Sonasoft

Why company-wide ML initiatives **fail**

Sonasoft®

Executive Summary

Machine learning is seen as a cornucopia that will rescue the fortunes of any company. Many companies are throwing huge resources into company-wide ML initiatives. But all too often, these initiatives fail. Here, we explore the main reasons for this. Then we present Sonasoft NuGene, an AI bot factory that overcomes all these problems.

Missing expertise

ML initiatives require complex ML models. One of the first issues you will hit is a shortage of experts who can undertake this task. ML modeling needs skilled data scientists and data engineers. But sadly, these are in extremely short supply.

Complexity of deployment

Deploying ML models within your organization needs significant computing resources. Usually, you need to run ML models in the cloud on specialist virtual servers. Even when you can run your model locally, there are deployment issues to overcome.

Lack of interoperability

ML models are developed in isolation. Each model's data undergoes different feature engineering processes. This means even if models share data, they can't run in lockstep. Solving this is time-consuming and requires significant coordination.

Sonasoft NuGene

NuGene is our in-house AI bot factory. NuGene creates ML models from your raw data without the need for feature engineering. These models are ready to deploy in edge devices or in our cloud. Most importantly, NuGene ML models work together, sharing data, and using continuous learning to improve their performance.

Introduction

We are often told that AI is the fourth industrial revolution. If this is the case, then machine learning (ML) is the power fueling the revolution. Over recent years, more and more companies have turned to ML to try and improve their efficiency, effectiveness, and profitability.

Machine learning can definitely transform your business. This might be through better demand forecasting that lets you improve your logistics. Or you might use anomaly detection for accurately spotting fraudulent transactions before they are cleared. You might even be extracting useful knowledge from your historical data.

Usually, single initiatives like these work well, but when you start a company-wide ML initiative, it falls flat. There are three key reasons for this, missing expertise, the complexity of the deployment, and a lack of interoperability. But before we go into these, it's essential to understand something of the process of creating ML models.

In this eBook, we will teach you some of the basics of machine learning. We will show how you can take a trained ML model and use it to create a functional AI bot. Then we will address the issues that commonly cause ML initiatives to fail. Finally, we will introduce you to Sonasoft NuGene, our AI bot factory.

A beginner's guide to machine learning

AI or artificial intelligence refers to any software that demonstrates some degree of intelligence. Remember the definition above? According to that, software becomes intelligent if it is able to acquire knowledge and skills and apply these to a problem. In other words, learning lies at the heart of AI. Or to be more exact, machine learning lies at the heart of AI.

What is machine learning?

In machine learning, the computer is taught to recognize certain patterns in data. It then applies this learning to spot the pattern in new data. There are 3 basic forms of ML.

- **Supervised learning.** Here you use known, labeled data to train the model. E.g. you might show a computer thousands of photos of animals and teach it to identify the ones that are labeled as cats.
- **Unsupervised learning.** This time, you give the computer a set of data and it simply tries to identify any interesting patterns. Typically, you use this to identify clusters of similar data (features) or results that don't fit the rest of the data (outliers).
- **Reinforcement learning.** Here you give the computer unlabeled data and ask it to identify features. Each time it identifies something correctly, it is "rewarded". Using this process, it develops a better and better model.

Creating usable machine learning models?

Many practical applications of machine learning actually combine more than one type of approach. For instance, semi-supervised learning overcomes one of the key issues in supervised learning. Namely, what to do if you only have a small amount of labeled data and cannot create an accurate model. However, you can use the labeled data to improve the performance of an unsupervised learning model. You can also combine supervised learning with reinforcement learning to improve the overall performance of the model.

The human in the loop

Creating machine learning models is very human-intensive. The following is a typical timeline for creating a fully validated supervised learning model.

1. **Understand the problem**. Define the exact problem you need to solve and check if you have suitable data to work with.
2. **Obtain the data.** Get the raw data and convert it into a suitable form for analysis
3. **Move the data to the cloud.** Although it is possible to run AI models locally on GPU-enabled laptops, this is inefficient. So, you really need to migrate the data to the cloud where you have access to (almost) unlimited compute power.
4. **Pre-process the data.** This includes cleaning, filtering, and manually labeling the data for supervised learning. This is one of the most labor-intensive stages of the process.

5 **Choose an ML model.** Your data scientist needs to use her experience to select a suitable ML model from the thousands that are now available.

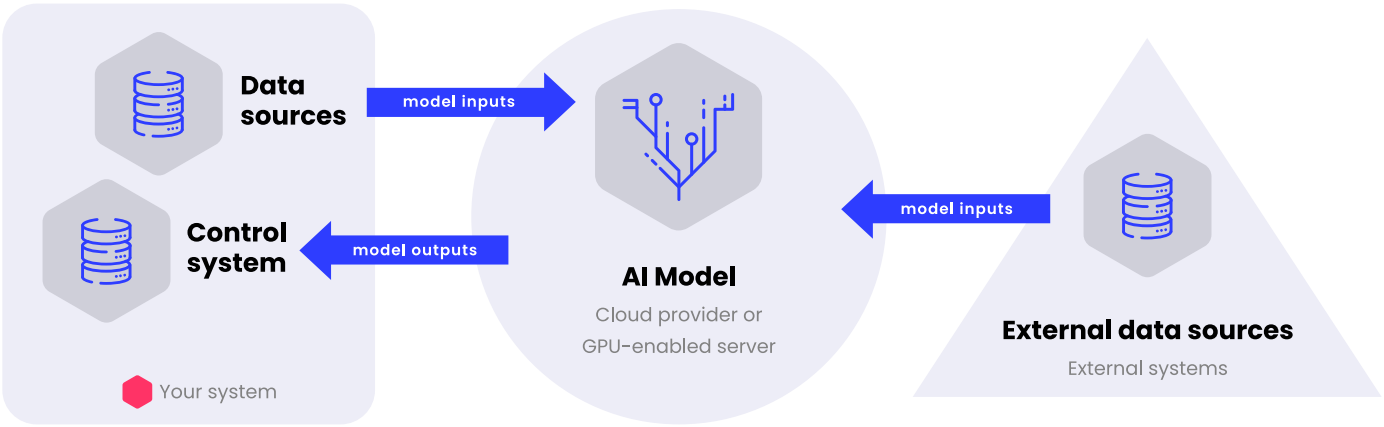
6 **Train and verify the model.** You can now go through the process of training and verifying your model as described above.

7 **Validate the model.** Finally, you have a trained model and can test whether it is suitable for the job needed.

Typically, you won't select the perfect model first time around. So, it is common to need to repeat steps 5-7 multiple times. Overall, this process might take you months to complete. Moreover, you can't easily parallelize it, so simply increasing resources won't solve it faster.

Anatomy of an AI bot

ML models are typically used for four types of tasks: forecasting, anomaly detection, knowledge discovery, and resource optimization. In order to use any ML model, you need to embed it into an AI bot. ML models need significant computing power, so you will need to use specialized hardware or cloud instances. The following diagram gives an indication of a typical AI bot.



A typical AI bot takes data from many sources and feeds the output back to your system

As you can see, AI bots have several elements you need to coordinate. Each type of AI bot has its own issues when it comes to deploying in your business.

Forecasting models

Forecasting models combine current and historical data to make predictions about the future. This can range from predicting electricity demand to forecasting future sales. The problem is, these models often take data from one system but use it to influence a completely disconnected system. As an example, a sales forecast model may be used to predict what stock is needed in which store. The inputs are all the sales figures from each store. But the outputs are then used by your logistics operation to order and ship the stock from central warehouses.

Anomaly detection systems

Anomaly detection allows banks to spot fraudulent transactions. System admins can spot anomalous logins or suspicious network activity. Heavy industry can use anomaly detection to identify failing machinery. All these models use knowledge of historical system behavior to understand when current behavior is anomalous. However, they often require inputs from a huge range of different systems and sensors in order to work properly. As a result, they are hard to integrate.

Knowledge discovery engines

Knowledge discovery is a powerful technique for extracting insights from raw data. It can be used for patent searching or patent discovery. HR can use it to identify potential candidates from large collections of candidate resumes. It can even help customer service agents provide relevant answers and close support tickets faster. The issue here is that knowledge discovery requires as much data as possible. Ideally, it should have access to all relevant data within your organization. But much of this data will be on legacy systems, potentially even on paper or old tape backups. Getting all this data into the knowledge discovery engine requires time and skill.

Resource optimization models

Resource optimization is about allocating your resources in the best fashion. For instance, delivery companies want to optimize their delivery routes to minimize their costs. Factories want to optimize their logistics to reduce the number of parts they have to store on site. In order to work well, your ML models will need to have access to data from a wide range of systems. You will then need to interface the model with your

resource allocation system. Whenever you need to integrate disparate systems, you are going to hit problems.

The problem with ML initiatives

As we have seen, creating a single ML model is completely feasible for most businesses given sufficient resources. The main issues come from getting suitable data to work with and integrating the resulting model into your system. But when you start looking at company-wide ML initiatives, it's a different story. Such initiatives generally look to deploy multiple AI bots across all aspects of a business. In most cases, the board is told the initiative will lead to dramatic improvements in efficiency alongside significant cost savings. However, months later, the project manager often has to come back to you and explain spiraling costs and endless delays. So, before you launch your own initiative, here are a few things to consider.

Missing expertise

ML initiatives require you to create a large number of complex ML models. Imagine a relatively simple initiative to boost customer support. This might well involve knowledge discovery, demand forecasting, and resource optimization models. Indeed, if you are doing it properly, you will need multiple different versions of each model. For every model, you need to employ a skilled data scientist and probably a data engineer to oversee the task. Sadly, such people are in high demand and can ask for high salaries[1]. If you have a project requiring five different ML models, you probably need to employ 5 or 6 data scientists and a couple of data engineers. That's not cheap!

Complexity of deployment

As we saw, deploying AI bots within your organization is relatively complex. You need sufficient computing power (either specialized servers or GPU-enabled cloud instances). Then you need to orchestrate all your data sources. This may require creating connectors to ensure your local and legacy data are available to the ML model. Often, this task requires real skill to get it right, especially if you are dealing with multiple incompatible systems. Realistically, you could be looking at several months of integration effort to get just a few AI bots up and running.

1 In the Bay Area, an entry-level data scientist can earn $120k and a data engineer $150k. Senior data scientists may even earn over $200k. Source Glassdoor and Career Karma.

Lack of interoperability

The final issue is the real killer. ML models are developed in isolation. That means each model's data undergoes different feature engineering processes. As a result, even if two models share a single data source, they have to run independently. Things get harder still if you need one model to feed another. For instance, you might use natural language processing to parse customer support tickets. This is then used to drive a knowledge discovery engine and provide relevant answers from support documentation. Chaining ML models like this is a common use case. But there is no easy solution for doing it manually.

Planning for success

If you are tasked with putting together a company-wide ML initiative, it is vital to plan properly. You need to take into account all the issues mentioned above. Here's four important things you need to consider

Be realistic about the time it will take. We aren't talking about a 3-month plan that will transform your business. ML initiatives can easily take over a year to get right.

Make sure you have a suitable team. You need to employ the best data scientists you can afford. Creating good ML models is as much about experience as anything else. So, senior data scientists are probably worth their salaries.

Turn to the online community to get support on the latest ML techniques and models. ML and AI are rapidly evolving fields. You need to keep up to date to get the best out of your own ML models.

Understand that you need the right infrastructure for running and deploying the models. This requires support from across your team to embed the resulting AI bots into your systems. Without buy in from the whole team, this process will be painful and liable to failure.

If you ask yourself the right questions, you can improve the chances of delivering a successful initiative.

Sonasoft NuGene—the AI bot factory

Fortunately, we have another approach that overcomes these issues and makes it really simple to implement your company-wide ML initiative. Over the past few years, we have developed our own AI bot factory. NuGene takes your raw data and identifies interesting patterns and correlations. From these, it forms some hypotheses about what is the underlying cause of the correlation. Uniquely, it then tests its own hypotheses for causation. Once it is sure the hypothesis is correct, it is able to start developing and testing ML models automatically. Finally, it embeds the best model or set of models into an AI bot that can be simply integrated into your system.

Shared data and continuous learning

NuGene's ML models use an approach called continuous learning to ensure they are always up to date. This means they constantly update their model parameters based on the latest data and whether the model output was valid. Even better, NuGene models can share data sources and can update each other without the need for human intervention. This means the models are always as accurate as possible. It also means they are able to outperform models that are developed in isolation.

Complex time-series data

Another feature of NuGene is that you can use time-series data. This is really important when you are dealing with forecasting. For instance, you might want to forecast whether a customer will default on their loan. To do this accurately, you need to know their current circumstances, repayment history, and the economic climate at the time. Both these last data sets are time-series. NuGene isn't limited to numerical data either. You can feed it video and audio data, and even IoT sensor data.

Continual updates

As mentioned above, machine learning is evolving rapidly. Our engineers and researchers are avid readers of the latest scientific publications. We ensure that NuGene is constantly updated with the latest ML models and takes advantage of the newest techniques. When NuGene generates ML models, it compares the performance of dozens of different approaches in order to select the best one. So, you can be confident that the resulting AI bot is as accurate as possible.

Simple deployment

All NuGene's AI bots can be deployed in our real-time cloud, NuGene RT. Alternatively, you can use NuGene Edge to deploy bots directly into suitable edge devices. This gives you a lot of flexibility in how you leverage the AI bots. In turn, it also makes it much easier to deploy them across your entire company.

Conclusions

Company-wide ML initiatives are not doomed to failure. However, they often do fail because the project manager and company bosses have unrealistic expectations. As you have seen, it is important to understand how much time and effort is needed. You also need to invest in a suitable team of data scientists. Finally, you need the buy in of your entire team, as they will need to help you deploy the solutions successfully. Alternatively, you could see how Sonasoft NuGene automates the process, generating deployable AI bots in weeks, not months. To learn more, contact our sales team.

Sonasoft

+1 (408) 708-4000

info@sonasoft.com

1735 N. First Street, Suite 110
San Jose, California 95112 U.S.A.