



How AI can exploit data sources of different quality and availability

August 2019

Thought leadership

Monolith AI

There are a lot of different sources available to collect data while developing a product, from quick rule-of-thumb estimates to real-condition experiments. Unfortunately, high-fidelity sources are normally more expensive than low-fidelity ones. As a result, the higher the fidelity of the data, the lower the amount of available data. Moreover, although most engineering companies use different sources, most of the time little to no correlations are made between the results of these different sources. The relationships between different sources can be extremely valuable, and AI can help to combine sources intelligently and fully exploit their relationships.

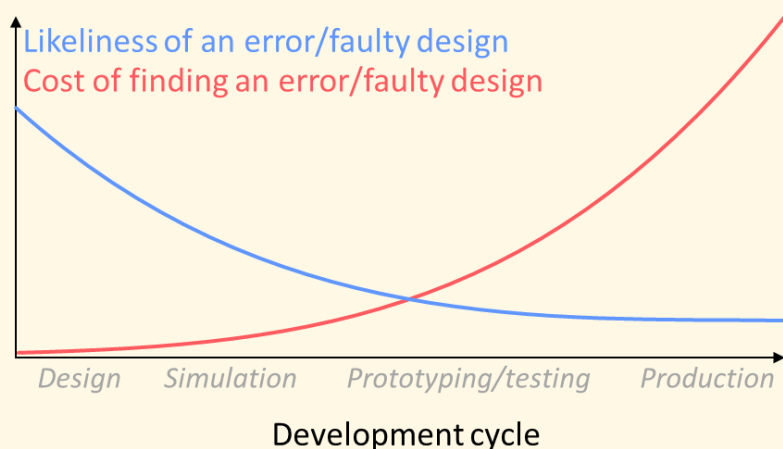


Introduction

In the past, before simulations were a thing, most products were developed by a trial-and-error approach. A lot of prototypes were built until one fulfilled all the required criteria. Some companies still follow this path, especially when manufacturing prototypes is cheap. For other companies, particularly where the cost of prototyping is very high such as aerospace, automotive, ..., the trend is instead now to run many simulations before prototyping. The figure below illustrates two of the main reasons for that trend:

- The cost of an error (time and money) increases a lot as one goes through the development cycle of a product;
- The likelihood of a faulty design is much higher at the early stage of the development.

Therefore, companies try to find as many errors and write off as many designs as they can at an early stage (whilst still exploring a range of possible designs which is at least as large as before). Preferably this is done before physical instances of the product are made, and definitely before the manufacturing process has been designed.



Main reasons why companies haven't invested in AI (Forrester, 2016).



02/ The limitation of single sources

Generally, more accurate and realistic data sources are more costly to gather. As a result:

You have a cheap source of data, but with limited accuracy

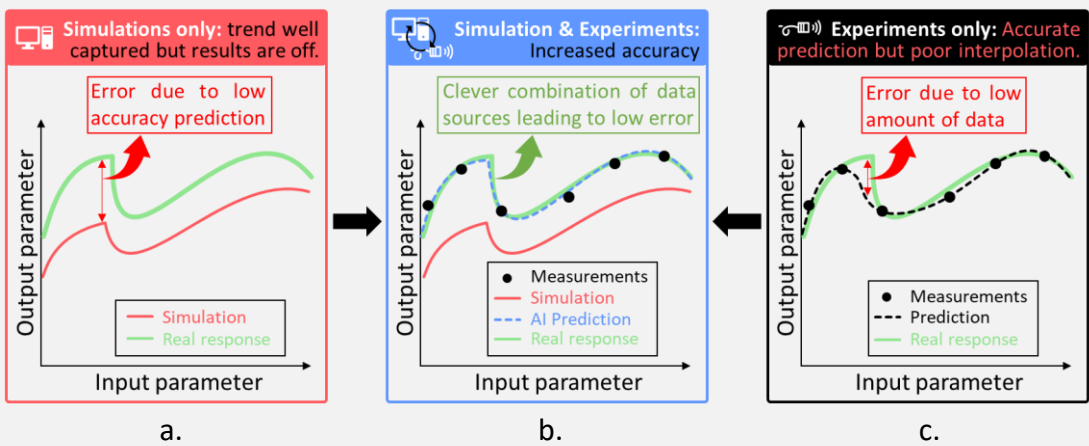
Examples of these type of model could be a set of mathematical equations, an analytical model, a 1D model for CFD prediction, a numerical FE model with coarse mesh. The entire design space might be easily covered by the cheap model, but with limited accuracy resulting in a limited understanding of the product behaviour (figure a. below).

or

You have an accurate source of data, but with limited availability

The source could still be virtual such as refined-mesh CFD and FE simulations, or it could be physical and require prototyping, such as wind-tunnel tests, track testing, tensile test, and fatigue tests. Your results might be very accurate, but you have limited sampling of the design space, also resulting in a limited understanding of the product behaviour (figure c. below).

Most engineering data sources will be somewhere between these two extremes. Most of the time, whatever data source is being used, there will be limitations and errors in the understanding of the product from single sources.



One-dimensional illustration for the inaccurate predictions from simulations (left panel) or experiments (right panel) on their own. An intelligent combination of both sources (middle panel) can result in much more accurate predictions.



03/ Combining different sources to improve predictions

Results from different sources are usually processed separately. Visual comparisons between two sources can help show if there is a clear mismatch. At best, the difference between the sources is qualified but rarely quantified and, when it is quantified, it is generally in a very simplistic way such as finding a single “calibration” coefficient.

Much more insight could be gained by combining these different sources. In the simple example illustrated in the previous page, the simulation (figure a.) finds a general trend that is similar to the real response but offset by some amount. Similarly, the experiments (figure c.) fit the real response more closely but gives fewer data points. An AI model can combine these two sources to predict the output accurately (see figure b.), taking the best properties from each source.

By knowing both the general response trend (from simulations) and a few high-accuracy points through which this trend should pass (from experiments), the model can predict the real response.

ML can be very powerful as it can detect very non-linear trends, whereas trying to build empirical functions to correlate multi-source data is much less reliable and not as universal.

04/ AI-enhance your current development process

AI doesn't need to replace your current tools to be valuable: it can also be used to enhance them. Some complex engineering problems may be too complicated to be modelled perfectly by AI with the current amount of data. For example, if your deep neural network fails to predict the drag within the required accuracy, you might think that AI is completely unsuitable for the problem and revert to running conventional FE/CFD simulations.

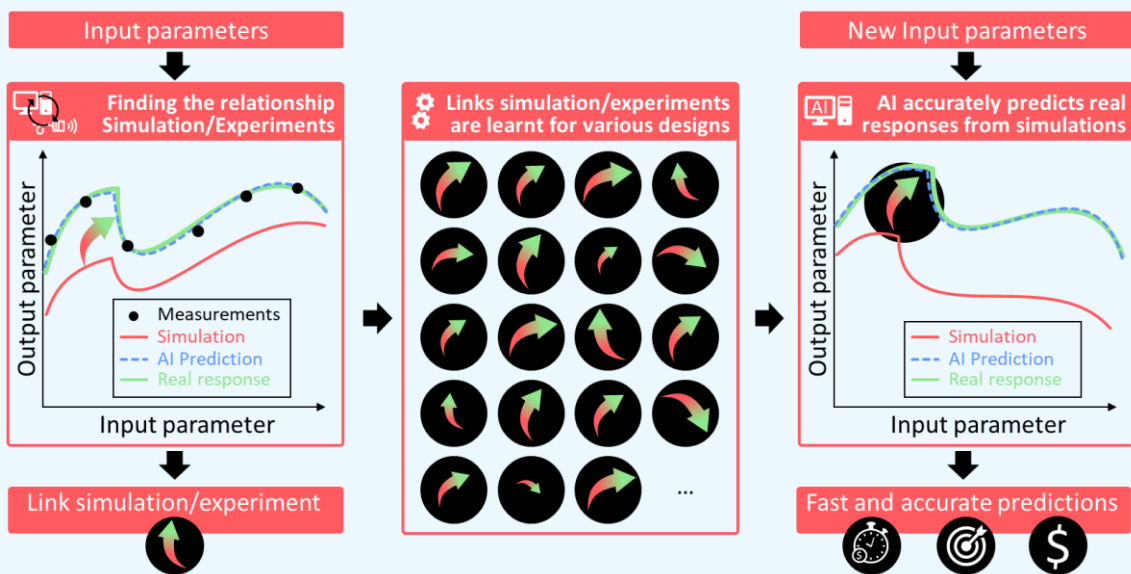
Let's continue with the example of a car. You might not be able to make perfect predictions of the drag of your car with individual datasets, but AI can still be used to link between the different available sources. Combining sources of different fidelities could allow AI to predict accurately the car behaviour, by learning to link the low-fidelity results to the high-fidelity ones. Think of this as a sort of multi-model calibration.





Imagine that this link can be extracted from the existing data and expressed as a simple arrow representing the offset in the figure below (in reality, this offset would be a more complex numerical array or matrix). The relationship could be learnt by an AI model for different car configurations or even between different cars. For a new unseen configuration, a low-fidelity simulation and the trained ML model could be combined to predict a high-fidelity result.

The AI can use the current design's simulation results to predict how this design will perform in the real world, by knowing the previously-learnt link between simulated results and final actual experimental results.

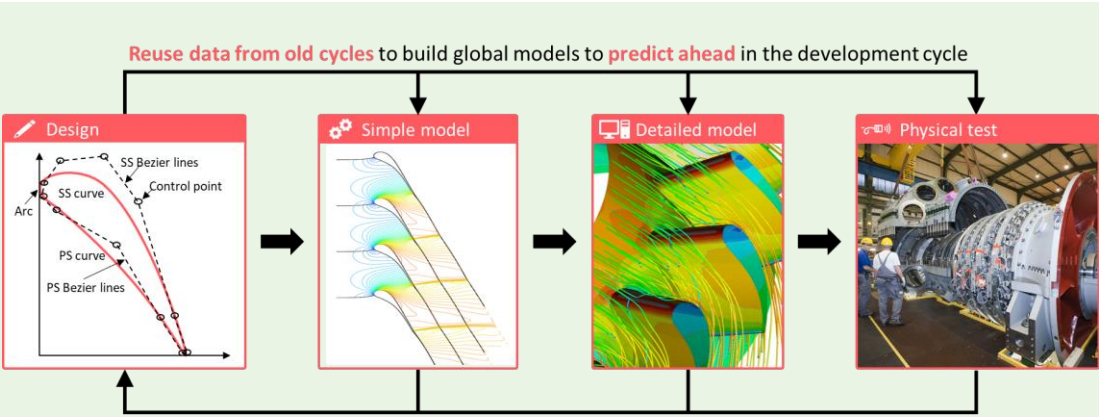


AI can learn the relationship (here illustrated by an arrow) between sources of different fidelities (e.g. simulations and experimental tests). Afterwards, the model can be used to produce high-fidelity predictions from cheaper low-fidelity results, without the need to run experiments.



05/ Conclusion

Clearly this can be extended to more than two sources. The ideal situation would be to combine the data from all processes (whether it is a simple analytical model, a numerical simulation or an expensive experimental test) on a single platform. In a new development cycle, the early-stage initial design results can then quickly be linked to future late-stage performances. The accuracy of such predictions will increase as more sources are collected during the development cycle.



Results from all sources are combined in order to predict performances ahead of time.



For more information
visit monolithai.com



MONOLITH