DELL Technologies

SOLUTION BRIEF

A federated learning platform for real-time artificial intelligence

How a secure framework helps drive more powerful analytical insights and provides a near real-time approach to edge computing

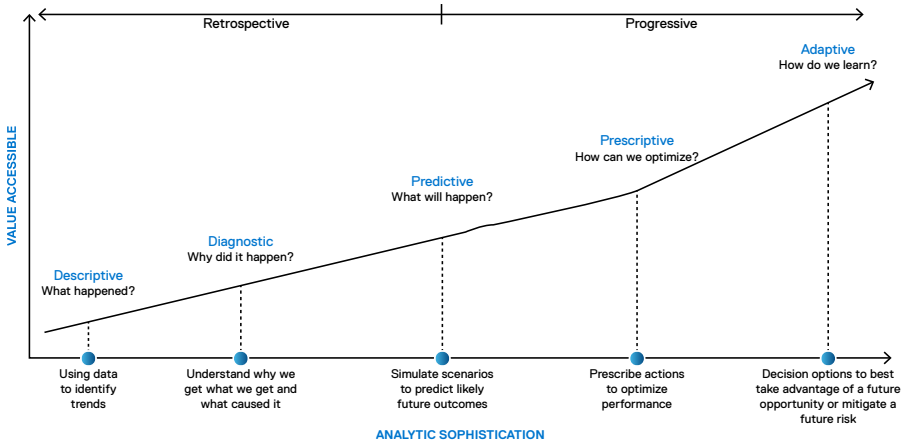
The edge is undergoing rapid transformation as demand grows for more widespread automation and system autonomy.

In response, machine learning (ML) and artificial intelligence (AI) are entering a third phase of development. In the first phase, experts defined rules for AI systems that generated descriptive and diagnostic results. In the second, big data and the cloud provided the means to enable a more predictive approach that could help augment a wider array of human decisions using powerful centralized systems. Now, in this third phase, AI/ML is becoming both prescriptive and adaptive with a wide range of intelligent edge systems delivering data-driven benefits extending to nearly all facets of human activity.

However, traditional AI/ML centralized and distributed architectures were not designed to support the data demands of modern autonomous systems, particularly as the largest volume of edge data is video based. These architectures have inherent limitations in being able to effectively gather, process, correlate, analyze, and utilize the information from edge devices in near real-time in order to produce the desired outcomes, while protecting the security and integrity of the data.

In a centralized framework, all data must be sent across the network to a data center or the cloud before deep learning models can generate the needed insights. This poses significant drawbacks especially when the value of data at the edge is time sensitive. A self-driving car, for example, can't wait for the centralized AI platform to decide if what its sensors have detected is an obstacle in the road and what action needs to be taken. That decision needs to be made immediately, in real-time.

DATA INSIGHTS – ANALYTIC MATURITY



THE THREE TYPES OF AI/ML ARCHITECTURE

CENTRALIZED

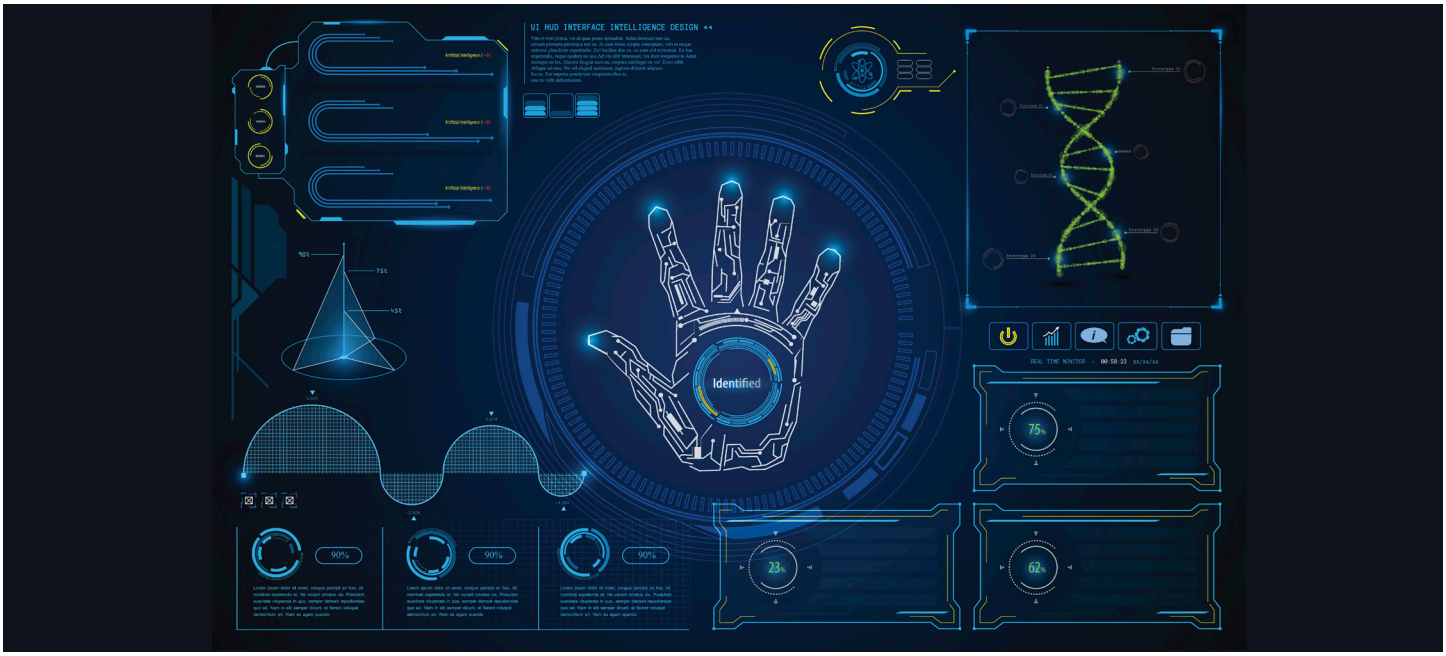
Edge devices send data to a central server (usually cloud-based), which builds a model via machine learning. The user or edge device then sends a request to make use of the available services, which are delivered from the cloud.

DISTRIBUTED

A global model will be sent to the edge device initially, but no further real-time communication to a central server takes place after this exchange. Edge devices can continue to build their own models independently using locally collected datasets, but local models in isolation lack input data diversity which can create biases.

FEDERATED

Each edge device performs its own duties in training a model with its own data, but then sends its results back to the initiator for aggregation. The central model is improved and can be sent back to the edge for better local modeling. Data stays local with only the trained model shared centrally and with peers. No raw data is shared across the network



Likewise, predicting the spread of an infectious disease is time critical and dependent on the sharing of general clinical learning while keeping patient data private. A centralized platform is limiting as it requires large network bandwidth utilization and exposes raw patient data to interception and exploitation. A distributed architecture, although faster and more secure, is not designed to be able to share consolidated results across regions, jurisdictions or even organizations.

This is where **a federated approach** to AI/ML processing benefits. It allows computational processes, AI, and ML algorithms to be run on data sets at the network edge as they are gathered while sharing only mathematical models, metadata, and results of these queries over the network to other edge devices, data center(s), or cloud. This interchange helps improve results by enabling the near real-time extraction of actionable insights from large, distributed datasets without exposing the data and intellectual property.

IT complexities when deploying AI at the edge

According to Moor Insights & Strategy, AI has become the number one workload at the edge.[1] Directly related to this trend is the rapid increase in the variety of endpoints delivering data. In fact, Gartner predicts the number of Internet of Things devices will grow around 15 percent every year until at least 2030, using wired, WiFi and wireless data connections.[2] These data sources will be in mobile ground systems, fixed towers and facilities, drones, handheld devices of users, ground sensors, biometrics, and even from livestock.

A robust and scalable enterprise IT infrastructure is required to bring the promise of AI, IoT, and autonomous operations to fruition. Moving greater computing power to edge devices is a start as it situates processing closer to where data is being collected. However, this introduces new levels of IT complexity for organizations looking to deploy AI at the edge. Challenges include:

- **Network Bandwidth**: Shifting compute and data to the edge means that endpoints utilize more bandwidth. Enterprises have in the past allocated higher bandwidth to central data centers and lower bandwidth to endpoints. The move to edge computing puts greater emphasis on having more bandwidth across the entire network, not just focused on the core.
- **Distributed computing**: Alongside the need for more evenly distributed network bandwidth is the greater distribution of compute power, which means that this edge infrastructure must be provisioned with enough resources. A remote micro data center may need similar resources to a centralized one, for example.
- **Latency**: Edge computing places the compute power closer to where the data is collected, which reduces both application latency and decision-making latency. However, the process of building learning models could still mean massive amounts of data will need to be moved back to the central core computing resource. Even where the compute is more evenly balanced between core and edge, data will travel in both directions, raising security concerns on data sharing and dealing with access rights.
- **Security**: Centralized compute and applications can benefit from centralized security, with the main resources residing inside a highly protected space. When computing and applications are moved to the edge, the security requirements are distributed to the edge as well. The edge needs the same level of security, reflecting its location and traffic patterns. Users may need access rights to a much larger number of devices. A report by Frost and Sullivan[3] argues that this is further complicated by the lack of native data decryption on many edge devices to keep costs and deployment speed down. An edge computing deployment needs to take this into account, so that adequate protection against cyber threats is implemented.
- **Backup**: Backing up data held centrally is a much simpler process than backing up data when it is held across multiple edge devices. Enterprises will need to develop an overall data protection strategy that takes the distributed locations into account. This has network bandwidth implications that are just as critical as the choice of storage medium. Backing up over the network may not be practical or preferable.

- **Data accumulation**: Data handling and compliance rules can also be problematic when the information is held in many distributed locations rather than centrally. Data governance is an increasingly critical area in modern computing, which can create considerable corporate liabilities. An edge computing scenario will also need to consider how data is moved over the network during its lifecycle, ensuring this conforms with regulations.
- **Control and management**: Enterprises need to maintain an identical set of procedures and protocols for management and control across all physical edge locations. This will require a suite of orchestration tools able to manage the potentially disparate nature of edge devices.
- **Scale:** Placing more compute at the edge naturally increases the scale of everything IT teams need to manage. Not only are there more devices to administer, but more compute overall, a greater distribution of network infrastructure, a larger storage capacity, more distributed management, a more expansive security domain and a wider array of application licenses. Enterprises must consider the increased scale in every aspect of IT provision.

Overcoming the challenges of distributed and centralized models

The Dell Technologies federated AI platform combines the best features of centralized and distributed AI architectures to help address the challenges posed by the increase in computing at the edge as data proliferates. The endpoints can react to incoming data immediately in real-time as with a distributed system, but they can also collect data, build models locally, and send the results of those models back to the initiator to improve the core model. The new optimized data models that incorporate results from all devices can then be shared across the network.

MORE ACCURATE INSIGHTS WITH LESS 'DATA BAGGAGE'

A distributed system, because it is not feeding data or results back from client devices, cannot provide a central overview including statistics, analytics, and insights. As a result, training models are subject to biases and system accuracy markedly diminishes. A centralized system has the advantage of a central overview, but a federated system is also able to deliver a similar benefit without requiring the transmission of excess data across the network. When considering that only an estimated 1 percent of the monitoring data is useful to derive business insights, a federated approach helps prevent transmission of terabytes of irrelevant data to the cloud or data centers and send only the relevant, actionable data.[3]

The accuracy of a federated system is also increased by having access to all relevant models, results, and metadata available throughout the broader operational ecosystem—whether from other devices or the core or cloud. This more accurate model enables real-time insights for better and more timely decisions.

THE SAME PREDICTIVE PERFORMANCE AT A SIGNIFICANTLY REDUCED BANDWIDTH

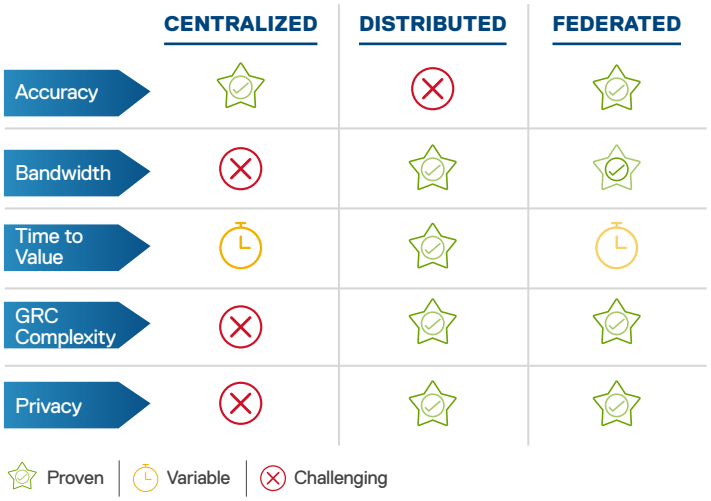
A potential concern with a federated AI architecture is that it won't deliver the same precision as a centralized model. Despite advancements in compute power at the edge, particularly with the application of extremely powerful GPU acceleration, the central server is still likely to outperform what is possible at the edge in raw compute terms. But as with accuracy, the use of all the raw data is not required when it comes to precision.

In fact, tests performed by Dell Technologies on two different federated environments have shown that the average precision of a federated system was comparable to a centralized one. In contrast, the bandwidth required to produce these results on the federated system was 99.96 percent less.[4] This results in substantial cost savings compared to a high-bandwidth centralized system, especially when considering not just data and wireless charges, but savings to the network infrastructure across the entire solution.

GREATER SECURITY AND REGULATORY COMPLIANCE

When edge devices in a federated system share their models with other devices or with a centralized server, the raw data is kept local, with only a pointer included. The movement of data, including personal identification and other sensitive information can remain within geographic, social, or political boundaries to better accommodate governance, risk and compliance (GRC) factors. Model insights are also homomorphically encrypted to maintain the ability to manipulate data without it being decrypted and visible. Only the results are shared, not the data itself.

COMPARISON OF BENEFITS FOR THE THREE AI/ML ARCHITECTURES

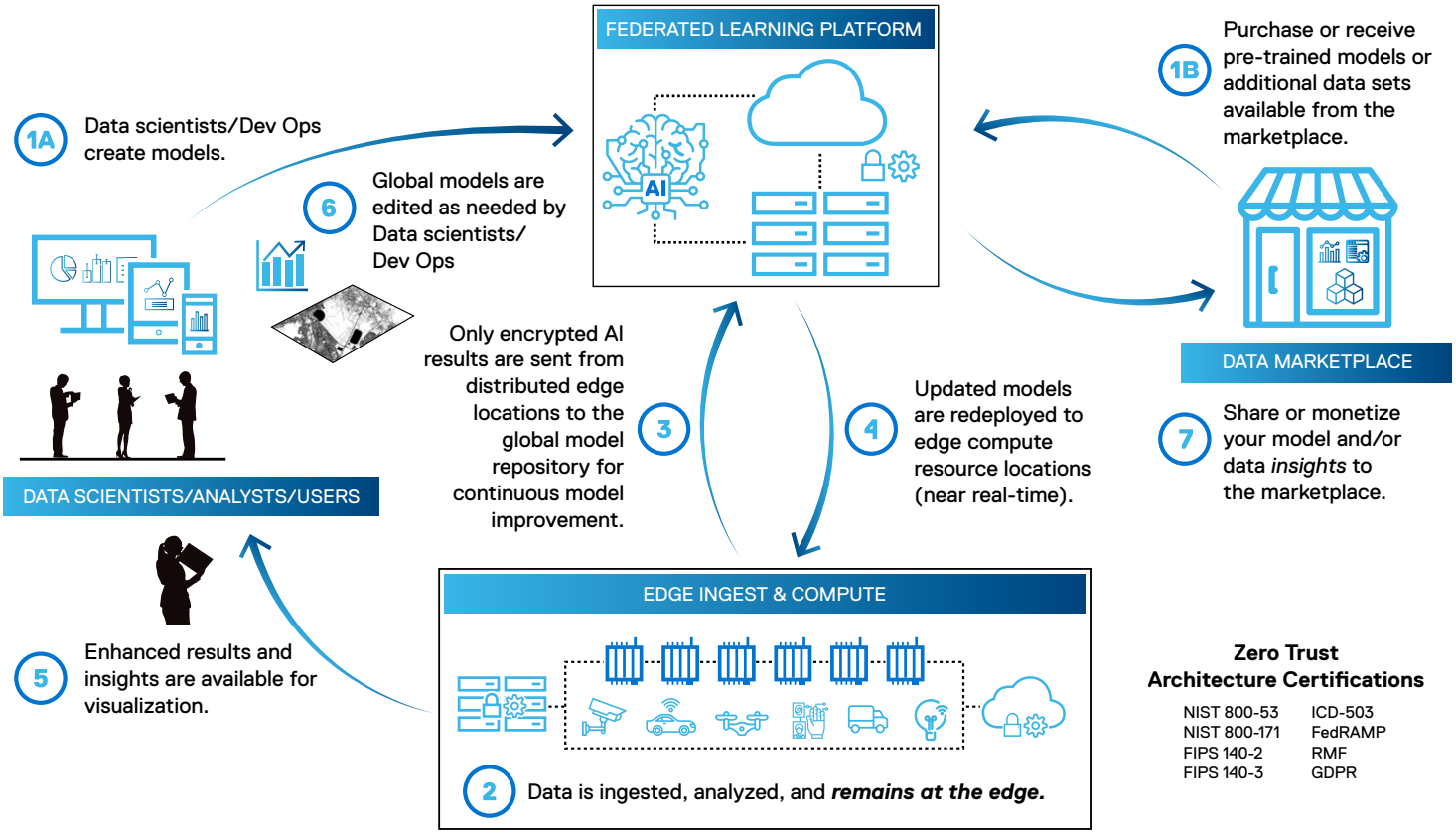


A simplified, scalable solution for your federated AI environment

The Dell Technologies Federated Platform is a prebuilt AI platform bringing together an ecosystem of lab-validated workloads, no-code/low-code analytics capabilities, edge devices, hyperconverged and storage platforms, and industry-specific software in a scalable architecture. We work closely with a curated partner ecosystem to provide the right solution designed to meet the needs of your AI/ML requirements. You also gain access to a marketplace where models and algorithms can be exchanged to help improve outcomes and generate greater business value from your AI/ML environment.

EDGE COMPUTING USING THE FEDERATED FRAMEWORK

The flow of information in the Dell Technologies federated model keeps data in place from each edge device while enabling secure data sharing of metadata, results and models across users, devices, data centers, and the cloud.



THE FEDERATED MODEL

The Dell Technologies federated learning model moves the compute to the data. Training takes place at the edge for network cost and privacy advantages, with no need to move data to a centralized lake. The interface also remains at the edge, which allows for more flexibility in connected and disconnected states.

SIMPLICITY FROM DESIGN TO IMPLEMENTATION

The platform delivers low-code or no-code deployment, creating CIO and CEO-level dashboards and reporting with ease. The client interface enables data scientists and analysts to manipulate the data and generate the desired insights. There is support for most common open-source AI frameworks. The automation of basic AI tasks includes support for the entire lifecycle from data ingestion, training, and model deployment to results visualization.

A CURATED PARTNER ECOSYSTEM

Dell Technologies provides a rich partner ecosystem for its edge strategy. This allows the delivery of best-in-breed solutions across many domains from a curated pool of over 100 technology, AI, and service partners. But our experience delivering integrated solutions means there is no lock-in to curated partners, with organizations able to build a solution that precisely meets their needs using their choice of platform components.

A CENTRAL MARKETPLACE FOR AI MODELS

A significant benefit of keeping the data local and sharing only the model in the federated system is the ability to exchange models and algorithms between edge devices, endpoints, and even organizations. Federated learning can be shared horizontally between platforms, or vertically via common entity IDs.

The AI Marketplace in partnership with other third party vendors provides trading or sharing of models using blockchain-enabled smart contracts. This allows organizations to trade data and algorithms, with support for cataloging, usage payment and auditing. The ecosystem already consists of 10,000 providers and enterprises.

A metro edge is available in 50+ metro areas across more than 25 countries with a secure network fabric across these distributed locations. Multiple clouds are supported with sub-1ms access across multiple clouds.

A COMPREHENSIVE SECURITY STRATEGY

Data manipulation is performed on a secure Kubernetes platform, using a zero-trust architecture. The lineage of data and model training is tracked via the customer's choice of blockchain, to provide an auditable usage trail. Automating the logging and registration of analytics increases the trust, transparency, and traceability of the entire analytics process. The platform is compliant with the principal data protection and privacy standards including NIST 800-53, FIPS 140-2 and ICD-503 security and privacy standards.

THE DELL TECHNOLOGIES FEDERATED PLATFORM

The analytics platform extends from edge to core to cloud including governance, marketplace software, AI technologies, edge solutions, and the supporting infrastructure.

GOVERNANCE

Support for consortium and single-party governance

- Manages which parties can participate
- Identifies minimum requirements for participation
- Includes exchange agreements and management
- Supports controls for consortium-based models
- Enables single-party governance.

MARKETPLACE

Environment to manage and enable the exchange of algorithms and data

- Supports different trust models
- Blockchain-based smart contracts/auditing
- Secure data/algorithm sharing sandbox
- Support for third party analytics frameworks

AI TECHNOLOGIES

Management and support for the complete AI lifecycle

- Management of AI lifecycle and workflow
- Data visualization
- Training dataset building
- Model development and training
- Model serving and refresh cycle
- Federated analytics
- Model validation and test
- Support for the entire environment, including third-party analytic frameworks

INFRASTRUCTURE

Physical secure environment for AI-data exchange

- Global high-speed private network fabric to connect the various data sources and destinations
- Hyperconverged / distributed (GPU) platforms including tactical clusters
- Cloud and network neutral

EDGE

Full suite of edge computing solutions

- Ruggedized clients, networking, servers, and edge PCs
- Intelligent management solutions that can employ automation and technologies that allow for flexible, scalable deployments and reconfiguration

MODERN, SCALABLE ARCHITECTURE

The platform sits on a tried, tested, and trusted basis of hardware and software. This includes next generation storage platforms, software-defined open networking, and Dell Technologies VxRail hyperconverged servers hosted in agile cloud data centers, which also host tactical clusters close to the edge. Our range of structured and unstructured storage appliances provides the necessary capacity for the exponential rise of data.

Our family of compute platforms including networking, ruggedized clients, servers, and edge PCs helps capture, analyze, and gain insights collected from edge devices. Combined with our robust ecosystem of systems integrators and AI solution providers and a marketplace available for sharing of data and models, we help you design, deploy, manage, and scale your AI solution as you grow.

AS-A-SERVICE SOLUTIONS TO SIMPLIFY DEPLOYMENT

The APEX subscription-based model allows for the full power of the platform without requiring an upfront investment. We work with your team to deploy the training model and we manage edge devices as part of a complete, secure infrastructure platform, built on VMware technology. Our ongoing partnership with VMware and our partner ecosystem translates to consistency across infrastructure, data, applications and security combined with seamless interoperation between all components.

THE DELL TECHNOLOGIES ADVANTAGE

The Dell Technologies comprehensive portfolio spanning from edge to core to cloud makes this the ultimate platform for federated AI. A retail enterprise can optimize its sales both in-store and across a chain, reacting in real-time to buying trends. A financial institution can analyze customer behavior and prevent fraud without releasing individual identities. A healthcare institution can investigate trends in well-being without releasing patient data.

It provides the right AI solution for delivering maximum insight accuracy and lowest real-time latency without the bandwidth overheads of a centralized system.

Our global network of industry experts coupled with a federated AI solution designed for simplicity, scalability and efficiency means that you can reduce the risk, cost, and complexity of implementation.

As demand for automation and system autonomy grows across industries, organizations are needing a platform to bring real-time analytics to the edge. That's what the Dell Technologies Federated Platform delivers, along with the ability to unlock the value of your data to deliver game-changing actionable insights and desired business outcomes.

Learn more about our AI solutions.

Contact an AI expert.

Connect with us.

1. https://www.gartner.com/en/documents/4004630/forecast-internet-of-things-endpoints-and-communications-worldwide-2020-2030-2q21-update
2. https://www.delltechnologies.com/asset/en-us/solutions/infrastructure-solutions/industry-market/delivering-the-ai-enabled-edge-with-dell-technologies.pdf
3. https://www.frost.com/frost-perspectives/challenges-of-adopting-edge-computing/
4. "Machine Learning for Predictive Maintenance. A Boeing 747 Bleed Air Valves case study". Test performed in the Dell Technologies lab.

© 2021 Dell Inc. or its subsidiaries. All rights reserved. Dell Technologies, Dell, EMC, Dell EMC, and other trademarks are trademarks of Dell Inc. or its subsidiaries. Other trademarks may be trademarks of their respective owners.

DELL Technologies