AI Guides

Three Gates For AI Project Success

By: Jen Schellinck, John Stroud and Tristan Shaeen
AI Guides
February 20th, 2024

INTRODUCTION

Data powered intelligent automation technologies allow organizations to grow and expand their impact while keeping head count and overhead relatively low.

Until recently, leveraging these technologies was typically reserved for the largest organizations, with considerable up-front resources (time, finances, and data science expertise) at their disposal. They undertook this work either on their own or with the assistance of a university or prestigious third-party advisor.

Today, even small organizations are able to access these intelligent technologies.

Strong leaders are already adept at determining "who" is the best person on their team to face a new challenge or take advantage of a new opportunity. This allows them to channel their energy to where it is most needed, while giving their employees the necessary ownership to lead projects with autonomy and intrinsic motivation.

We believe a similar approach can be taken with intelligent technologies except in this case the "who" is not necessarily a person. Instead, the "who" may be a technology that can act like a fellow co-worker to support and collaborate with an organization's existing employees.

Before Investing in AI Technology Ask, "Would We Hire Someone to Do that Job?"

Before investing in AI, it's crucial to ascertain that there is a reasonable possibility of a return on investment from the AI technology. One way to frame this question is to ask what sort of job the AI can do, and whether you would be willing to hire a person to do that job.

To determine this, the case for using any particular technology in a company must be grounded in an understanding of the business needs being met by the technology, and how these identified needs contribute to the overall functioning and profitability of the company.

Once this connection is clearly understood, a detailed analysis of the technology requirements and overhead can be carried out with this ultimate end-goal kept firmly in mind. In the case of AI technologies, technology functionality, digital infrastructure and data requirements (what we refer to as the three gates) must also be satisfied, but it is only worth satisfying them if this effort makes sense within the context of the expected gain from adopting the technology.

Ensuring a Return on Investment when Adopting Technology

In order for there to be an ROI, a new technology must do at least some of the following:

- Increase the rate of business (e.g. more clients each month), which might involve some combination of:
    - gaining more clients/customers,
    - improving current products and services, thus attracting more new or repeat clients/customers,
    - allowing for the provision of new products and services that attract new clients/customers.
- Allow for an increase in the amount charged for current business products or services.
- Increase efficiency and effectiveness, thus either increasing the capacity to provide current products and services or more generally increasing profit.

THE THREE “GATES” OF AN AI PROJECT

Although the focus of AI solutions is often on the functionality of the AI engine, several other elements (data, digital containers, helper programs) are equally as important for making AI work.

Specifically, to successfully use AI three elements are required:

1. an AI engine (which is an algorithm implemented as a program in a particular a programming language), and which must be housed by a computer (on-premises or on-cloud) that can run the engine.
2. data, which is fuel for running the engine.
3. a series of connected digital containers (e.g databases, file systems) where the data can sit, both before being transformed by the AI engines into new data or information, and after it has been transformed, along with helper programs to move data from container to container, prepare the data and feed the data into the engine itself.

For example an AI engine might transform data about client behaviours, sitting in a database, into predictions about how many clients you will have in the following weeks, which can be saved as a report in a file system.

Increasingly, AI engines are pre-made. Once constructed, these engines may be embedded by vendors into software solutions that provide other functionality as well. This is similar to buying a car with the engine already in the car. AI engines may also come stand-alone - this would be similar to buying an engine that is not yet housed in a car or other vehicle.

For the rest of this discussion, we’ll largely assume the AI engine in question will either (1) work out of the box, using company data as fuel, or (2) be initially available “partially out of the box”, but need to be tuned on company data, after which the engine can be used by feeding it new data. Creating an AI engine from scratch has requirements that are similar to those used for running an existing engine: data, an AI engine (program) that creates AI engines (also programs) and digital containers to hold the newly produced AI engine and the data used to create and test the engine.

Gate One

AI Engine Functionality for Supporting Business Needs

AI engine functionality itself can be broadly broken down into six main categories:

- Communication and Creation
- Forecasting and Prediction
- Classifying
- Monitoring and Detection
- Task Automation
- Pattern Detection and Knowledge Discovery

Drilling down, these high-level functional categories can encompass a wide variety of AI assisted activities. For example, the Communication and Creation functionality might include translating from one language to another (including translating from human languages to machine languages) or creating a summary of an article. Monitoring and Detection could include detection of anomalies (unexpected behaviours or situations) or sending an e-mail when a certain file gets downloaded.

There may be some overlap in categories (for example, sending an e-mail upon file download could also be considered task automation, in part). A good use of these categories is to provide ideas for how to turn existing tasks into AI supported tasks, so if a particular task fits under more than one category, this is not an issue - rather it adds strength to the idea that the particular task could be supported by AI technologies.

Gate Two

Data Requirements for Identified AI Functionality

Digital technologies are created with, and supported by, data. In the case of AI technologies, data requirements can be substantial. Specific quantities of the right types of data are required to act as the right kind of fuel for each AI engine, and these requirements can be difficult to predict in advance of engine construction. As well, separate from its role in AI use cases, data is increasingly being viewed as a company asset in and of itself.

Data as Company Asset: As technology continues to develop, data that might previously have been difficult to use may become more relevant and valuable over time. If the opportunity in the moment to collect and organize this data is lost, this potential future value is also lost. Specifically, as noted, collecting, maintaining and properly organizing data is a valuable company activity, even separately from its role in AI use cases.

As a result, investing in data collection and management, and the associated digital infrastructure and skilled employees, can be worthwhile as part of a general digital maturity strategy. This work will allow a company to take advantage of any new technology developments rapidly, as soon as they arise, and to re-purpose data previously collected for other uses in novel and dynamic ways.

Importantly, unlike many physical assets, the same data, if properly collected and organized, can be readily re-used for new and possibly previously unanticipated purposes while still being used for existing purposes. Because of this, the collected data will likely only increase in value for the company, as it can be used both for its original purpose and simultaneously any new uses as they arise. Having the data in hand and properly organized will support the discovery of these new potentially valuable use cases.

Thinking about data as an asset in its own right is particularly important when dealing with point solutions, which may trap data within the solution. In this case, if the solution becomes outdated or the SaaS discontinues the product, the data assets stored by the solution may be lost in the process. Since historical data can be extremely valuable, this could represent a significant loss for the company. If an organization has substantial capability with respect to managing its own data it will be well positioned to recognize, consolidate and then gain ongoing value from its data assets, particularly when considering the recommended AI Use Cases.

Digital Twins: In order to collect and organize data in a way that makes it most likely to be valuable over the long term - i.e., usable for new AI use cases as they arise, as well as for other analytic use cases - we recommend adopting a strategy based on the digital twin concept.

First appearing in the manufacturing space, the digital twin concept combines a number of pre-existing concepts and technologies to create digital twins of existing physical objects, where the digital twin reflects, sometimes in real time, the current properties and states of the physical object itself.

In its most fully realized implementation, digital twins of real-world objects are supported by sensors which feed data into data repositories maintaining the digital twins. This data can then be used to have situational awareness of the physical objects, in real time, and also, by providing this data to AI engines, used to predict potential future states of these objects - e.g., whether or not they are likely to break down, which can allow the manufacturer to replace a part before this occurs.

Moving out of manufacturing, the digital twin concept is now being applied more broadly, to larger systems (e.g., airports) and to physical objects that may not provide data at such a high level of fidelity, but for which data can still be collected on a regular basis (e.g., clients). With this in mind, identifying a company's key objects (objects critical to the functioning of the company) and collecting and organizing data about these objects through the lens of obtaining comprehensive high fidelity, up-to-date data on these key objects will lead to an increase in possible use cases and data assets with greater short and long-term value.

AI Engine Data Requirements. Each of the different AI engine functionalities discussed above have particular data-as-fuel requirements for operation (note that different and/or additional data may be required to create the engine itself). Although specific engines will require more specific types of data, the following gives a good overview of the relationship between engine functionality, the type of data required to fuel the engine and the resulting output of the engine.

Communication and Creation

Fuel: Text (unstructured, semi-structured, structured), Media (e.g., pictures)
Result: Text (unstructured, semi-structured, structured), Media (e.g., pictures, videos)

Forecasting and Prediction

Fuel: Historical information about object properties - specifically those properties for which we would like to predict future values. For example, historical data on the purchasing history of clients.
Result: New data representing predicted future values.

Classifying:

Fuel: Objects represented by rows in a database (for example clients with information about their purchasing history), text selections (for example newspaper articles) or media files (for example photos).
Result: Information on which category each of the object belongs to, or new categories, depending on the engine. Note that many tasks can be re-defined as a categorization task. For example, a recommender engine takes as input a store catalogue, and divides it into which objects you might like and which ones you probably won't.

Monitoring and Detection:

Fuel: Current information on events, activities or processes as they happen, represented by data coming into a database (e.g., transaction data), media files (e.g., video, photos)
Result: alert e-mails, tagged list of the original inputs also indicating what has been detected by the engine

Task Automation:

Fuel: Digital inputs for a particular task (e.g., files, e-mails, webforms)
Result: Output of particular task (e.g., files move to a new folder, completed form downloaded)

Pattern Detection and Knowledge Discovery:

Fuel: Information about objects of interest (e.g., properties and states), typically represented by rows in a database (for example information about clients collected during previous projects).
Result: Description of detected pattern - this may be in the form of a mathematical equation or text description, provision of new knowledge - this may be in the form of a mathematical equation or text description.

Gate Three

Infrastructure Requirements for AI Use Cases

There are three main options for housing and using AI engines (some solutions may fall somewhere between these options in terms of their functionality): point solutions, commercial AI engines and custom AI engines. Here we provide brief descriptions of these solution types.

Point Solutions: These are stand-alone solutions (typically SaaS cloud solutions, which can be accessed for a monthly or yearly fee) where everything, possibly except the new data used to power the AI engine, is fully contained within the solution itself.

Commercial AI Engine Solutions: In this case, the solution (often a Software as a Service (SaaS) or Platform as a Service (PaaS) cloud solution – e.g., no-code PaaS) is designed to pull in data from an external data container (e.g., company data stored on premises or in the cloud), transform the data internally and then either store the output in another designated data container, or store the data internally until it is pulled out via some external process (e.g., via API, scheduled download or manual download). For many small and medium-sized business, with relatively smaller infrastructure, these can serve as a "Goldilocks" option: not too small and not too big.

Custom AI Engine: Here the AI engine and data containers are managed in-house (although the AI engine may have been initially built elsewhere). In this case, to run the solution in-house the company will require data containers for the input data and results, a server on which to house and run the AI engine and additional helper programs to extract the data from storage, prepare the data to be fed into the engine, and place the results into a new container.

Benefits and Disadvantages of Different Solution Types

In the case of point solutions, the data or information output by the AI engine also stays in the solution, and results will be displayed by interfaces built into the solution. There may be no easy option to extract data from the point solution once data has been added to it or produced by it. This means that those using the point solution must rely solely on the existing functionality of the point solution to meet their needs. And, if the technology becomes obsolete, or the SaaS company folds, the data stored in the solution may be entirely lost. At the same time, a major benefit is that the end-user is also not required to provide any of the digital infrastructure (data containers, helper programs) or to operate and maintain the AI engine. These solutions often have the AI feature as an add-on to some broader functionality that is more the focus of the solution. For example, the solution may be an accounting system and the AI engine may generate some predictive analytics reports using accounting data entered for tracking day-to-day business transactions.

The advantage of using commercial AI engine solutions, rather than custom solutions, is that the creation of AI engines often requires large amounts of data and processor power. Using an already created engine allows companies to avoid this step, while still getting the benefits of AI engine functionality in a relatively versatile fashion, while also retaining more control over their data. In particular, this allows the company to use the data for multiple purposes and respond to advances in AI engine functionality by replacing one AI engine with another as the technology improves. However, connecting the AI engine to other digital infrastructure may not be trivial, and this option may also require the company to provide some of its own digital infrastructure in-house.

With commercial AI engine solutions and custom AI engines, the company can more readily add the AI component to an existing chain of programs. In the case of point solutions, the company has limited control over the both the AI engine and the data stored in the solution. However, in the case of point solutions, if the functionality of the solution and the fashion on which the output is provided is satisfactory, and if the company doesn’t need access to the data outside of the application/solution itself, either currently or in the future, this may be a satisfactory option.

If a particular AI use case is likely to be shared by many companies with similar data, this greatly increases the likely availability of either point solutions or commercial AI engines, especially if the data itself has a similar format or structure across companies. The more unique a use case is for a particular company, the more likely a custom AI engine will be required, in which case the company must have data available to create the engine itself, as well as data to act as fuel to run the engine.