AI REQUIRES EXPERIENCED SUPPORT

New Electronics talks to distributors about the support they are providing engineers, in the industrial space, when it comes to using artificial intelligence. By **Neil Tyler**

Over the past ten years artificial intelligence (AI) has moved from pure research into a technology that is now revolutionising the world of information technology.

Deep learning has the ability to process images, video and speech, and deliver results similar to those of humans and is now a core part of numerous online services.

The potential of AI is immense and is seen by designers of industrial systems as an important tool to improve the efficiency and performance of their systems because of its ability to detect trends and anomalies in complex data in real time.

"The concept of using AI in a production environment can be very daunting," explains Richard Jeffers, Technical Director for RS Components in Northern Europe.

"Engineers will need to learn a whole new language, which will be very different to what design and maintenance engineers may be familiar with. They will also need to be able to see through the hype and understand the problems that the technology can help solve. These are just some of the common issues."

According to Jeffers customers need support from the industry, and distributors are in the perfect position to offer this.

"Rather than just focusing on the technology, it's important for distributors to understand the problems a customer is looking to overcome, and then look at the technology that can help them," he suggests.

However, an understanding of the

tools and development platforms available to engineers is vital.

"The availability of cloud-oriented development tools such as Caffe, MXNet and TensorFlow provide a major boost for developer productivity," suggests Michaël Uyttersprot, Market Segment Manager Artificial Intelligence and Vision, Avnet Silica.

"They make it easier to evaluate different AI techniques, particularly for those based on deep learning. The shape of the pipeline in a deep-learning network, for example, has a strong influence on its performance for a given application. When it comes to an image-recognition pipeline, for example, it will often be quite different to the networks employed for time-series data such as audio streams. It is important to be able to try different structures on sample data in a convenient manner, and these environments make it possible."

Even when the prototype has been shown to be working in a workstation

Above: The potential of AI is immense and is seen by designers of industrial systems as an important tool to improve both efficiency and performance

or cloud implementation, the developer, in this case an embedded developer who wishes to make use of AI techniques in their application, still faces challenges. One of the main issues is that of local processing resource.

"Deep learning has a requirement for a large number of matrix multiplication operations for any of its operations. Training can be readily offloaded to powerful cloud servers. But inferencing requires low-latency responses that will, in most cases, mean execution either on the target device itself or on a gateway module," according to Uyttersprot.

"There are a growing number of manufacturers, such as NXP, building support for deep learning into their embedded processors that are capable of high-throughput matrix multiplication. However, these will have significantly less compute power than the cloud servers in use today."

Embedded systems

To implement machine learning on embedded systems requires a number of additional steps beyond those required of developers working on cloud systems, which is creating a skills gap in the industry.

"But this is a gap that can be filled with help from partners, such as design-in distributors, who can provide in-depth assistance," explains Uyttersprot.

"One approach used by Avnet Silica is to take the available components and wrap machine-learning software IP around them to create building blocks that can easily be integrated into a target system. In assembling these building blocks, distributors can pass on the benefit of their experience."

One of the key lessons from deploying machine learning in embedded environments is that, when used for inferencing, deep-

learning models do not rely heavily on arithmetic precision and they also often exhibit high levels of redundancy. When models are trained in the cloud, the default is to use floating-point arithmetic.

"This is computationally intensive but allows for a smooth training process. Numerous research projects have demonstrated that the error rate of deep-learning networks only slightly increases even with dramatic reductions in arithmetic precision. For many networks it is possible to use low-resolution 8bit integers for neuron-weight calculations in place of the double-precision floating point frequently supported by cloud-oriented libraries and environments. Some experiments have shown that even binary or ternary resolution is sufficient for some applications."

Pruning is another source of greater efficiency. This technique analyses the trained neural network to determine how much influence each path has on the final result.

"With the right platform, developers can apply pruning and approximation to models developed using cloud-based libraries," says Uyttersprot. "Xilinx, for example, supports its FPGA-based platforms with a set of AI optimisation tools that can take a trained model and produce a version that will run efficiently on its hardware. But such tools are only part of the solution for embedded implementation."

The cloud-based environments do not take into account the depth of integration required to make AI work in the context of a real-time, embedded system. Conditioning those inputs to be suitable for an AI model that performs time-series analysis requires careful consideration.

"Such complexities demonstrate why the classic hardware-only distributor model, of yesterday, does not work effectively when customers need to deliver complex systems that rely heavily on software integration," according to Uyttersprot.

"As a result, Avnet Silica has been working on solutions for AI for several years and through platforms, such as Xilinx's Zynq UltraScale+ MPSoC, we have built development systems that make it easier to create AI-accelerated applications."

What data?
For artificial intelligence to work, it needs data.

"At RS, we've invested in the capability to support customers through the development and delivery of an Industrial IoT strategy, supported by our recent acquisition of Monition, a specialist reliability and condition monitoring business," says Jeffers.

"We have had experienced maintenance and operations professionals engage customers, suppliers and industry thought leaders in how they see industrial IoT unlocking value in the factory environment. Through these conversations, we know customers are looking for support on what data to collect, how to collect the data and then how to apply AI to the data."

To support the 'what data?' question, RS works with customers to conduct a 'Criticality Assessment and Technology Selection (CATS) Survey.

"Through this, we jointly identify the critical assets and assemblies in the customer site and the right technology and parameters to measure to get an early indicator of failure," explains Jeffers.

"The availability of cloud-oriented development tools provide a major boost for developer productivity."
Michael Uyttersprot

Below: Artificial intelligence is helping to revolutionise the world of information technology

"The easiest way to collect data is to harvest from existing PLCs, industrial PCs and historical databases. Only where the data does not exist, locked in the control environment, would we advocate installing new sensors."

According to Jeffers, having collected and aggregated the data, and executed any appropriate local processing, it can then be passed to the cloud for in-depth analysis.

"We are working with data from our distribution warehouses and with data from customers who are interested in co-developing solutions. Prior to deploying any AI tools, we pass the data through rule-based streaming analytics to identify any immediate issues. After this, we take a multi-faceted approach: physics based simulation of the real-world environment to build a digital twin of the system, e.g. we know that, as bearings wear, power consumption in a motor increases, and that can be represented in a model; machine learning to look for correlations in the data sets, and to build algorithms based on these correlations to predict future events; real world domain expertise to validate the outputs of the digital twin and machine learning and to accelerate the training of the system."

Having completed a round of training on a relatively simple data set, RS repeats the exercise on a more complex data set to understand how much of each model can be ported between use cases, and how much needs to be built new each time.

"RS is working to build a solution that can be applied to a range of common industrial plant and processes, to make it cost effective to deploy to a range of customer problems.

"It is this kind of support that distributors must look to offer to customers if AI is to thrive and progress in the industrial space," Jeffers concludes.