparent. Predictive models should be understood by their designers, and the process for avoiding unethical bias should be systematic in your organization.

As major players in this space try to create products that are usable by people without a data science background or deep coding experience, we find many of these tools operate with what we call "blackbox intelligence". Engineers prepare data to feed the tool, and the tool provides output, but what is happening behind the curtain or inside the "blackbox" is inaccessible. **If you are not able to see the processes, train the model transparently, and understand contributing factors to the outcomes and *their weights* – you can't learn from the process.**

This is, of course, intentional on the part of these companies. They want to facilitate machine learning outcomes for individuals who do not have the background to *understand* the underlying processes, the ethics concerns, or how to calculate statistical significance.

What we're looking for is a balance – tools and systems that are judicious about how they display what goes on under the hood. **Transparency is a must, but tailored to the right audiences. Not everyone needs to see it all – but the nuts and bolts must be accessible.**

Training ML models requires patience and judgment. With a blackbox process, you may not realize that one of the variables you've given is skewing results, causing recursion loops, or making common-sense predictions – for example, a result that identifies that the customers most likely to spend over $100 are the customers who have already spent over $100.You need meaningful, transparent outputs because you still need data scientists to read those outputs, and perhaps find patterns in data that aren't being stored today.

For example, an algorithm may identify that geography is a contributing factor to late cart abandonment, with the highest rate being for customers in Wyoming – but a human might be able to draw conclusions that a machine can't based on information that the dataset did not have available. Perhaps the human realizes that cart abandonment is actually highest, not just in Wyoming, but in rural areas – but the machine didn't have information about what geographies were rural. The human has the context to seek out a correlation with long shipping time estimates, which may not have been a part of the data set. But if you just rely on the literal outputs, and only the variables available to the machine – or if you don't have appropriate context for those outputs – you may never get down to the actual problem.

A final step in the human-centric ethics of AI is to be conscious of how models are trained, and whether companies providing them were ethical in their sourcing of data – for example, their data sourcing methods respect copyrights and facilitate the ability of people to learn if their personal data is being used in AI systems, and to opt out at any time.

SECTION 2

TARGETED STRATEGY

Identify problems before solutions

AI is often seen as a hammer without a nail. For leaders in many industries, the contrived problems that automated systems, LLMs, and classification models can solve are exciting, but are illusively vague and difficult to map to their own business's goals. This perception is often the result of focusing on the solutions that AI can enable without first identifying a specific, relevant problem that one's business is facing – the nail.

As any carpenter or handy individual will tell you, there are a lot of different types of hammers; choosing the right kind is highly dependent on what you are trying to do. For this reason, it is crucial to define the business problem you aim to solve with AI before attempting to architect the solution.

Are you trying to improve customer experience through a chatbot?

You will need an architecture that supports the deployment of an LLM model with curated content sourcing.

Do you need a custom UI for this chatbot and the ability to host it on your company's website?

You will need to identify a deployment service that integrates with a front-end framework, of which there are many.

Defining criteria like these will help you bring your use case into better focus, and will directly guide decisions about the design and implementation of your AI solution.

AI with a destination

There are generally two types of machine learning models – supervised and unsupervised.

Supervised

As the name implies, supervised models operate within constraints defined by their designer to predict a specific outcome. Supervised models have a specified destination that is always measurable, and as a result can have real business value assigned to it.

Unsupervised

By contrast, unsupervised models generate outputs that aim to identify patterns without the need for guidance from their designer. Exploration is the name of the game with unsupervised models. Due to their comparative ease of deployment, these models are often the first forms of machine learning that organizations experiment with. These early efforts may surface some previously unknown opportunities for customer segmentation, market behavior, or a whole host of other trend-based insights, but they often stall out at this stage due to a lack of clarity on how to measure impact. Attributing a dollar value to unsupervised machine learning initiatives is difficult, and when it comes to deploying AI with a positive, attributable ROI, unsupervised models alone are rarely the right approach.

SECTION 3

SCALABLE ARCHITECTURE

New functionality doesn't always mean new tools

With the increased interest in AI workloads, many new tools and platforms have hit the market. Many of these are great and offer unique advantages, but they are not a requirement to enable enterprise AI. Cloud platforms that you likely already use – Snowflake, AWS, GCP, etc. – have rapidly incorporated broad support for many types of custom AI applications and native ML and LLM models. Try using these tools, especially at the beginning of your journey, as you get your bearings in the world of AI, rather than onboarding to a new platform before you even understand how you're going to use AI.

For many, these may be all you need. If you do select another platform, make sure that it is able to scale with your organization. To our earlier point, don't adopt a blackbox tool with limited flexibility just because you don't have full-time data scientists now.

At its core, a scalable architecture is one that can evolve and adapt to changing demands without compromising performance. By leveraging current technologies, companies can build on their existing infrastructure, integrating AI capabilities that enhance productivity and creativity. This approach not only saves costs but also minimizes disruption to workflows, enabling teams to stay focused on their core responsibilities. Moreover, this mindset encourages innovation within established frameworks. Employees can creatively utilize familiar tools, enhancing their productivity while fostering a culture that values human ingenuity. By marrying existing technologies with new AI functionalities, businesses create a hybrid ecosystem that combines the best of both worlds—effective automation alongside human expertise.

SECTION 4

GOVERNANCE & COMPLIANCE

AI outputs are a product. Own them.

Following the Data Mesh philosophy, all data is a product owned by the team or organization which manages the system that creates it. Therefore, the outputs of AI systems should be monitored, managed, and evangelized by designated owners, the same way that a database management team is responsible for the tables in that database. Similarly, tools and platforms used to deploy AI workloads and make their outputs available to business stakeholders should support clear reporting on ownership, troubleshooting, and recurring QA results so that end users have a clear picture of exactly what they are consuming.

Predicting regulations for predictive AI

As of Q4 2024, the majority of standing international law regarding the development and use of AI focuses specifically on generative AI, and almost entirely regulates the ***type*** of model being used as well as the source content being handed to it. This is a trend that DAS42 has followed and advised on as regulatory drafts have become law around the world. These regulations are a point of focus for the developers of generative AI – Meta, OpenAI, etc. – but their impact on organizations employing predictive AI and most LLMs is minimal. However, lawmakers in the EU and nearly a dozen U.S. states are rapidly shifting their focus to the legality of data inputs and prediction targets for predictive AI. So what does this mean for you as the current or future user of predictive AI?

Organizations that use automated systems to determine outcomes for customers or third parties need to prepare for a world where regulatory audits might occur. The "logic" behind predictive models and their individual predictions needs to be human readable, to be compared against standing regulations on protected classes – hiring decisions, promotions and workplace advancement, credit and loan approval, etc.

Feed your AI models, but beware – some food is illegal

Of similar importance to model explainability is the type of data used to generate predictive outputs. Regulatory protections like CCPA and GDPR have clear definitions of what data represents PII (Personally Identifying Information) and therefore cannot be used if the individual which that data pertains to has opted out. Even beyond current regulatory guidelines, there are ethical and moral implications to the data used to generate predictions. For all of these reasons, it is crucial to keep tabs on what data is and isn't available for predictive workloads.

We often find that businesses underestimate what data can be identifiable. The obvious are things like credit card numbers, full addresses, and social security numbers, but even demographic information can result in a record being identifiable. If you have age, sex, race, ethnicity, and zip code, there are going to be multiple instances where those fields in combination with one another uniquely identify a person.

Data tagging, often applied in an analytics oriented data warehouse, is the best way to definitively categorize and selectively exclude certain data elements for certain uses. **Developing a framework for tagging PII, and other protected types of data is a key step to ensuring that no regulatory or ethical guidelines are broken when designing a predictive model.**

SECTION 5

ASSESSING IMPACT

The most common question asked when attempting to garner organizational support and funding for an initiative is "what is the projected ROI and business value?" There's a reason for this: AI can be costly, and so organizations must focus their efforts to build where they can expect to see explicit impacts on business KPIs, especially early on in their journey with AI.

Let's review a brief example:

EXAMPLE

A company with a subscription-based business model needs to improve retention of their customers. The company decides to use a supervised classification model to predict when a customer is likely to churn, and successfully deploys the model. The model runs on a daily basis, and the customers it predicts are likely to churn are automatically targeted with promotional pricing to encourage retention, or receive direct outreach from an account manager to discuss current pain points. Since every customer prevented from churning has a known average value to the business, the company is able to calculate the ROI for the AI initiative that enabled these predictions. At this point, should the business deem it useful, unsupervised models could be used to surface nuanced insights about customer behavior across unconventional segments – insights that would be difficult to identify through manual analysis alone. These exploratory findings have the potential to inform future strategic decisions and may play a role in improving the predictive accuracy of the initial churn prediction model.

Being able to attribute impact is what takes AI from being just an experimentation tool, to being a driver of positive business outcomes. By defining what problems to tackle with AI before picking a type of AI to employ, designing solutions with measurable business impact, and later using unsupervised models to guide strategic decisions and unearth complex insights, organizations can ensure that their AI initiatives have a positive ROI and are aimed at improving specific aspects of their business.

Give your AI a performance review

The overhead cost of AI initiatives is never zero, so continually assessing bottom-line impact is key. Just like any initiative in your business, the time, resources, and dollars spent to enable an AI deliverable have to be worth the value it brings to your business. For this reason, the age-old pursuit of accurate attribution is the same as it ever was. Did the output of your Customer Churn ML model mean the difference between keeping a customer and losing them? If so, was the per prediction cost of that model less than the additional lifetime value of that customer? These are the types of questions that need to be answered consistently to ensure that AI and ML are net positive to your business.

Automating the analysis to answer these questions is key. Most modern AI/ML platforms provide ample information about the running cost of your AI workloads, which can be used in combination with internal analytics on business outcomes (customer LTV, average per unit savings, etc.) to determine if the juice is worth the squeeze.

Even AI needs to be put on notice from time to time

AI and machine learning are not set-it-and-forget-it workflows. Models have to be regularly reviewed, analyzed, and re-trained. It's a never-ending job.

Sometimes, in the course of this regular reassessment, you will find that these AI workloads are not resulting in positive ROI. Would you part ways with a promising employee because they missed a goal one month? Probably not. Instead, you would give them extra guidance and a detailed plan for how they can reach the expectations of their role. AI is no different.

In the case where your analytics show that an AI workload does not have a positive ROI, there are many levers to pull to correct for this.

Perhaps your ML model is not producing the quality of results that it did last quarter; seasonal aspects of your business and changing consumer behavior need to be factored into a new version of that model.

Perhaps your customer support chatbot is costing more than the additional customer purchases that are being attributed to it; consider dropping to a leaner version of the same LLM model or reducing the number of tokens used to derive responses.

Perhaps the number of attributes that you collect has grown significantly but the quality of your predictions hasn't; more advanced ML model types may be able to leverage these additional attributes in ways that your current model can't.

Iterating on model types, inputs, and the scale of deployments is key in order to adjust for undesirable outcomes. It is also important to account for the non-technological costs associated with AI development, such as the human labor and cognitive load associated with deploying and maintaining these systems. By focusing on iteratively improving and investing in your AI workloads, the same way that you would a human employee, you will see continued growth and efficiency gains in your AI initiatives.

Get in Touch with DAS42 to Further Your AI Journey

Harnessing AI and machine learning in the right way will help your business optimize its resources, operate with more agility, and exceed expectations of your customers. The DAS42 team will work closely with you to harness the full power of AI and ML, crafting strategies that perfectly align with your business goals.

We can help with:

- Workshops and AI/ML readiness reviews
- Gap analysis
- AI roadmapping

CONTACT US