Artificial intelligence: a killer app for edge computing?

Artificial intelligence and edge computing are both buzzwords within the industry in their own right. But, they should also be considered together – AI as a key use case for edge computing, and edge computing as a key enabler for AI to deliver on performance and keep costs down.

Tilly Gilbert, Consultant

December 2019

Artificial intelligence (AI) is a data-heavy, compute-intensive technology – a perfect candidate for edge computing. Edge benefits AI by helping overcome the technological challenges associated with AI-enabled applications and it is specifically well positioned to deliver on:

1. **Reduced data transfer to the central cloud** – machine learning algorithms must ingest very large amounts of data in order to detect trends and provide accurate recommendations. Rather than streaming all of this to the cloud, more processing can happen at the edge, thus reducing backhaul costs. This is particularly important for use cases which require analysis of high-definition video, where streaming to the cloud would need huge amounts of bandwidth.
2. **Real-time decision making (with reduced latency)** – where machine learning is triggering real-time actions, latency must be kept to a minimum. Rather than streaming all raw data to a remote cloud for centralised processing, edge computing can enable these decisions to be made close to the source of the data and resulting actions triggered at the edge.
3. **Local data storage and processing** – by using edge computing, sensitive or proprietary information, such as customer location data, is stored locally rather than in the cloud. By performing AI at the edge, only aggregated data sets and key insights need to be streamed to the cloud and the rest of the data is remains local.

Why would you use edge computing for AI and advanced data analytics?

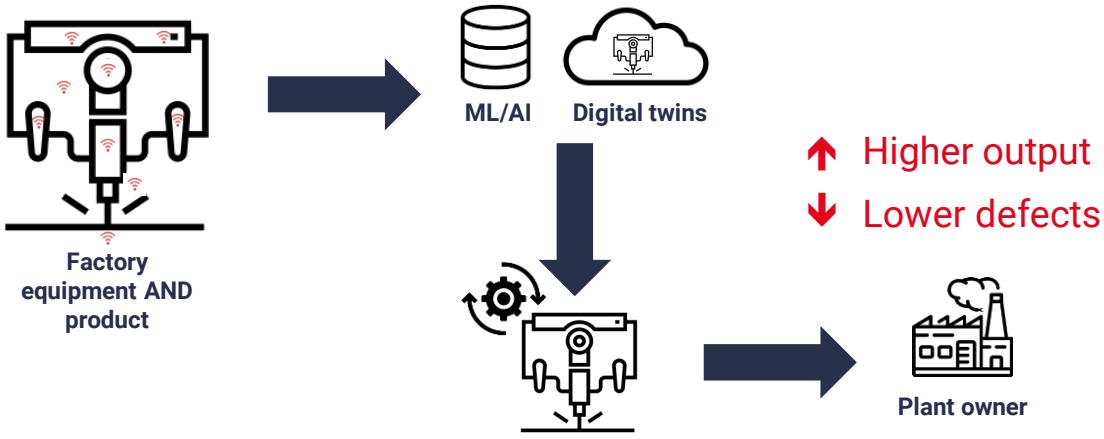
Currently, where artificial intelligence is used, it is either happening on a device (e.g. on smartphones for virtual assistants like Siri or Cortana) or in the cloud. On-device AI has real limitations for IoT use cases where devices are unlikely to have the compute, storage or battery power to ingest, label and analyse very large amounts of data. On top of this, machine learning algorithms must be fed massive amounts of data aggregated from many devices (rather than just one) for outcomes to be accurate and useful.

Equally, as the number of IoT devices increase, it will not be feasible in all cases to rely on the cloud to process and analyse data for real-time decision making. Streaming and storing huge amounts of data may become prohibitively expensive in backhaul and cloud storage costs. Instead, the big data analytics can occur closer to the end-device at an edge computing location. Only updates to algorithms need to be sent back to the cloud in order to synchronise learnings across multiple sites.

What are examples of applications that would use AI and edge computing?

Use case 1: Precision monitoring and control of manufacturing machinery

Precision monitoring and control of machinery is one example of a use case that would be well suited to using AI at the edge. This needs to leverage these technologies because it not only requires very large amounts of sensor data to be collected and analysed, but, off the back of this, changes to the machinery (fine-tuning of movements, temperature reduction, vibration control etc.) and overall manufacturing process need to be made in real-time. In a high-speed production line, latency must be kept to a minimum, and therefore doing the data processing closer to the manufacturing plant is highly valuable.



Source: STL Partners

Use case 2: Video analytics - video surveillance, facial recognition and flow analysis

There are several specific use cases in various verticals that will make use of video analytics. For example, in retail, video analytics can be used to track customer footfall and analyse the buying patterns of particular customer profile types to improve products, product placement and customer service. In comparison, in smart cities, video analytics may be used for surveillance and tracking of criminals using facial recognition software. With recent high court rulings deeming the use of automated facial recognition as lawful in the UK, it seems likely that adoptions of this video analytics application will rise. Advanced pattern and facial recognition requires significant compute power and doing this at the edge, rather than in a centralised cloud, will reduce latency and backhaul costs.

Increasing levels of AI and video analytics and multi-camera coordination required

	Object detection	Object recognition	Object tracking
Security & surveillance	Intrusion detection; Incident detection	Facial recognition; Access control	Criminal tracking
Counting & tracking	Traffic optimisation; Footfall analysis; Queue mgmt. / resourcing		
Fault detection		Equipment monitoring; Quality control	
Customer engagement		ID verification; Customer recognition	Heatmapping / tracking

Source: STL Partners

Challenges to using edge computing for AI

The key challenge to using edge compute sites (either on-premises or network edge) for artificial intelligence is the need for heavy duty storage and compute power. Most AI use cases use GPUs in the centralised cloud to provide this, but it is not clear to what extent GPUs will be part of operators' edge compute roll-out strategies, nor a key component of an on-premises edge at enterprise sites. However, Google has developed its own edge TPUs[1], purpose-built ASICs[2] designed to run AI at the edge. This may prove a viable alternative to the heavy duty compute capabilities of the centralised cloud.

Tilly Gilbert is a Consultant at STL Partners, specialising in edge computing, artificial intelligence and automation.

Get in touch with the author to learn more

tilly.gilbert@stlpartners.com

Or visit STL Partners' Edge Hub

www.stlpartners.com/edge-computing

[1] TPU: Tensor Processing Unit – Google's custom ASICs for machine learning algorithms
[2] ASIC: Application-specific integrated circuit – a customised chip design for a particular use, rather than general purpose