JPMorganChase

EMERGING TECHNOLOGY TRENDS:
JPMorganChase Perspective

Global Technology Strategy, Innovation and Partnerships

Executive Summary

Technology innovation serves as a catalyst for transformative change. It propels the advancement of technology, unlocking novel capabilities, products, markets, experiences, and productivity opportunities for our esteemed organization.

Through collaborative efforts with leaders across the firm and the external technology ecosystem, the Global Technology Strategy, Innovation, and Partnerships team ensures that JPMorganChase remains meaningfully connected to these innovative and emerging trends that will shape our strategic outlook in the immediate and long-term future.

This document outlines comprehensive thematic overviews and industry perspectives, representing pivotal areas of innovation. Last year, we witnessed the emergence of foundational models and there was an all-out AI arms race as hyperscalers started building their own language models - with emphasis on scalable and secure solutions.

AI continues to spearhead the charge, with an array of use cases that continue to emerge. Notably, there is a growing investor interest in AI agents, LLMOps (governance), and specialized AI hardware. Hyperscalers are actively revamping their infrastructure and distinguishing themselves through the development of innovative hardware and specialized offerings. Additionally, AI-driven super apps are rapidly gaining traction as a comprehensive destination for employees, enabling them to seamlessly access and engage with digital content while enhancing their interaction with the physical office environment. Finally, it is worth noting that emerging components of AI/ML security, such as scanning open-source models for vulnerabilities or malicious code, are gaining prominence as organizations strive to ensure the integrity and robustness of their AI systems.

To read more about how JPMorganChase is driving innovation, **click here** to view materials from our 2024 Investor Day.

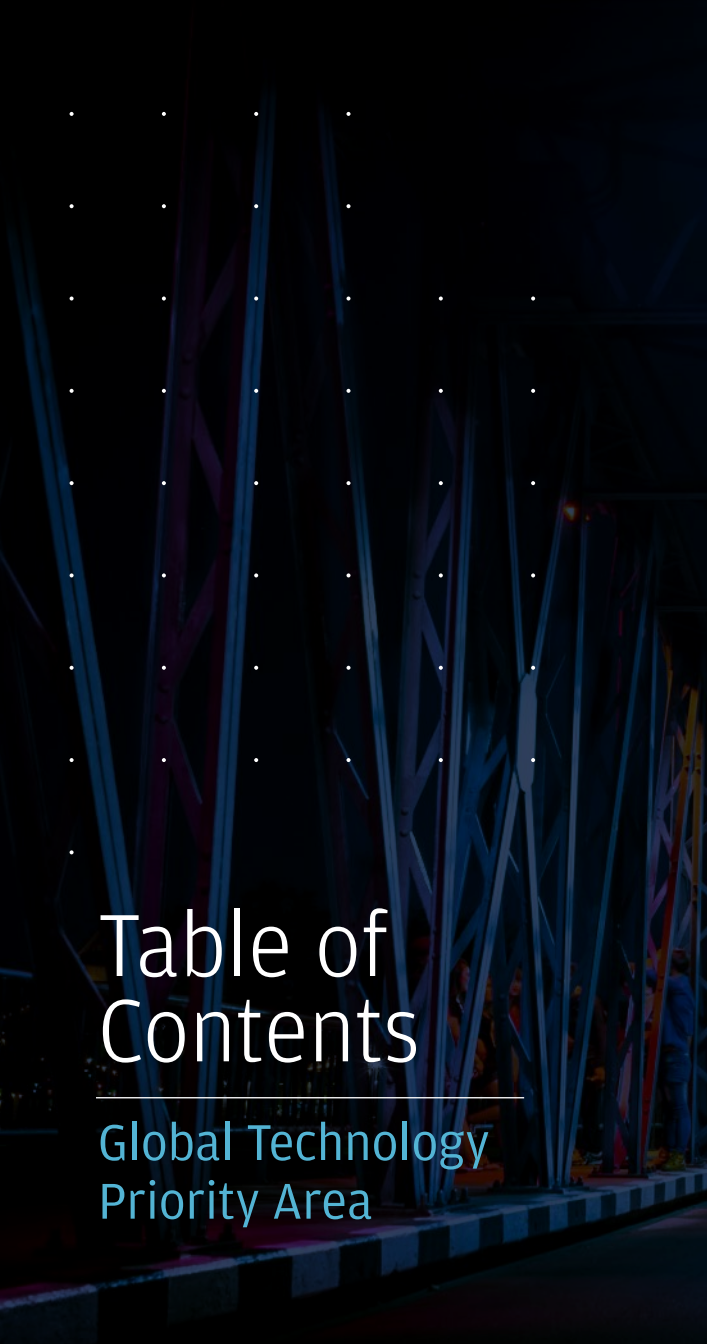


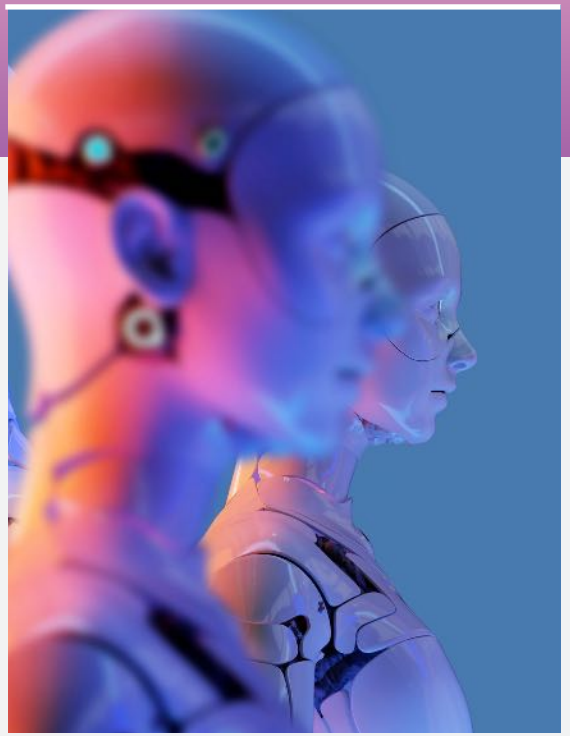
Table of Contents

Global Technology Priority Area

Modernizing our Tech

Rapid enterprise adoption of cloud, the rise of Generative AI (GenAI), and fierce competition within the industry is driving a fresh wave of infrastructure innovation. Leading public cloud providers are revamping their infrastructure and differentiating through new hardware development and specialized offerings. Simultaneously, the software ecosystem is creating novel solutions that enhance cloud security and efficiency, further accelerating the migration process.

AI Demand Transforming Infrastructure



While the last decade was focused on building large, web-scale data centers filled with commodity infrastructure, the current era is focused on providing access to workload-specific computing environments that access a myriad of specialized technologies to improve overall price performance for modern AI use cases.

The public clouds, as an example, are investing heavily in their core infrastructure to entice new AI startups to their platforms over competitors, be it for research, training or inference. Billions of dollars are being spent – and as a result the industry is seeing a renaissance in infrastructure innovation. From compute, network and storage, to new data center designs, every layer of the stack is being optimized to support the tremendous scale that new AI use cases require.

In addition to hardware, Infrastructure as a Service (IaaS) providers are also investing heavily in new software techniques (e.g., scheduling, capacity reservations) focused on allowing customers to more efficiently orchestrate their AI workloads on otherwise costly AI infrastructure in a bid to make their platforms the most economic choice. Still, it's the hardware performance growth that's causing the most change.

With the explosion of demand, the focus on AI infrastructure is not only dominated by the large public clouds, but there is a growing number of specialist providers focused solely on the delivery these capabilities. These "AI Clouds" represent a new category of competition for the large providers.

Market and Industry Perspectives

- Significant supply chain issues have arisen as demand has soared, and understanding this dynamic and future capacity requirements will require ongoing planning.
- The public clouds are making significant investments in their infrastructure to compete. These range from AI chip developments, enhancements to networking and storage performance, and new data center designs.
- Adjacent to public cloud innovation is a growing ecosystem of 3rd party specialist solutions that offer like GPU specific clouds, high performance networking and storage, and cutting-edge chip design.

Chip Competition Continues

The semiconductor industry has been rapidly evolving in recent years, and the pace of change will continue to increase in 2024. As is typically the case in this market, the drivers of innovation are focused on increasingly improving price to performance; but with demand far outstripping supply, the ecosystem is changing to accommodate more and more entrants with a greater emphasis on workload specific chips.

Starting in 2024, major public clouds will have meaningful in-house chip development programs. These efforts can be split into broadly two categories – more efficient, generic computing chips using ARM based architectures, and domain specific chips for particular workloads (e.g., AI Inference).

Compared to the CPU market, GPU competition has been relatively stable with current leading incumbents. There is movement, however, as vendors have made considerable progress with their hardware performance and have focused on software integrations with popular AI frameworks. With competition comes even more innovation, with some players even using AI itself to develop its new chip designs.

Lastly, startups are bringing to market chips that are even more specialized; for example, chips focused on a specific type of algorithm, which may lead to a further proliferation of chips available in the market.

Market and Industry Perspectives

- At this point, all three major cloud providers are developing in-house silicon, be it for general purpose computing or AI. This trend is only expected to accelerate as customers begin adopting and further investment can be made.

- Matching workloads to optimal chip architectures will likely be a significant part of an overall public cloud efficiency program. For the right workloads, ARM based chips offer significant power savings, and in the AI ecosystem there is significant concern over the long-term operational costs of inference, which could be improved by domain specific CPUs.

- The rapid innovation taking place at the hardware layer, born out of significant demand for high performance compute, is likely here to stay for the medium term. Longer term, abstraction technologies may play a role in simplifying workload placement – but in the short-term organizations will benefit from investing in both their understanding of, and use of these new workload specific architectures.

Quantum Computing

The semiconductor industry has been rapidly evolving in recent years, and the pace of change will continue to increase in 2024. As is typically the case in this market, the drivers of innovation are focused on increasingly improving price to performance; but with demand far outstripping supply, the ecosystem is changing to accommodate more and more entrants with a greater emphasis on workload specific chips.

Starting in 2024, major public clouds will have meaningful in-house chip development programs. These efforts can be split into broadly two categories – more efficient, generic computing chips using ARM based architectures, and domain specific chips for particular workloads (e.g., AI Inference).

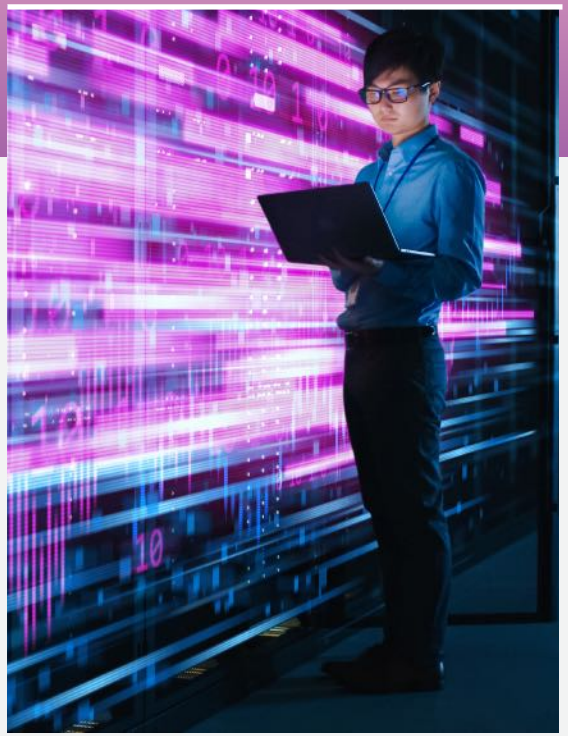
Compared to the CPU market, GPU competition has been relatively stable with current leading incumbents. There is movement, however, as vendors have made considerable progress with their hardware performance and have focused on software integrations with popular AI frameworks. With competition comes even more innovation, with some players even using AI itself to develop its new chip designs.

Lastly, startups are bringing to market chips that are even more specialized; for example, chips focused on a specific type of algorithm, which may lead to a further proliferation of chips available in the market.

Market and Industry Perspectives

- Advancements in hardware and algorithms are reducing the number of qubits required for computations and increasing the speed and accuracy of results.
- Big tech firms have solidified roles across various layers of the quantum technology stack: hardware layer, circuit compilation layer (hardware interface), and algorithmic layer. Start-ups are also rising to pioneer advances across quantum hardware.
- Numerous ISPs and Telcos around the world are exploring QKD and starting to offer QKD as a Service (QKDaaS) solutions.
- Finance is among the first industry sectors likely to benefit from quantum computing due to the abundance of high-complexity financial use cases, where computations need to be completed in real time.

DevFinOps



As more enterprises migrate to cloud, often they face rising cost challenges due to the improper use of the on-demand compute and resources the cloud offers. For years, public clouds and their partners, emerging startups, and the open-source ecosystem have developed tools to help users manage their cloud costs through enhanced cost visibility, cost optimization recommendations, reserved instances, compute forecasting, and auto-resizing workloads.

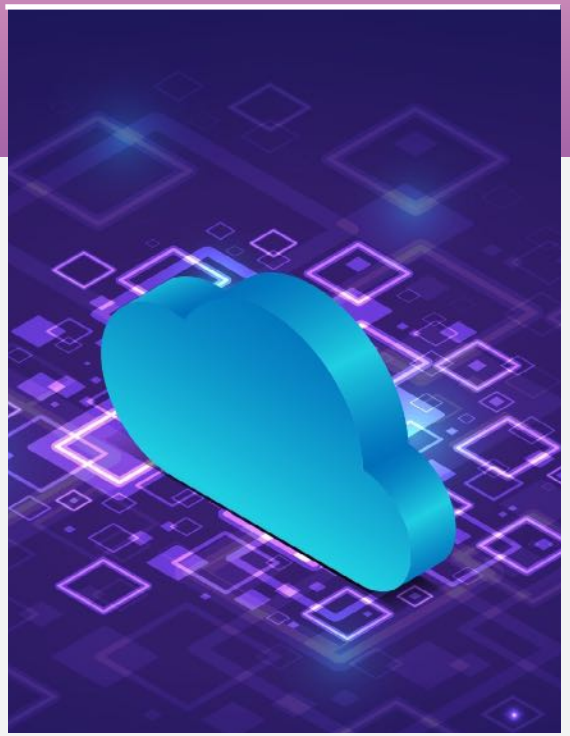
While the accountability of FinOps has traditionally been under the sole remit of an organization's Finance or Operations department, there has recently been a shift in responsibility to the actual users of the technology – developers and engineers.

Cloud architecture and costs are tightly related and poor architecture typically results in an inefficient use of resources (e.g., overprovisioning / lack of autoscaling). As such, there is a growing push to put cloud cost management in the hands of architects and developers who can design and build with costs in mind. As applications are developed, there are tools that will provide an estimated cost of the resources in their CI/CD pipelines before it is deployed into the cloud. This enables developers to build more cost-efficient applications from the start of the development process, and allows Finance, Operations, and Engineering to partner more closely together on managing an organization's cloud spend in the long term. Ultimately, a truly successful FinOps solution and implementation will be through a combination of proactive day 1 design efforts with transparency across these partnering teams.

Market and Industry Perspectives

- FinOps solutions are increasingly being marketed towards the developer community as the technology ecosystem places a high value on developer mindshare. In addition, evidence suggests that it can be difficult to implement optimization techniques after code has already been pushed to production.
- Incumbents in the cloud and observability space have shown renewed interest in the FinOps market with several acquisitions from large incumbents. In parallel, many DevOps solutions are developing their own FinOps solutions as part of their offerings.
- The FinOps Foundation was founded in 2019 and has grown to a 12K+ person community and represents more than 3.5K companies, including AWS, Google Cloud, and Azure. The Foundation published the first cloud billing standard known as the FinOps Open Cost and Usage Specification (FOCUS) in November 2023 and has been backed by the big cloud providers and more.
- JPMorganChase has been focused on migrating more workloads to public clouds as part of its multi-cloud strategy. As the firm's overall public cloud spend continues to grow, there has been increased focus on cloud optimization.
- We are proud to be the first FinOps Foundation member to complete the FinOps Assessment and achieve the FinOps Certified Enterprise certification, a key measure of advanced FinOps practices.
- Going forward, there may be opportunities to provide additional cost visibility directly to developers so that they can analyze the expected resource consumption of their applications *before* they push their code.
- Ultimately, the accountability for managing cloud costs will fall to a number of groups (i.e., SREs, product and platform owners), but developers are the first line of defense against inefficient cloud architectures.

Bring Your Own Cloud



While the adoption of SaaS continues to grow, the quest for better economics and efficiency has influenced innovation in customer data storage. In traditional SaaS models, data is ingested into the platform for use on that platform only. This leads to potentially high ingest and consumption costs but also encourages duplicate data (i.e., observability data being replicated into a SIEM). A new model has emerged often called “Bring Your Own Cloud” (also known as “Compute Over Object” or “Bring Your Own Bucket”). In this new architecture, compute and storage are decoupled – data is stored in a customer-controlled object storage location and compute runs independently (i.e., serverless queries), scaling up or down to meet the elastic needs of a business. With BYOB, customers are also in control of data ownership and retention policies, while also leveraging the lowest cost storage platforms.

BYOB spans multiple domains and use cases, notably observability, streaming, and data analytics. Data analytics vendors and open-source libraries can now remote query data in a BYOB environment. There are emerging solutions for streaming as well, such as being able to leverage BYOB for data ingestion and reduce the need to manage compute architecture (e.g., disk). More prominently, BYOB serves as a solution for observability and SIEM data. With open source and proprietary data storage techniques, BYOB vendors can compress this data to enable longer retention for regulatory requirements and extended performance/outage review windows.

Market and Industry Perspectives

- Many companies leverage solutions that allow them to remain on prem where data storage and compute are done in secure, private data centers. However, on prem solutions have shifted towards SaaS offerings as momentum with public cloud adoption has grown, with many deprioritizing on-prem support or ceasing it completely. With BYOB, companies are able to take advantage of new SaaS offerings without the high ingest and consumption costs, and also gives companies more control over their data.

- Reducing costs with better data privacy while also reducing vendor lock in would benefit data management, observability, and security teams. This market is emerging, but has seen a growth in solutions available today.

AI Powered DevOps Lifecycle

In the first half of 2023, AI for the Software Development Lifecycle (SDLC) primarily focused on code generation. Since then, tools have progressed from single line code auto-complete to more complex multi-line code generation and beyond. Chat interfaces have been added to Integrated Developer Environments (IDE). In the past year, we've seen a surge of AI tools now being offered across all phases of the DevOps lifecycle – plan, build, code, scan, test, deploy and observe.

AI has been introduced into the earliest part of the DevOps lifecycle during the planning phase, with content automation assistants aiding in planning software. AI code suggestion solutions have now evolved to generate code comments, tests, and documents, with more advanced solutions offering code explanation and re-factoring to help explain and update legacy code. Once code has been committed, there are tools to help with code review (i.e., suggesting line by line changes and explaining why).

The use of AI generated code will emphasize the need for stronger security scanning and debugging tools to ensure that the suggested code is safe to use and within compliance standards.

Once deployed to production, monitoring the health and status of these applications is becoming more efficient with the use of AI observability tools. Today, these tools largely focus on generating program language queries to quickly find data or summarizing incidents and events. Over time, these tools will offer more advanced AI-Ops capabilities to pinpoint the source of outages and recommend or execute playbooks to restore service quickly.

While the first generation of GenAI tools leveraged general purpose LLMs, new deployment models

now allow for the fine tuning or domain adaptation of LLMs leveraging a company's *private code repositories and documentation*. In theory, this will allow tools to deliver company-specific and even application-specific suggestions and assistance, ultimately enhancing the current solutions available today.

Market and Industry Perspectives

- GenAI for developers is arguably the most mature category in the GenAI domain. With solutions being released as early as 2019, the breadth of these tools has expanded as adoption has increased.
- As more AI tools come to market and adoption increases across the industry, it will become the expectation of developers and engineers to have AI tools available to them in job roles.

Innovative Products & Services

Customer and employee expectations continue to evolve, with GenAI unlocking a new set of opportunities for how we will build, design and consume experiences in the future. Whether it's how we build new FinTech products to how we design the latest enterprise SuperApp or how we consume content via an infinite workspace canvas, the latest wave of change will redefine expectations on what it means to have a best-in-class experience. Customers and employees will stand to benefit from new technologies to not only provide them with more seamless journeys across channels, but a highly personalized experience tailored to their individual needs.

Building a Data-First Strategy in a Cookieless Future

Marketers have relied heavily on data collected by third-party cookies to deliver marketing campaigns and attract customers. But key regulations, such as the General Data Protection Regulation (GDPR) and the California Privacy Rights Act (CPRA), security concerns, data privacy and the importance of consumer trust, have all contributed to the end of third-party cookies.

How marketers have historically targeted and attracted customers e.g., with retargeting or ad targeting using third- party cookies, is undergoing a fundamental transformation. Although a cookieless future may seem initially daunting, it opens new opportunities for marketers to develop a unified zero and first-party data strategy to drive hyper-personalized customer experiences.

First party data gathers information about a customer's direct interaction with a company's website and apps, from what they are browsing, clicking, viewing, reading to even the device they are using. Zero party data is information voluntarily and proactively shared by a customer via e.g., a survey, feedback form or preference center, that provides detailed insights into a customer's, unique preferences.

Imagine you're launching a new credit card towards frequent travelers. With first party data, you can understand previous transaction history, account balances and demographics. On top of that, you conduct an online survey to those same customers to specify their preferred airlines, hotels or perks. This access to richer data and insights can create highly targeted marketing campaigns to drive improved acquisition and a personalized customer experience.

Market and Industry Perspectives

- Major technology players have already started restricting third-party cookies for Chrome users with the goal to phase out third-party cookies for 100% of users by the second half of 2024.
- As a result of restricting third-party cookies, the use of contextual ads will grow and will be targeted towards a user's immediate interests as they are browsing a site, without having to rely on their personal data e.g., someone reading an article on tennis may be more likely to purchase tennis equipment.
- Multiple technologies have also emerged to enable enterprises to collect more holistic customer data, including Customer Data Platforms, Customer Experience Management solutions and Digital Experience Platforms. Data Clean Rooms also enable companies to share data with partners to obtain additional customer information for co-marketing campaigns.
- Universal Identifiers – a single identifier able to identify a user in the digital marketing ecosystem across devices and platforms - are also starting to gain traction as viable alternatives to third party cookies.

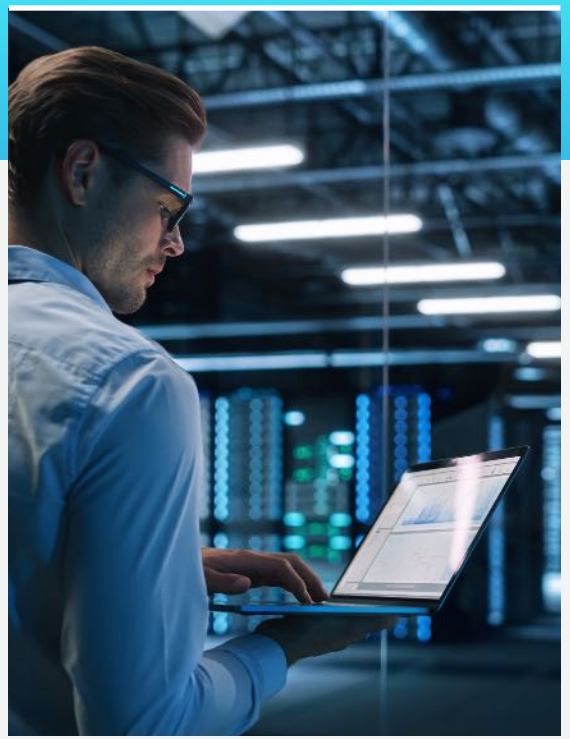
Distributed Ledger Technologies: Tokenization of Traditional Assets

Distributed Ledger Technologies (DLT), including blockchain, are transforming financial market infrastructures from payments platforms to investment management solutions. Despite the collapse of several notable crypto-native firms over the past two years, DLT technology and infrastructure have continued to advance with further focus placed on risk management and resiliency.

Technology enhancements have ranged from cryptography advancements using zero-knowledge proofs (ZKPs) to improve scalability and privacy, to better user experience through wallet and key management developments, and new interoperability solutions connecting blockchain layers and networks together. This has led to more enterprise-ready blockchain solutions paving the way for further institutional adoption.

One key area that is anticipated to see significant adoption in 2024 is the tokenization of traditional financial assets on blockchain-based platforms. Tokenization, as it relates to DLT, is the representation of ownership of an asset on a computer program called a smart contract on a blockchain. Our blockchain business unit Onyx by J.P. Morgan has been a leader in this space. The Onyx Digital Assets (ODA) blockchain platform providing asset tokenization and settlement has recently surpassed $1 trillion in tokenized asset settlement since its inception in 2020.

Limited interoperability of tokenized assets across different blockchains has historically restricted the scalability of securely moving value across different blockchain networks operated by other financial institutions. However, Onyx by J.P. Morgan



completed a proof of concept in collaboration with multiple industry participants successfully showcasing how interoperability could be achieved. Click here to read the report published Onyx by J.P. Morgan on how cross-chain interoperability could transform the asset and wealth management industry through a new paradigm for portfolio management.

Market and Industry Perspectives

- To further tokenization adoption, agreement on standards would allow for compliance rules to be embedded in smart contracts. A blockchain tokenization standard that was recently adopted for Ethereum, one of the most actively used blockchains, is called ERC-3643 which introduces compliance mechanisms and controlled advanced functions into tokenized assets. Onyx by J.P. Morgan has developed a token standard called “ODA-FACT” to provide a consistent way to represent fungible financial assets on the ODA blockchain with built-in control functions that are necessary to operate and manage regulated financial instruments. It is agnostic to the identity solution leveraged in conjunction with the token, enabling the use of either an on-chain or off-chain identity mechanism.

- There are currently several notable industry collaborations taking place to drive blockchain technology advancements to further tokenization adoption. One example is Project Guardian which is led by the Monetary Authority of Singapore (MAS) and includes buy-side firms, sell-side firms and policy makers exploring use cases to tokenize assets on open, interoperable networks. Onyx by J.P. Morgan has been an active contributor to Project Guardian leading two proof of concepts as part of the initiative.

- While Onyx by J.P. Morgan has surpassed $1 trillion in tokenized asset settlement on ODA, public blockchain networks also recently cleared an important benchmark. As of the end of March 2024, more than $1 billion in U.S. treasury notes and products have been tokenized on Ethereum, Polygon, Avalanche and other public blockchain networks.[1]

- Estimates for the total addressable market (TAM) for tokenization of financial assets varies, but most predict it will be a multi-trillion dollar amount by the end of the decade. Industry reports estimate a roughly $10 – $20 trillion TAM for the tokenization of real-world assets by 2030, which some argue is an underestimation. In terms of revenue, Onyx by J.P. Morgan recently co-authored a paper which describes the $400 billion in potential revenue that could be created by using tokenization to more effectively distribute alternative investments to individual investors.

[1]Coindesk, “Over $1B in U.S. Treasury Notes Has Been Tokenized on Public Blockchains”

About Onyx by J.P. Morgan

- Onyx by J.P Morgan continues to offer some of the largest and most mature institutional blockchain solutions operating at scale within financial services, via Onyx’s business verticals Onyx Digital Assets, Coin Systems, Liink by J.P. Morgan℠, and Blockchain Launch.

- Onyx Digital Assets (ODA), a multi-asset blockchain platform that enables tokenization of traditional assets, is settling on average $1-2 billion in tokenized assets daily. In October 2023, ODA also collaborated with JPM Securities Services to launch the Tokenized Collateral Network (TCN) to enable increased mobility of collateral through tokenization.

- Coin Systems, a blockchain network providing next-generation payment rails for value transfer, enabled Euro-denominated payments for corporate clients in June 2023 using JPM Coin. It is a permissioned system serving as a payment rail and deposit account ledger facilitating the movement of liquidity funding and payments in right time.

- Liink by J.P. Morgan℠, a blockchain network for secure and privacy-preserving information exchange, is increasing the reach of its Confirm offering, which verifies account information using blockchain technology, by expanding to more jurisdictions globally.

- Blockchain Launch, Onyx’s research and development team, engaged in its first-ever Digital Identity hackathon, in September 2023 drawing over 700 external developer participants and 70 projects submissions.

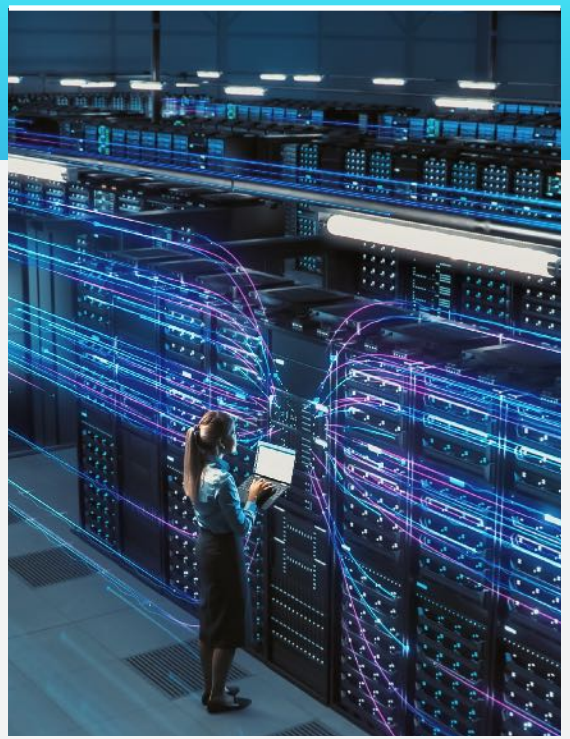
AI-Driven Digitized Workplace

To date, workplace applications available on mobile devices have been largely used for simple tasks like reading the intranet or booking a desk, and tasks that are one-directional with minimal user stickiness.
These apps and experiences will evolve to enable employees to better engage with the physical office through digital experiences.

Every touchpoint employees have with a physical element of the workplace will likely have a digital and connected interface, be that natively or through a standalone app on the employee's phone. Some organizations may even choose to build a 'SuperApp' which encompasses all of the digital touchpoints an employee has with the physical workplace.

These next generation workplace digital experiences will open a host of new possibilities including calendar integrations that can auto-suggest desk/meeting room bookings proactively based off an employee's calendar; facilitate navigation with tools like indoor digital wayfinding (e.g., "How do I get to Meeting Room 2A?"); ordering their lunch from their desk as well as enabling self-service capabilities with smart lockers, vending machines, and improved visitor management (e.g., visitor status, automated parking status/availability, coffee preference etc.).

To take this further, an Intelligent concierge – driven by AI - will integrate these capabilities into a unified experience that will help employees locate meeting spaces, book appointments, give directions, hail rides and supports day-to-day tasks. This will allow employees, including visitors, to interact with the building using natural language to easily navigate the physical space via their mobile device, enabling a seamless workplace experience.



Market and Industry Perspectives

- With the proliferation of SaaS applications, hybrid working and an increasing number of communications channels, employees will need a central place to find the services they need in their work lives. Organizations have taken note and have started either building their own proprietary workplace application, leveraging vendors to fast-track the development of a single application and one-stop shop for their employees.

- SuperApps have been around for several years, allowing consumers to seamlessly access multiple applications and services within a single user interface. Whilst some organizations continue to use multiple applications for each experience e.g., dining, mirroring the consumer experience we know today, others have started to take the route of building a SuperApp for a one-stop shop for their employees to access the services and amenities they need.

- Players in the market have started focus on creating a unified experience, integrating once modular components e.g., visitor management, into the wider Workplace app for a more holistic workplace offering and mobile platform.

Powering Seamless Employee Help and Servicing

The rapid advancements in conversational assistance enabled by GenAI in consumer tech will lead to employees expecting a similar experience at work to get help and information.

Navigating portals, chatbots, and documentation can be cumbersome and time-consuming. With the benefits of GenAI becoming available more broadly, employees expect to have their questions, requests and needs addressed in a natural language interface. GenAI will extend into areas like help and employee onboarding where employees can be serviced anywhere in the world for any request within a single pane of glass.

The employee experience of the future will use integrations with core systems e.g., talent management, learning management, HR data platform, collaboration tools, IT service management, workforce management., with orchestration across platforms invisible to the user presented in a single, conversational user interface (UI) to enable seamless employee assistance. Whether that's integration via existing communication and collaboration platforms, servicing employees in a seamless way will continue to be a large area of focus for enterprises on how to deliver best-in-class experiences.

These capabilities will be further extended by grounding based upon factual data sources like a company's public facing website and internal knowledge repositories to increase the accuracy of responses. A simplified journey will give employees the ability to self-service with ease, reduce total tickets and pressure on service centers, empowering employees with a personalized, consumer-grade experience.

Market and Industry Perspectives

- AI Copilots have proliferated over the past couple years with vendors in this space playing a key role in orchestrating actions across platforms to automate workflows over a wide range of use cases, including the IT help desk and employee onboarding.
- To improve the user experience, players in the market have started to integrate with major communication and collaboration players to provide a single, unified interface and hub where employees can interact with their AI Copilot to complete actions within minutes.
- The shift towards a single UI will not only improve communication but overall productivity, allowing users to navigate less tools and save time navigating multiple portals or help and request tickets.
- The employee onboarding experience – a critical aspect of the employee lifecycle and to ensuring employee retention - will also stand to benefit from this technology with access to the technology (including software and hardware) that they need from Day 1.

Enabling Immersive Productivity with Spatial Computing

Spatial computing is not a new concept; it gained popularity in 2003 when MIT Researcher Simon Greenwold described it as a specific sort of immersive tech that uses deep knowledge of the physical world to deeply blend the physical and digital worlds around the wearer. The tech is beginning to catch up with the theory to support true spatial computing experiences giving these devices new ways to leverage the blended environment.

A new generation of mixed reality displays use advanced sensors, AI, and display technology to let wearers work in the integrated physical and digital environment without the need for traditional hardware like a mouse and will soon offer a more complete Spatial Computing experience. While it has been possible to be productive using headsets released in recent years, limitations have been the cumbersome nature of the device, low resolution screens, and limited understanding of the physical world around the wearer. Devices are rapidly improving in these areas, making headsets a viable computing option for work and play.

The use cases are many and varied. Devices can be used to overlay digital information onto the real world to create new experiences and bring traditional desktop applications into the user's physical environment where they can interact with applications and spatial data in an infinite landscape, bridging the physical and digital world - no laptops or Virtual Desktop Infrastructure (VDI) required.

Eventually, users will also be able to navigate software applications with subtle micro-gestures or expressions (e.g., jaw clench or eyebrow raise), placing users more seamlessly within the computing environment to enable a fluid human-machine interface, instead of abstracting users from it.

Market and Industry Perspectives

- The evolution of mixed reality and the metaverse toward spatial computing will continue to be a major topic for the year ahead and will open new opportunities to collaborate, connect, and work. Multi-screen expansive virtual workspaces could mirror traditional at-desk setups to create interactive workspaces anywhere, enhancing how we interact with our applications and data. Vendors are currently offering solutions for general purpose headsets, as well as dedicated productivity-focused spatial workspace solutions to market to enterprises.

- While advanced gestural recognition, e.g., hand movements, is possible with spatial computing, the application of technologies that use micro- gestures or expressions to navigate workplace technology remains nascent. However, this could open new, more intuitive ways to interact with computing devices.

- Although spatial computing hardware is becoming more advanced (e.g. sensors and cameras) this technology is still far from widespread adoption. Initial use cases include training, hybrid meetings and events, immersive analytics and storytelling, and collaboration opportunities by immersing users in interactive 3D simulations. The command-and-control capabilities of spatial computing, whether using a headset or earbud, also opens tremendous potential to support employees with assistive technology requirements e.g., those who use wheelchairs or any wearable devices, the flexibility to navigate their workspace experience in a potentially more seamless way.

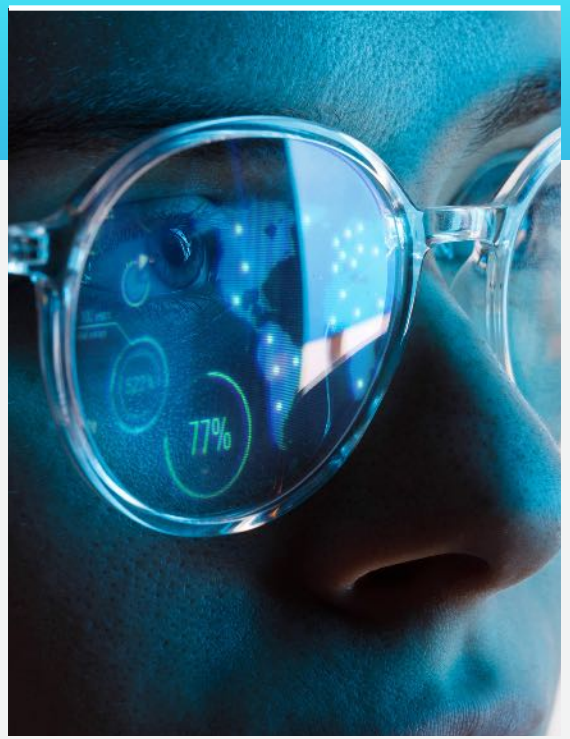
GenAI's Solution to Skills Gap Identification & Learning Journeys

GenAI has the potential to disrupt today's status quo of a one-size-fits-all approach to skills acquisition, by forming an end-to-end process for skills gap identification to personalized learning journeys for each individual employee while in the flow of work.

While traditional corporate learning technologies and approaches do not always prove effective for **employee outcomes** (e.g., failure of employees to acquire skills in skill priority areas, not personalized), nor are they always **cost effective** (e.g., enterprises may have limited funds and resources for corporate learning & development), GenAI has emerged as a potential enabler to take Learning and Development (L&D) to the next level. For example, GenAI can enable corporate L&D to save time by auto-generating personalized content based off existing knowledge materials.

As learning content becomes outdated, there is increased demand for learning content creators. GenAI companies in this space can generate entire course content (e.g., text, audio, video) in a shorter turnaround time compared to a traditional content creator albeit via a human-in-the-loop process. GenAI can also assess an individual's learning preferences (e.g., learning topics of interest) and role-specific topics to focus on, which will enhance the quality of content delivered to employees.

GenAI also has the potential to transform skills mapping. It has promise to aggregate skills through resumes, core HCM, professional sites (e.g., LinkedIn) and infer additional skills through work product and everyday tools. From here, skills can be mapped to a standardized taxonomy to further



identify gap areas to be successful in an existing position or a mobility opportunity. With human in the loop, employees and managers can confirm the skills and the depth of competency suggested by GenAI. Any gaps can be quickly identified with the end-user receiving an automated prompt to complete the learning module in their flow of work.

Market and Industry Perspectives

- While in the initial stages, large-scale players are beginning to take advantage of their vast app ecosystems to infer skills in the flow of work. One example from a large enterprise software provider is a learning experience platform built on top of existing learning management systems, that offers the ability to infer skills via other commonly used programs, which then maps inferred skills to a skills knowledge graph built on a skills taxonomy. We expect large scale players to augment traditional Learning Management Systems (LMS) providers by promoting learning content, rather than replacing them.

- Emerging players are differentiating themselves from large-scale alternatives by being further along in providing sophisticated analytics for skills & learning. For example, one provider offers insights into skill supply/demand, most effective courses, most frequent skill gaps, and hiring trends. While industry consolidation is inevitable, we expect providers who are most flexible in ingesting disparate data sources (outside of a closed app ecosystem) to be successful.

- From an **employee perspective**, GenAI's promise of personalization (i.e., more voice in preferred learning content) and easy access (in the flow of work) could enhance retention & mobility, while improving employee satisfaction.

AI-Powered Operating Systems

As we launch our devices, whether a laptop or mobile phone, a grid of applications is launched at the start of each session. With the advent of GenAI, how we have come to interact with the traditional Operating System (OS) is poised to change.

But what does that mean in practice? Traditional Operating Systems are built on top of hardware (e.g., an iPhone) and work as an abstraction layer to allow users to interact with and run applications. At the heart of the OS is a kernel, designed to manage a system's resources and act as a bridge between hardware and user-level applications, managing the processor, memory, files and access controls.

An AI-driven OS would see the LLM embedded directly into the kernel. This would enable user context and intent to be shared across the device, enabling the AI OS to invoke multiple applications and access files to solve a single user prompt, interacting with the user using natural conversation (whether by text or voice) – e.g., "Send the latest PowerPoint deck on public cloud migration to my manager and mark the task as complete on Monday. com." If an app is not downloaded by the user, the AI OS would find a suitable app available or could even build one. To take this a step further, customized user interfaces – whether webpages or applications – could also be auto-generated based on a user's request.

AI at the OS level would also allow users to effectively consume resources, e.g., compute and reallocating memory and storage, based on a user's individual needs across every application with AI embedded in each operation and process. Although it is early, AI OS has the potential to change how we interact with computing devices as we know it.

Market and Industry Perspectives

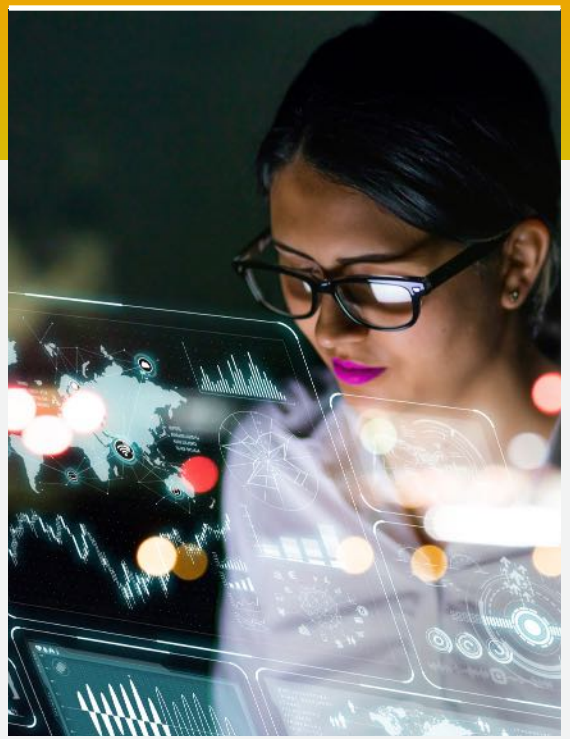
- An AI OS opens tremendous potential to uplift end-user productivity. Although we are still far away from seeing this come to fruition, there are already early developments in the market by vendors to suggest we have some of the components already to make AI OS a reality.

- An AI-powered OS would fundamentally change how employees use applications today across their devices – a type of change we have not witnessed for decades. Although this type of innovation is years away from being in the enterprise, it would allow employees to use any device they want and any applications by using a natural language interface to action complex tasks, potentially leading to significant productivity benefits. However, with the AI OS able to weave together actions across multiple applications and potentially develop new user interfaces instantaneously – amongst other future capabilities – significant efforts would be required on building out guardrails to protect the firm's data and continue to use AI responsibly.

Delivering Data & AI/ML at Scale

Hyperscalers, established firms, and emerging startups are racing to build models, platforms, and enabling technologies like AI Agents to drive the creation of next-generation use cases. However, the industry is also recognizing that to fully harness the power of GenAI, enterprises must tap into their own data. Enterprises are reassessing practices like data discovery, quality, and entitlements. The consensus view is that data management serves as the ‘picks and shovels’ in unlocking value from GenAI.

Delivering Automation with AI Agents & Orchestration



We know LLMs can generate content like marketing copy, images, video, etc., can summarize complex reports, and be a powerful tool for Q&A, but what if apart from these tasks, AIs could execute actions? This is a core idea of AI Agents. AI Agents are based on LLMs and can perceive diverse sets of input data (text, image, or speech), reason effectively (plan, search, learn from experience), and take action – without human intervention.

AI Agents action the next steps in a workflow on behalf of the user via a single prompt by creating tasks, executing them, collecting any feedback and effectively prioritizing based on the task list until the objective is complete.

To complete multi-step tasks, AI Agents can leverage a variety of tools like Excel, PowerPoint, enterprise systems, the Internet, and other Agents. Each AI Agent, armed with a specific tool and role, is guided by an LLM-powered orchestration engine, which directs the optimal decision-making sequence.

The potential of AI Agents is vast. Imagine a scenario where an AI Agent is given an objective ('prompt') e.g., "Create a social media campaign on LinkedIn targeted towards new graduates and post twice weekly for the next month." The AI Agent can produce a to-do list and action each step autonomously, including the ability to collect and analyze data based on previous campaigns, generate ideas, schedule posts and review performance.

Ultimately, the next transformational technology in the GenAI race will come from AI Agents and their ability to unlock substantial value across a variety of complex, multi-step workflows across platforms.

Market and Industry Perspectives

- Due to the success and popularity of open-source AI Agent frameworks, there has been a surge in investments from hyperscalers, model providers, and emerging startups towards the development of commercial AI Agents. This trend is expected to continue, with enterprise software vendors incorporating AI Agents into their platforms and core offerings, along with developer focused “build your own” Agent frameworks, which provide tooling for users to build agents for specific use cases.

- AI Agents hold the potential to provide significant value for large enterprises, especially for complex, multi-step workflows that are manual time-consuming tasks. For instance, in banking, the Know Your Customer (‘KYC’) process, is a complex multi-step workflow that includes document validation, background checks, identity verification and more. Similarly, in Loan Application Processing, employees must assess multiple financial documents, client credit history, and other risk factors. These are just two examples of tedious processes that can be automated using AI Agents.

GenAI Powered Data Management

In the financial services industry, data management has historically been neglected, leading to issues such as data siloes, duplication, and sprawl. It is crucial to understand that the root cause of these issues lies primarily in people and processes, rather than technology. However, the emergence of GenAI has brought about a significant shift in this landscape, and enterprises are now realizing the importance of having highly discoverable, classified, transformed, and machine-readable data to fully harness the potential of GenAI. This allows data consumers throughout the organization to leverage diverse data for model training and GenAI applications, fostering innovation.

Fortunately, GenAI-enabled data management tools are emerging to address these challenges. Data discovery is one area where significant progress is being made. Leading cloud data platforms and emerging startupds are aggressively investing in LLM-powered Q&A capabilities, which simplify access to data by eliminating the need for SQL expertise. However, the challenge remains to ensure that the discovered data is the correct data. To solve for this, vendors are investing in using LLMs to understand data structures, which leverage ontologies, and establish data relationships to enhance the traditional manual methods of data classification.

Another area where GenAI-powered innovation is making significant strides is in the transformation of unstructured and semi-structured data. While some vendors are utilizing GenAI for structured SQL data transformations, the ability of LLMs to handle unstructured data has led to the emergence of vendors that automate unstructured data into chunks. To effectively leverage enterprise data with LLMs, it is essential to prepare data in chunks that align with the context window size. As a result, data chunking is expected to become a critical component of data transformation and management.

Market and Industry Perspectives

- Virtually every enterprise data management software vendor is integrating GenAI into their core capabilities. This integration is enabling the democratization of data through natural language interfaces. By reducing the reliance on SQL expertise, these capabilities simplify data access and the complexity of using each platform.

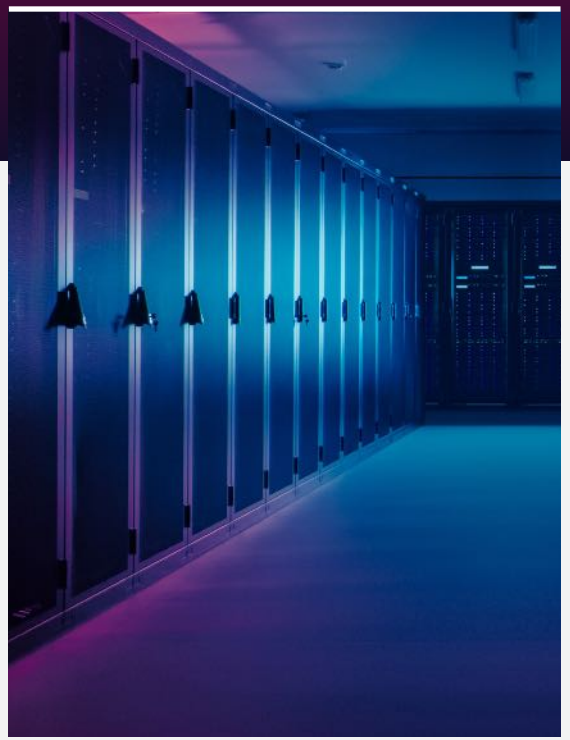
- For companies to fully realize the benefits of GenAI, they must prioritize data management. It is widely recognized in the industry that to enhance the accuracy of LLMs, a significant amount of diverse and high-quality data is required. To enable this, we must establish an environment where data producers take charge of their data's utility and publish it in centralized catalog accessible throughout the organization.

- In the current GenAI gold rush, data management has been recognized as the picks and shovels. Companies must carefully consider how to incentivize data producers to share high-quality data in a discoverable and controlled manner for our data consumers.

Protect The Firm

With accelerating changes in technology and the cybersecurity threat landscape, trends in security are focused on surfacing contextualized and prioritized risk, guiding proactive and AI-assisted remediation and continuously validating that cyber defenses are optimized for dynamic change.

Identity Security Posture Management



The cybersecurity industry has focused on contextualizing risk and the convergence of security posture management capabilities – across cloud infrastructure, applications and data – to optimally and efficiently prioritize cybersecurity risk reduction. Identity Security Posture Management is emerging as an additional pillar, providing visibility to true access and effective permissions across disparate systems and applications. This is helping organizations visualize identity-related risks and right-size access across their environment.

This visibility enables a policy-based approach to address over-privileged or under-utilized access and can help organizations accelerate default-based access patterns by modeling existing access for a given role and simplifying the request and recertification process. This visibility can also be connected to other forms of cloud risk posture, helping piece together a comprehensive risk picture and end-to-end awareness of potential attack paths.

These techniques also improve identity security monitoring by modeling access chains to understand potential attack paths and blast radius of standing access. This can be valuable in an incident response or threat hunting context to understand permitted access of a given identity at a point in time.

Market and Industry Perspectives

- As companies increasingly adopt SaaS services and continue to deploy more applications across multiple public clouds, the orchestration of Identity and Access Management (IAM) becomes more complex.

  Identity and credential compromise continue to be common elements in most breach activity.

- This category is poised to grow further in importance and vendor investment, particularly as other themes like generative AI and data science bring new identity challenges that will require broader and deeper visibility of who can and is accessing what. Emerging solutions are also focusing exclusively on the risk of non-human identities, as those proliferate and are more frequently subject to targeting and compromise as well.

Security Data Fabric

As the amount of data across the security ecosystem proliferates and is currently managed in siloed tools or targeting specific use cases, Security Data Fabrics are emerging to drive an integrated approach to maximize the value of disaggregated security data. While Security Information and Event Management (SIEM) tools have served as this aggregation point for security operations teams, the fabric concept aims to address a broader range of security outcomes, including unlocking security insight, aggregating risk context, facilitating prioritization and ultimately providing a platform for supporting AI driven security use cases.

New approaches seek to leverage a cloud-based data lake to aggregate and ingest data from a range of sources with a precise decision about what data is selected, then normalizing, correlating and creating logical context to facilitate downstream processes, interactive queries, and AI processes. Ultimately, these tools can help to facilitate an extended posture management or XPM-like approach by aggregating posture data across tools to drive connected context and facilitate “next best action” recommendations to optimally reduce risk.

Market and Industry Perspectives

- Several startups have emerged with a specific focus on providing a security data fabric that is distinct and separate from a SIEM.
- Data warehouse vendors are embracing these security use cases as well, with a strategy focused on partnering with security-focused vendors that are building on top of their platform as well as building security enabling features themselves directly into their platforms.
- Public cloud providers are investing here as well to leverage their native data services to manage security data beyond the scope of what they generate natively (e.g., ingesting from other security vendors) to ultimately position them to be the leader in AI security co-pilot like services.
- Major security platform vendors are similarly building out these data fabric capabilities to power their AI co-pilot ambitions as well.

Securing AI/ML

As the world navigates the acceleration of AI, so do threat actors, driving accelerated innovation in how organizations secure models as well. There are several emerging components of AI/ML security including scanning open-source models for vulnerabilities or malicious code, implementing security testing to ensure robustness from adversarial attacks, monitoring model inputs and outputs for sensitive data exfiltration or adversarial prompts as well as firewall or proxy type techniques to block unwanted or malicious activity. Also, AI/ML model integrity capabilities (e.g., AI Posture Management / Bill of Materials) are emerging to ensure that components of the AI pipeline are appropriately validated – from the data trained on, to the models and AI platforms used.

As AI is increasingly embedded in software and applications, this integrity verification will be important to consuming AI through third party software. Ultimately, security policies set by organizations based on their risk appetite and model use cases will help to manage the AI risk posture and monitor model behavior. Increasingly, platform providers are introducing guardrail type approaches which aim to prevent model risks from occurring, according to a pre-defined policy.

Market and Industry Perspectives

- New security practices for AI are quickly developing and players are emerging rapidly in this space. Emergent AI security solutions are mapping their defensive capabilities to frameworks like MITRE ATLAS (Adversarial Threat Landscape for Artificial Intelligence Systems), which focuses on identifying threat vectors and correlates methods of mitigation.

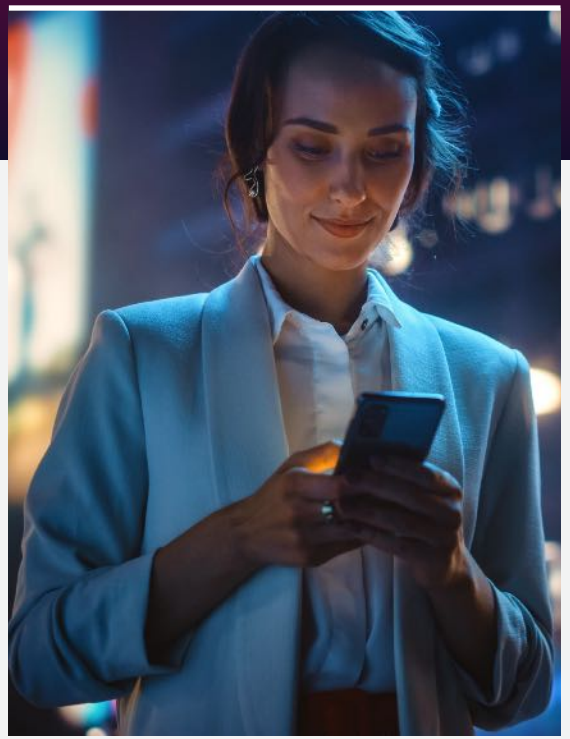
- To date, most emerging companies in this space have a specific product focus across the many new categories of AI security. Over time, we expect companies to consolidate these point features into an aggregated platform model. AI cybersecurity companies are also working through various architectural patterns to seamlessly integrate into multiple AI environments while minimizing latency and maintaining performance of outputs. To this end, we are also closely monitoring the introduction of AI security capabilities from platform providers that integrate natively within their AI services but are unlikely to work across AI platforms.

Automated Security Validation

Growing complexity in technology environments and evolution in threat actor tactics drives constant change in an enterprise's attack surface – making it challenging to effectively manage security exposure. Emergent automated security control assessments and continuous pen-testing techniques aim to provide validation and assurance across an organization's end-to-end security control environment by continuously testing chained threat scenarios to identify control weaknesses or gaps.

This approach aims to improve security posture by verifying that controls operate effectively in concert in a layered combination vs. traditional approaches, which test point control effectiveness or verify that control coverage exists. Tools in this emerging category leverage the latest threat intelligence and adversarial tactics to continuously find, fix and verify the exploitable attack surface by revealing proven attack paths, explaining how weaknesses lead to specific impact and prioritizing and detailing the fixes that should be made immediately with consideration for the compensating controls that also exist.

By operating continuously, these tools also verify that fixes made in the environment effectively mitigate the risk. By validating controls end-to-end, an organization can get a better picture of priority areas for remediation and performance against real world tactics and techniques.



Market and Industry Perspectives

- The market in this area has developed around three primary approaches. First are pure play solutions that focus primarily on the continuous threat simulation, agnostic to the underlying security controls.

- Second is the marrying of active testing with existing posture management techniques, for example cloud posture tools which feature a dynamic scanner to validate if perceived weaknesses are exploitable by known threat techniques.

- The final approach is a broader exposure management platform which is being pursued by larger security vendors, which intends to connect their broad view of enterprise risk posture to better pinpoint and recommend the highest priority risk areas of focus. As a result, security platform vendors have begun acquiring products in these categories to complement this platform focus. In the end state, this approach will likely be most impactful, but these vendors' ability to operate agnostically with controls and data provided by other vendors could be a challenge.

- Automation through simulated threat activity would improve understanding around the impact of control gaps in a layered set of controls, as well as pinpoint the optimal and most efficient risk reducing action. Ultimately, these tools help move towards an exposure and risk management approach to security vs. a compliance-based approach of remediating lists of findings which can also make risk reduction efforts more efficient and explainable. Additionally, there are varied approaches forming in the market and additional research will be required to decide which, if any, is most appropriate to pursue.

JPMorganChase

Global Technology Strategy, Innovation and Partnerships