VDC Research

July 2025
IoT & Embedded Technology

Accelerating Edge AI Innovation

Tools & Strategies for IoT Differentiation

by **Chris Rommel**, *EVP*, and **Dan Mandell**, *Director*

License to Distribute:

arm

Introduction

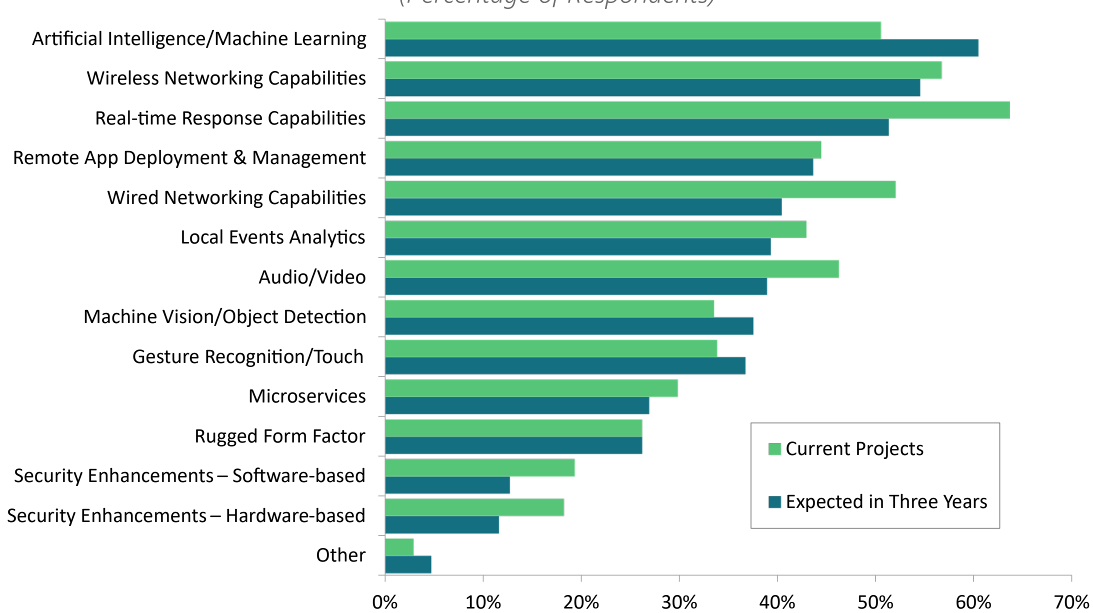
From edge gateways to end nodes, the IoT redefined product development and operational strategies. Engineering firms conditioned to define product value through physical system performance had to adapt. Connectivity – and the capabilities born from it – rapidly recast traditional product functionality goals. Services such as over-the-air updates, functionality provisioning, predicative maintenance, and location/context-aware services fundamentally changed the mechanisms for customer engagement and value delivery.

> By 2028, AI will be the dominant technology used across IoT projects

A new frontier for differentiation has now emerged – edge AI – pushing development organizations to adapt and evolve even more rapidly than they did during the rise of the IoT. Embedded systems are becoming more intelligent, with more sophisticated workloads. Intelligent automation and autonomous systems' functionality requirements are challenging system builders to adapt their development skillsets, processes, and tools. In the same way artificial intelligence has reshaped the enterprise computing landscape, AI is now redefining embedded computing needs and capabilities, creating a fabric for High Performance IoT (HPIoT) devices and networks that allows for the combination of advanced processing, low latency, and reliable, secure operation. While the transformation to edge AI capabilities is uneven and displays a range of maturation, engineers, developers, and product leaders recognize its arrival and impact. Edge AI development, however, is moving many organizations into uncharted territory where they lack the resources and/or expertise to efficiently bring solutions to market. Furthermore, traditional threads of compute and design requirements remain, with many organizations still needing to navigate real-time, security, and safety-critical development hurdles.

Exhibit 1: Capabilities/Features of Current IoT Projects

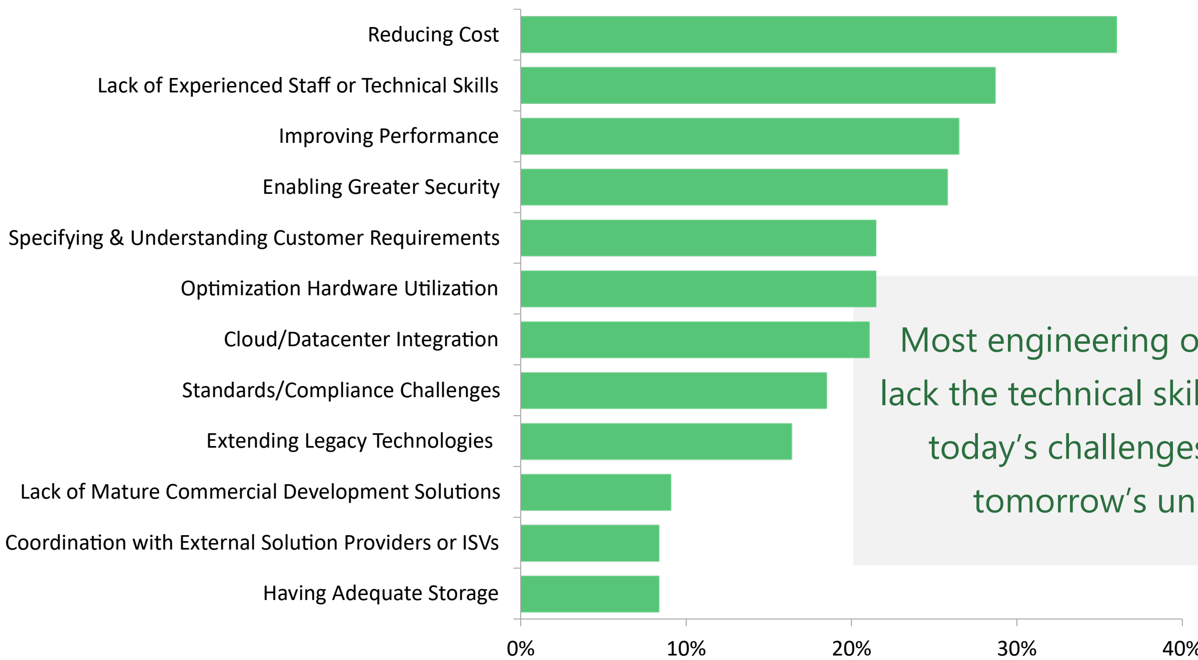
(Percentage of Respondents)



The pursuit and implementation of edge AI fuels not only tremendous opportunity, but also new challenges that require a revaluation of existing tools, components, and resources. In fact, next to cost, lack of experience is cited as the top issue impacting software and system development challenges [See Exhibit 2]. Other commonly cited issues highlight the importance of identifying solutions supported by robust ecosystems, such as the performance to meet dynamic workload needs, the availability of capable hardware, as well as existing security concerns, the impact of which can be increasingly acute with the advent of more complex and data-driven applications. With engineering organizations forced to innovate and deploy at unprecedented rates, informed development solution choices impact not only efficiency but also an organization's strategic direction. In many instances, enterprise and client markets have provided a foundation of technologies and processes to address edge AI challenges. Over the course of this paper, we will highlight key trends, challenges, and solutions for leveraging and adapting AI for the unique requirements of edge applications.

Exhibit 2: Top Factors Driving Software and System Development Challenges

(Percentage of Respondents)



Most engineering organizations lack the technical skills to navigate today's challenges, let alone tomorrow's unknowns

Background on VDC Research

VDC has been covering the product development and deployment technology market since 1971. The analysis and supporting discussions in this paper are based on VDC's ongoing research in this market, findings from a global survey of 275 edge AI decision makers, and in-depth interviews with key IoT product development organizations. This research, sponsored by Arm, offers insight into leading business and technical trends impacting IoT organizations as well as the best practices implemented to address them. The respondents are based across a range of industries including aerospace and defense, healthcare, industrial automation, retail automation, and smart cities, among others.

© 2025 VDC Research

Tech Changes Needed to Fuel HPIoT

IoT Forcing Reevaluation of Incumbent Technologies

The IoT is forcing a universal reevaluation of incumbent technologies. The evolution of device requirements is necessitating new technology choices – from operating systems, to security, to processors, and beyond. These decisions have become even more critical given how new systems will enable emerging high-performance workloads and IoT business goals, from collecting operational data, delivering new connected services, and interfacing with existing infrastructure. It is now critical for development organizations to identify technologies that can help them navigate these rapid areas of change.

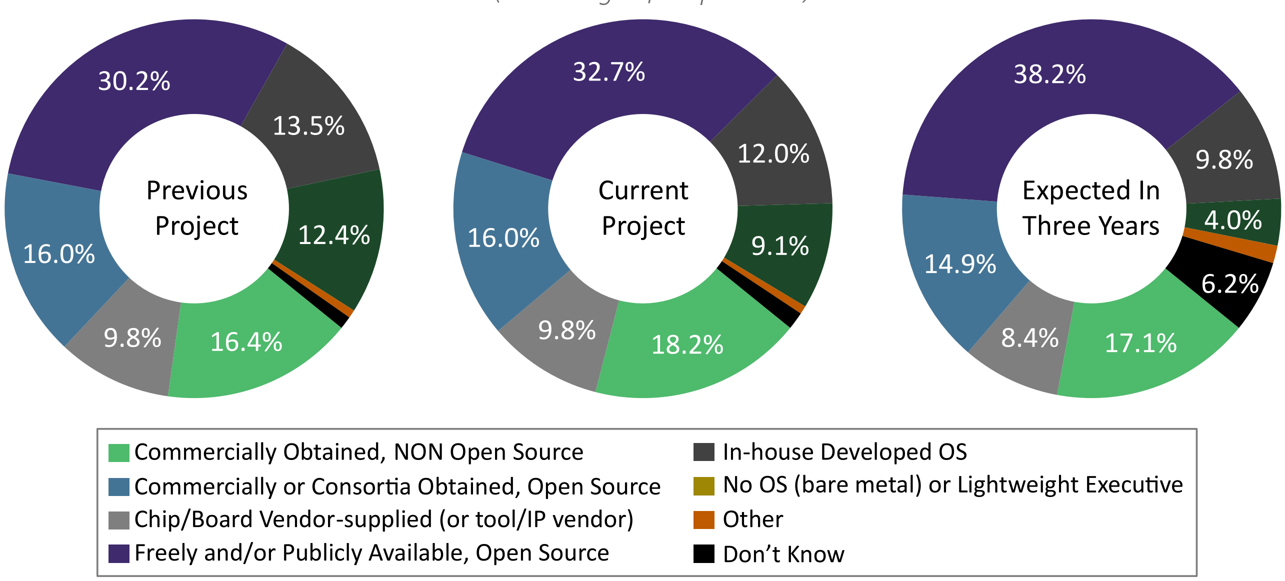
> Over 50% of IoT engineers plan to use an open source OS in 3 years

Modern Edge Workloads Demand Open, Scalable OS

For years, fragmentation was the unifying theme within the embedded market. Diverse system form factors, configurations, and operating requirements engendered the creation of highly specialized solutions. Nowhere was this dynamic more apparent than within the operating system (OS) landscape. Dozens of different operating systems, designed to meet specific sets of resource constraints or performance requirements (e.g., determinism/RT-latency, footprint, lack of a memory management unit, or industry-specific middleware feature sets and certifications) emerged.

Exhibit 3: Primary Operating System Used

(Percentage of Respondents)

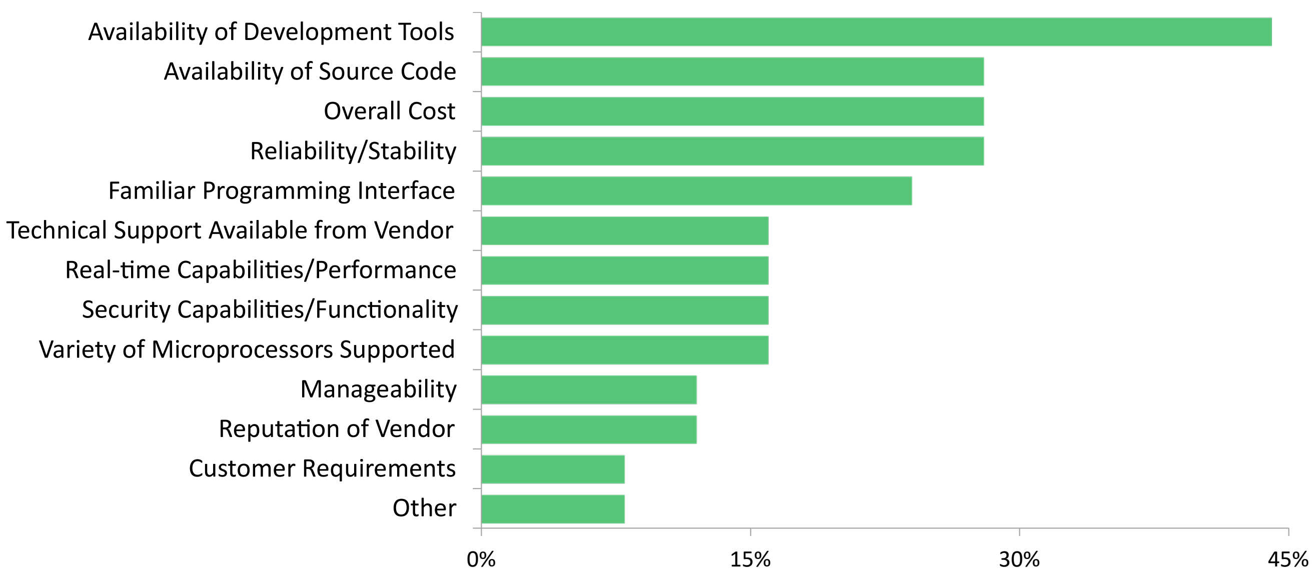


Now, however, the market is changing. New requirements for advanced functionality and edge AI are driving the adoption of rich operating systems. Open source OSs such as Linux have gained increasing adoption in the embedded marketplace, gradually displacing entrenched perceptions of the general-purpose run-time utility. The combination of increasing contributions targeted at the real-time domain, a maturing ecosystem of solutions, and increasing comfort and experience with Linux have driven its adoption to new heights, with more than 50% of the IoT engineers surveyed planning to use an open source OS in three years. Furthermore, increasingly sophisticated system-on-chips (SoCs) and a growing comfort with virtualization technologies now provide OEMs with multiple mechanisms to integrate rich, open source OSs with RTOSs. These combinations afford engineers the ability to leverage the capabilities and ecosystem of solutions like Linux while maintaining an underpinning of greater reliability, security, and safety with an RTOS.

When organizations adopt a rich OS, other factors become much more important to not only their selection of the runtime, but also their solution set in general – from the bill of materials to the solutions used across the software development lifecycle. For example, when organizations migrate away from systems without a formal OS, tooling support is paramount. As organizations adopt new solutions, they must ensure both optimization for their target hardware as well as efficient software development for their next-generation workloads. In addition to standalone tools and those provided by OS vendors, semiconductor ecosystem participants such as Arm and its licensees provide a range of SDKs for software and ML model development. For example, Arm's NN SDK together with CMSIS-NN and Arm Compute Library provide optimized inference engines and neural network kernels that are specifically designed for Arm-based devices. Likewise, the tight integration between these tools and LiteRT (previously TensorFlow Lite), TFLite Micro, and Ethos-U NPUs help provide hardware acceleration for ML workloads. The combination of these solutions with a strong partner network can offer OEMs the flexibility and functionality needed to accelerate their development for a range of workloads and hardware systems.

Exhibit 4: Most Important Characteristics for OS Selection when Migrating from Using No OS

(Percentage of Respondents)



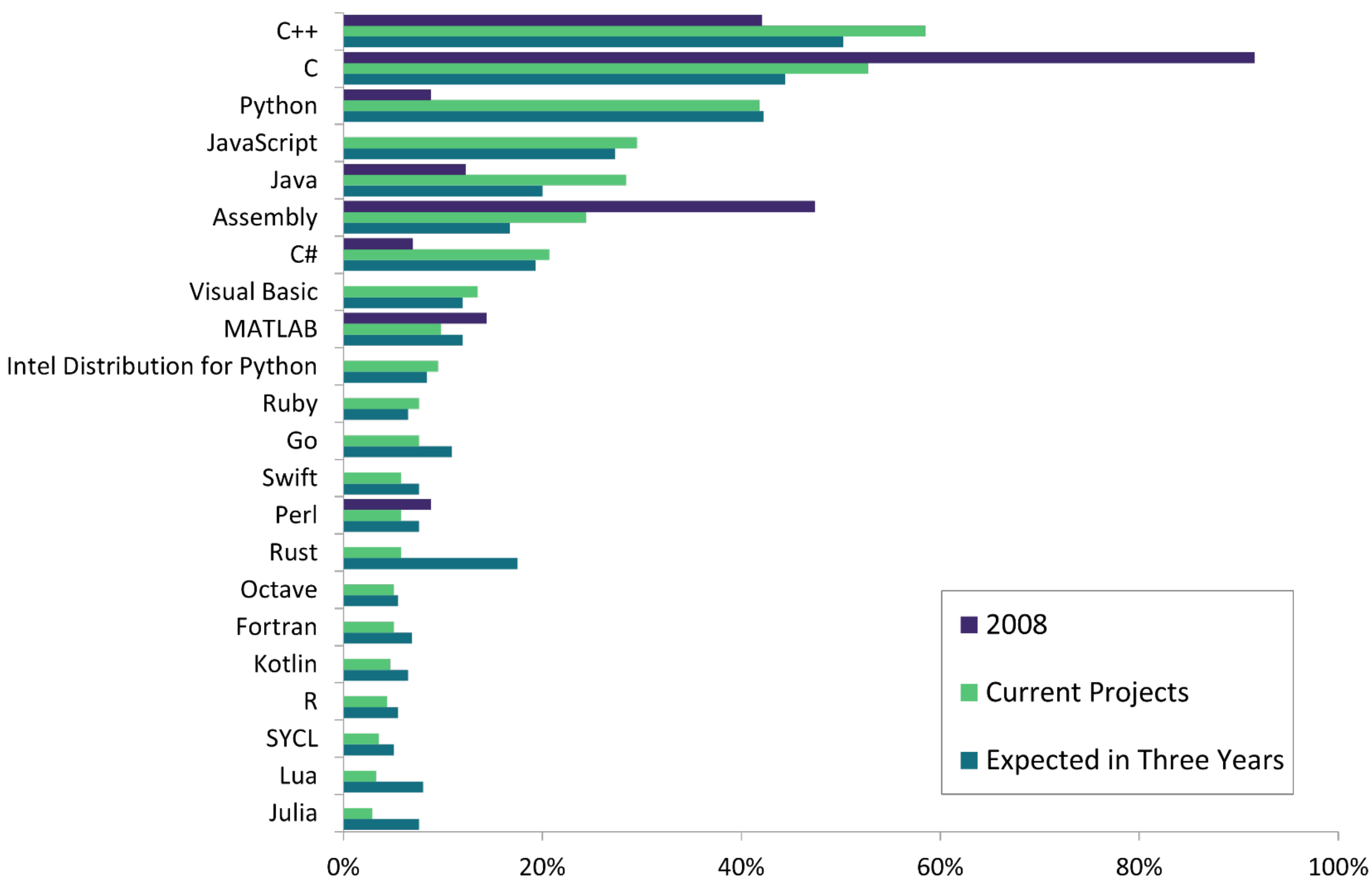
Changing Languages of Innovation

For decades, embedded software development was predominantly conducted in C and other lower-level languages. However, embedded engineering is no longer synonymous with small-footprint, fixed-function, single purpose devices. Today's devices often need to perform multiple functions and have multiple processors or cores to meet their system's current demands. Over the past decade, the requirements for high reliability and determinism that perpetuated the use of C, while still present, were subjugated in favor of object-oriented and higher-level languages such as C++ and Java. In fact, many of the common software development languages used today had little to no usage just a decade ago, highlighting the erosion of traditional embedded market inertia and accelerating technology changes.

> Python and languages targeted at high-performance compute are the mediums of edge AI innovation

Exhibit 5: Software Development Language(s) Used

(Percentage of Respondents)



The requirements for edge AI and the maturation of the ecosystems and libraries to support it have again recast the lexicon for innovation and content creation. The use of higher level languages will continue within this market due to the desire to abstract the complexities of new hardware, maximize the use of existing and future assets, and increase access to off-the-shelf connectivity and middleware options to further reduce time and cost of development, and enable more re-use across devices. Now, however, the rapid increase in use of Python, as well as of other languages targeted at heterogeneous compute and/or high-performance compute (e.g., Julia, MATLAB, R, SYCL, etc.), is changing development organization needs. In particular, the ecosystem advances with AI frameworks built with and supporting Python, such as PyTorch and TensorFlow, have fueled a resurgence in its use, beyond its traditional scripting use cases. This shift aligns with the overall acceleration in the rate of AI innovation and reinforces the value proposition of higher-level languages. Access to extensive libraries – many of which were initially developed for enterprise and client use cases – allows engineering organizations to tap into new areas of ecosystem expertise. In order to build and deploy across a range of edge devices, engineering organizations must identify platforms that offer integration with standard AI frameworks while offering compatibility with legacy code bases and IP.

Application Story

Growing Efficiencies with AI in Agriculture

This organization, based in Argentina, began designing a machine to dry cacao, a key ingredient for chocolate. The machine uses a combination of hardware and software to monitor and control the drying process, including temperature, humidity, and rotation of the cacao. A series of LED lights are used incorporating machine learning to optimize the drying process and product quality.

Their desire to automate a previously arduous task led them to integrate a range of legacy and new technology. The hardware uses a microcontroller (initially 8-bit, now upgrading to 32-bit) for sensor inputs and motor control, along with a more sophisticated Arm SoC running Linux. For software development, the company used C for the microcontroller and Python for the SoC, though they admitted not yet fully integrating machine learning libraries like TensorFlow.

> "Our future priorities are to expand our AI/ML knowledge and integrate it into our systems, as well as to explore cloud-based solutions, as this technology is still relatively new in our country."

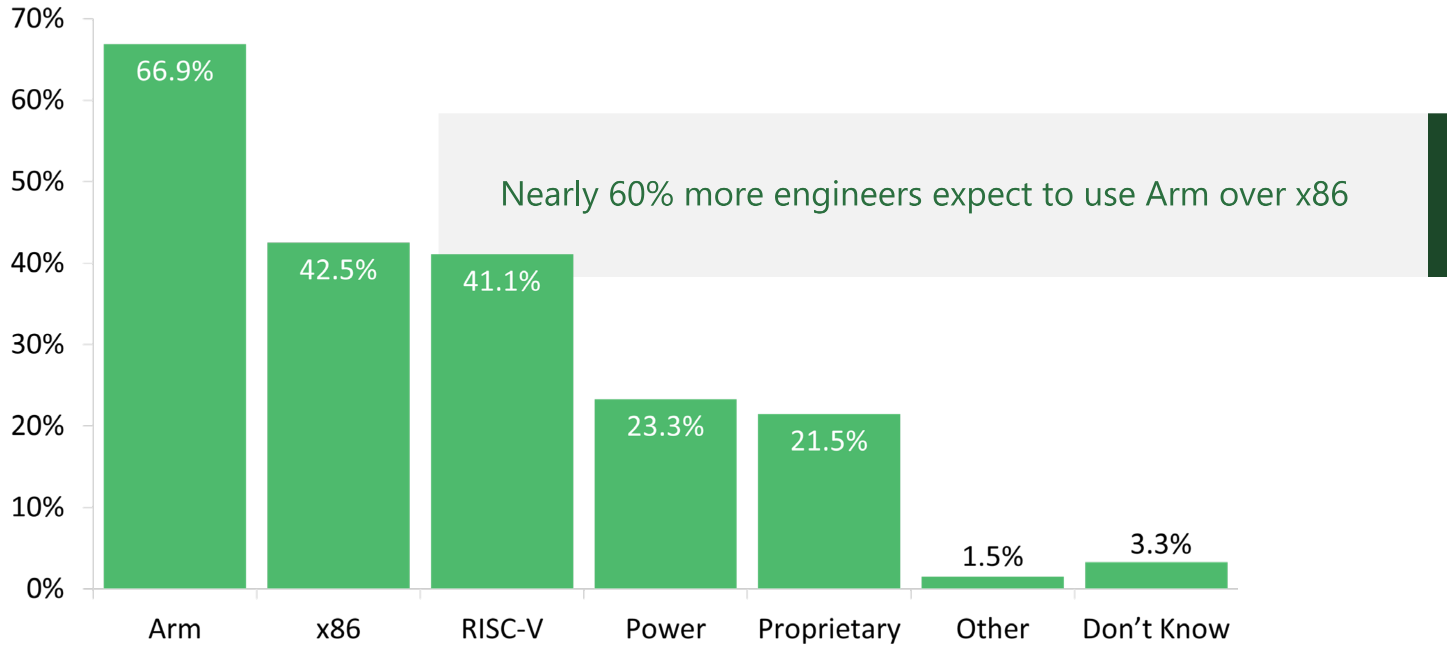
As this organization looks to upgrade its solution in the future, it will be critical to incorporate more advanced AI capabilities and models. Operating and innovating within a unique market segment, training data from their own deployments will be critical to optimizing their application.

Arm Powering a High Performance Future

The IoT market is incredibly complex and heterogeneous, fueling a diverse range of product form factors, functionality, and business models. There are a number of criteria that drive the selection of specific hardware and software solutions. To respond to new requirements, engineering organizations must consider both current requirements as well as investments made during past projects in legacy technology and IP. One such way this trend has manifested is through the choice of processing technology. The embedded and edge market, however, yields many more processor families than within the traditional IT/server landscape. Although there has been some level of consolidation around Arm and x86 processors over recent years, the sheer number of chipsets and the increasing pace product evolution present challenges to engineering organizations that do not effectively plan for the future.

Exhibit 6: Processor Architecture(s) Expected to be Used in Three Years

(Percentage of Respondents)



Over the past two decades, Arm has come to represent the leading choice among engineering organizations. Arm is not only already cited by greater than 50% of engineers as the primary architecture for their current projects, but over two-thirds of engineers expect to use the architecture within their projects in three years. Organizations can no longer afford to lock themselves into closed platforms that limit corporate agility or innovation. With a portfolio scaling from small-footprint MCUs to highly performant CPUs, GPUs, and NPUs, Arm presents a range of solutions to semiconductor suppliers and OEMs alike, offering benefits of:

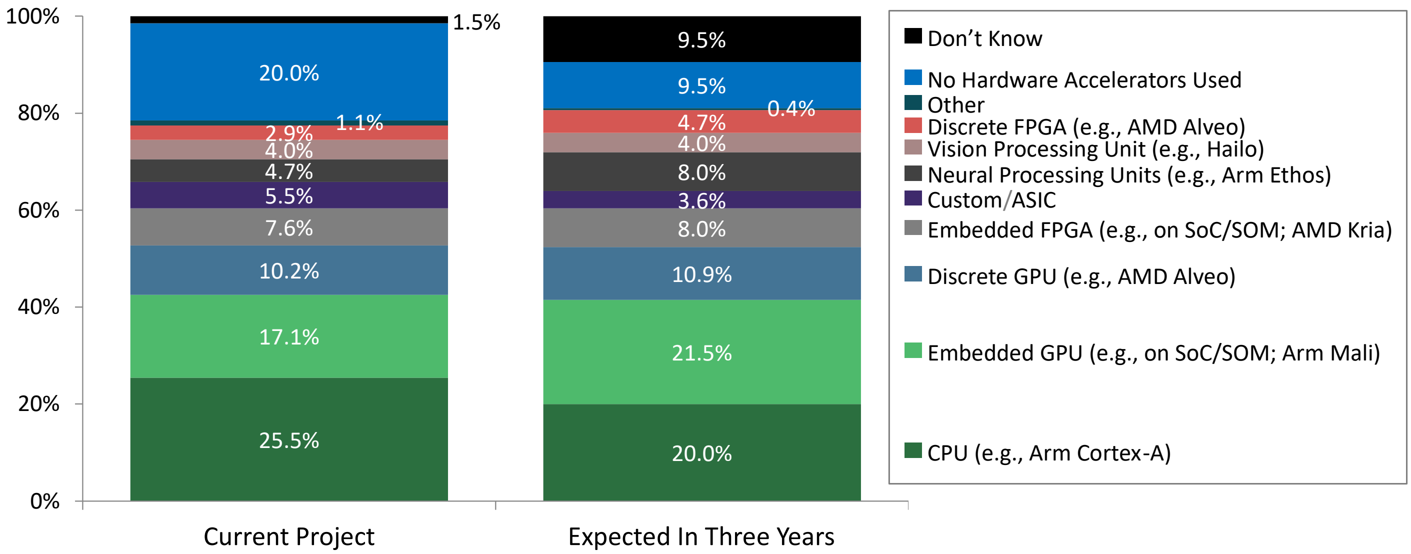
1. Power Efficiency
2. Scalability from Edge to Cloud
3. Performance and Compute Efficiency
4. Ecosystem and Vendor Adoption

© 2025 VDC Research

Finding a flexible, scalable, and standard-based architecture has become a critical foundation for engineering organization success. Organizations must now identify processing solutions capable of addressing both their current and future design requirements. To this end, more engineers expect increasingly complex systems and SoCs, with a higher number of peripheral computing elements to manage their edge AI workloads [See Exhibit 7]. Accelerators can range from GPUs to NPUs to additional CPUs – all of which spawn new layers of both performance and design complexity for engineers.

Exhibit 7: Hardware Accelerators Use

(Percentage of Respondents)



Use of NPUs such as Arm Ethos will nearly double in just 3 years

These new options, which can be critical for both end system performance and efficiency, underscore the need for engineering organizations to identify not only best in class semiconductor technology, but also that which can ensure their development organization's efficiency. Already, the availability of programming tools ranked third overall when selecting a primary processor, behind performance and price. The growing complexity and time-to-market urgency for edge AI application development are necessitating even greater scrutiny on technology and tool selection. The IoT market is already responding in kind. Engineers with Arm-based designs, which are expected to represent a majority of projects in the future [See Exhibit 6], cited the availability of tools as a driver in their processor selection at an even higher frequency than the market as a whole.

Application Story

Seeing the Future of AI for Healthcare

Developing a new eye-tracking system that provides assistive typing technology for people with disabilities like cerebral palsy, this organization intended to leverage the newest technology available to create a new, long-term platform for future product development and innovation.

The new design aimed to reduce cost, size, and noise compared to previous versions. In order to address more complex requirements, the new eye-tracking system uses a single Arm-based processor instead of a series of 8-bit microcontrollers that relayed information to a PC in the previous design. This new design allows for more processing power and integrated features like quadrature decoding that are critical for rapidly interpreting eye movement for intended communication.

> "Our system is not low cost... but for the processing components, the cost versus performance compared to our prior design is unbelievable. It allowed us to fundamentally change our design and the user experience."

Tools for Success Required

Increased software complexity, scarce development resources, and more demanding end-user requirements are challenging engineering organizations to adapt. Exploding code bases and new functionality requirements necessitate OEMs to find new ways to simplify and accelerate software development. In fact, developers rated identifying ways to reduce complexity and abstract the coding process as the most important characteristic of edge AI projects, highlighting the importance of solutions identified as valuable in that regard [See Exhibit 8].

© 2025 VDC Research

Exhibit 8: Ranking of Importance of Characteristics for Edge AI Projects

(Average of Responses)

1. High Degree of Code Abstraction
2. Power Consumption
3. Performance portability
4. Portability
5. Performance

To address next-generation edge AI requirements, engineering organizations must identify tools that can speed development and offer optimization for more sophisticated processing elements. Many organizations look to semiconductor and semiconductor IP organizations, such as Arm and its licensees, as a first step for development solutions. While many vendors have helpful solutions, Arm has consistently rated higher than its peers.

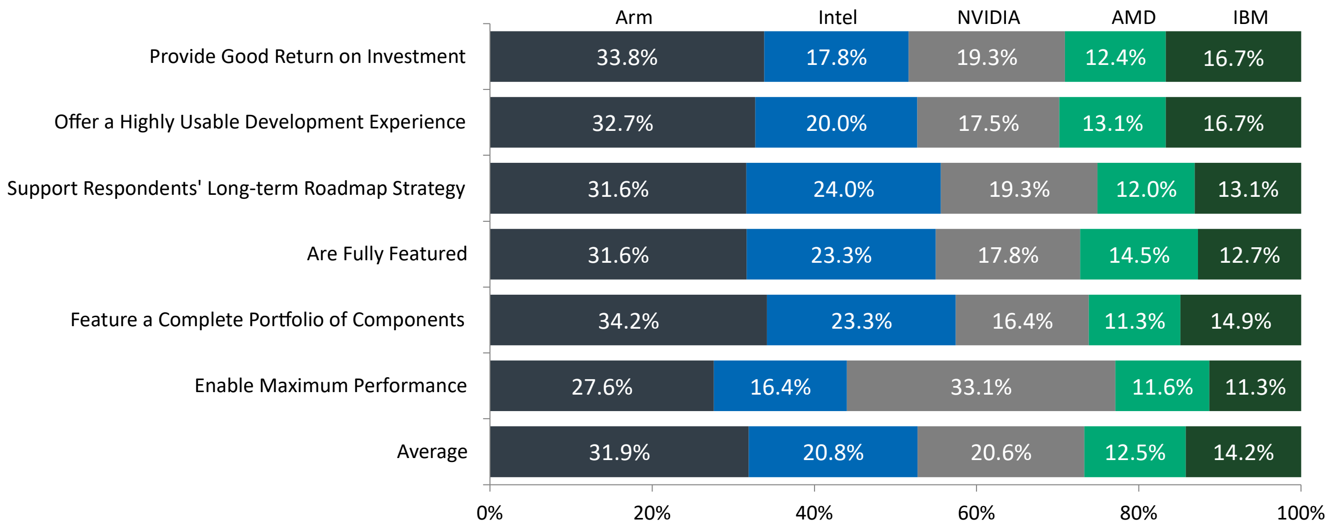
In fact, Arm has a long history of providing tools to help engineering organizations accelerate development – from RVDS to DS-5 to Keil to its more recent cloud-based Arm Virtual Hardware. Over time, Arm expanded its solution set to offer libraries and platforms that can help organizations 'shift development left' via high fidelity virtual prototypes. These solutions help accelerate the start of software development in advance of final silicon availability, removing traditional task serialization and bottlenecks. Moreover, the breadth of Arm's tool portfolio provides engineering organizations with a significant asset for scalability, spanning MCU-class devices to v9-based CPUs – as well as the virtual environments to allow developers to profile their workloads in each of these environments.

In its recent v9 Edge AI platform launch, Arm combined enhanced software reuse with native AI acceleration for Ethos NPUs, integrated security, and AI model portability between cloud and edge workflows. In addition, this new platform added significant performance uplifts versus previous generations. Now, engineering organizations are afforded additional flexibility, with the ability to choose from a range of solutions from v9-based AI acceleration or Ethos NPUs, depending on power and performance requirements. The combination of Arm's unique ecosystem position as a widely-embraced IP provider with its decades-long investment in software design capabilities has made it the highest-regarded tool provider amongst its peers [See Exhibit 9].

© 2025 VDC Research

Exhibit 9: Semiconductor/IP Provider Perceived to Offer Software Tools that:

(Percentage of Respondents)



Arm is recognized as a clear leader in providing Edge AI software tools, cited at a rate 50% greater than its closest competitor

Application Story

Automating Inspection with AI Aerospace

Traditionally, mundane maintenance tasks for commercial aircraft go unnoticed to air travelers. One such undertaking is the regular replacement of the thousands of countersunk fasteners on the exterior of aircraft fuselage. To remove old rivets, manual processes required time-consuming drilling that was rife with risk of damage to aircraft.

This organization's design brings safety and efficiency to that process. They are developing an inspection tool for an e-drill system that uses electrical discharge machining to remove the fasteners. Utilizing an Arm-based System-on-Module with a neural co-processor, the tool uses a real-time vision system to gauge the concentricity of the cut, eliminating the risk of damage to the aluminum aircraft skin. Currently, however, the user still must identify the types of fasteners used on the system to ensure proper cuts, introducing delays and additional chances for errors. Now, the organization is integrating machine learning to automatically identify the fastener type to add further efficiencies to the maintenance process.

"The challenges are always manpower, the labor, and the time it takes to develop these things… Everything that we do is so tightly coupled to the hardware and has real-time requirements, inputs, outputs and sensors, so any efficiency we can gain through integrated and AI-optimized solutions is a huge benefit."

© 2025 VDC Research

VDC's Summary & Recommendations

Engineering for Intelligence Demands Performance

The Internet of Things and the growth of edge AI are rendering many incumbent embedded engineering technologies and design processes insufficient and antiquated. Technology first widely deployed for client and enterprise systems – from GPUs to programming languages – is providing the edge industry a catalyst and mechanism for rapid evolution. Already, the growing role software plays within electronic system functionality and differentiation has upturned an ecosystem traditionally averse to change. Although many of the requirements for embedded devices remain unique and often industry-specific, others such as connectivity, enterprise integration, security, and artificial intelligence are becoming more important. The increasing number of interconnected embedded devices will place an even greater premium on the ability to effectively manage software upgrades and connected service delivery.

Future Innovation Driven by Present Software Choices

The fabric for continued edge innovation is woven by threads of software choices. The growing proportion of software functionality will magnify any existing development inefficiencies. OEMs will face mounting pressure to design and deliver software as a seamless part of the overall system architecture and product lifecycle, creating an infrastructure for continuous engineering and differentiation evolution. Engineering organizations now need new solutions that address these evolving requirements and speed development and time to revenue. Many leading firms are recognizing the foundational value that rich, open source operating systems can provide to help accelerate AI development – with over 50% of engineers planning to use an open source OS within the next three years. Beyond tapping into rapidly maturing ecosystems such as that around Linux, it is critical that organizations future proof their designs and architect their systems for flexibility and evolution. The pace of requirements change in the industry now necessitates the creation of devices with embedded software capable of post-deployment evolution.

Identify Partners Capable Supporting Edge AI Goals

Artificial Intelligence and machine learning advancements are rapidly expanding edge application potential. As a result, engineering organizations expect their need for highly optimized AI solutions to increase dramatically going forward, requiring increasingly complex systems and SoCs, with a higher number of peripheral computing elements. For example, surveyed engineers expect their use of NPUs such as Arm Ethos to nearly double in just three years, showcasing the pending impact of enterprise-grade solutions in the IoT.

© 2025 VDC Research

The recognition of value in edge AI solutions must be followed by the selection of a partner capable of providing the needed combination of products, scale, and support. Arm is an example of one such company that has built a foundation of products and expertise serving the IoT industry over the course of decades. With the complexity of IoT systems and their workloads on the rise, it is critical to identify a partner that is likewise invested in continued improvement and success. Arm has not only expanded its processing architecture portfolio to match industry needs but has also ensured its customers and OEMs alike have an ecosystem and tools for success.

About the Authors

Chris Rommel
Executive Vice President,
IoT & Industrial

Chris leads VDC's research programs and consulting engagements focused on development and deployment solutions for intelligent systems. He has helped a wide variety of clients respond to and capitalize on the leading trends impacting next-generation industrial and device markets, such as security, the IoT, and engineering lifecycle management solutions. Chris holds a B.A. in Business Economics and a B.A. in Public and Private Sector Organization from Brown University.

Email Chris at **crommel@vdcresearch.com**

Dan Mandell
Director,
IoT & Embedded

Dan supports a variety of syndicated market research programs and custom consulting engagements in the IoT and Embedded Technology practice. He leads VDC's annual research services for embedded processors, boards, integrated systems, IoT gateways, and other computing hardware. Dan's insights help leading technology providers align their go-to-market planning and competitive strategies with the dynamic embedded landscape and its constantly evolving buyer behaviors, technology adoption, and application requirements. Dan holds a B.S. in Information Systems Management from Bridgewater State University.

Email Dan at **dmandell@vdcresearch.com**

About VDC Research

Founded in 1971, VDC Research provides in-depth insights to technology vendors, end users, and investors across the globe. As a market research and consulting firm, VDC's coverage of AutoID, enterprise mobility, industrial automation, and IoT and embedded technologies helps our clients make critical decisions with confidence. Offering syndicated reports and custom consultation, our methodologies consistently provide accurate forecasts and unmatched thought leadership for deeply technical markets. Located in Southborough, Massachusetts, VDC prides itself on its close personal relationships with clients, delivering an attention to detail and a unique perspective that is second to none.

For more information, contact us at info@vdcresearch.com.

© 2025 VDC Research