AI SPRINT
AI IN SECURE PRIVACY-PRESERVING COMPUTING CONTINUUM

EUCloudEdgeIoT.eu

Computing Continuum and the role of AI as a complementary technology:

European market forecast and insights

Webinar report

May 24th saw the webinar "Computing Continuum's organisation and AI's role as a complementary technology: European market forecast and insights", co-organised by EU-funded projects AI-SPRINT and UNLOCK- CEI, representing the EUCloudEdgeIoT.eu initiative. The event aimed to provide valuable insights to all value chain actors involved in the Edge, Cloud, IoT, AI, and Data communities, including research organisations, associations linked to computing continuum technologies, developers and domain experts, governments and policymakers, citizens, and civil society. The event also aimed to allow attendees to better understand the demand market of edge computing and get insights into the spending, adoption level, and top use cases. Furthermore, the webinar provided insights into market trends that could impact investment plans for adopting Edge and AI, focusing on the healthcare, energy, and agriculture sectors and complementary technologies such as ML and federated learning.

DOI:
10.5281/zenodo.8046555

URL:
doi.org/10.5281/zenodo.8046555

Authors:
Claudio de Majo
Maria Giuffrida
Rita Giuffrida
(Trust-IT Services)

Editor:
Andrea Greco
(Trust-IT Services)

AI SPRINT
AI IN SECURE PRIVACY-PRESERVING COMPUTING CONTINUUM

EUCloudEdgeIoT.eu

Golboo Pourabdollahian (IDC) introduced the UNLCOK-CEI project, one of the two CSAs powering the EUCloudEdgeIoT community, focused on the demand side of cloud-edge-IoT technologies. In contrast, the other CSA project - OpenContinuum - deals with the supply side. Together, these projects aim to realise a pathway for the understanding and development of the Cloud, Edge and IoT Continuum, promoting cooperation between a wide range of research projects, developers, suppliers, business users and other potential technology adopters. Furthermore, both projects support a large community of Research and Innovation Actions (RIAs) projects developing various innovative infrastructures in the computing continuum through different interrelated initiatives. Such aims are accomplished through the interaction of different task forces clustering different members of project consortia and other open initiatives.

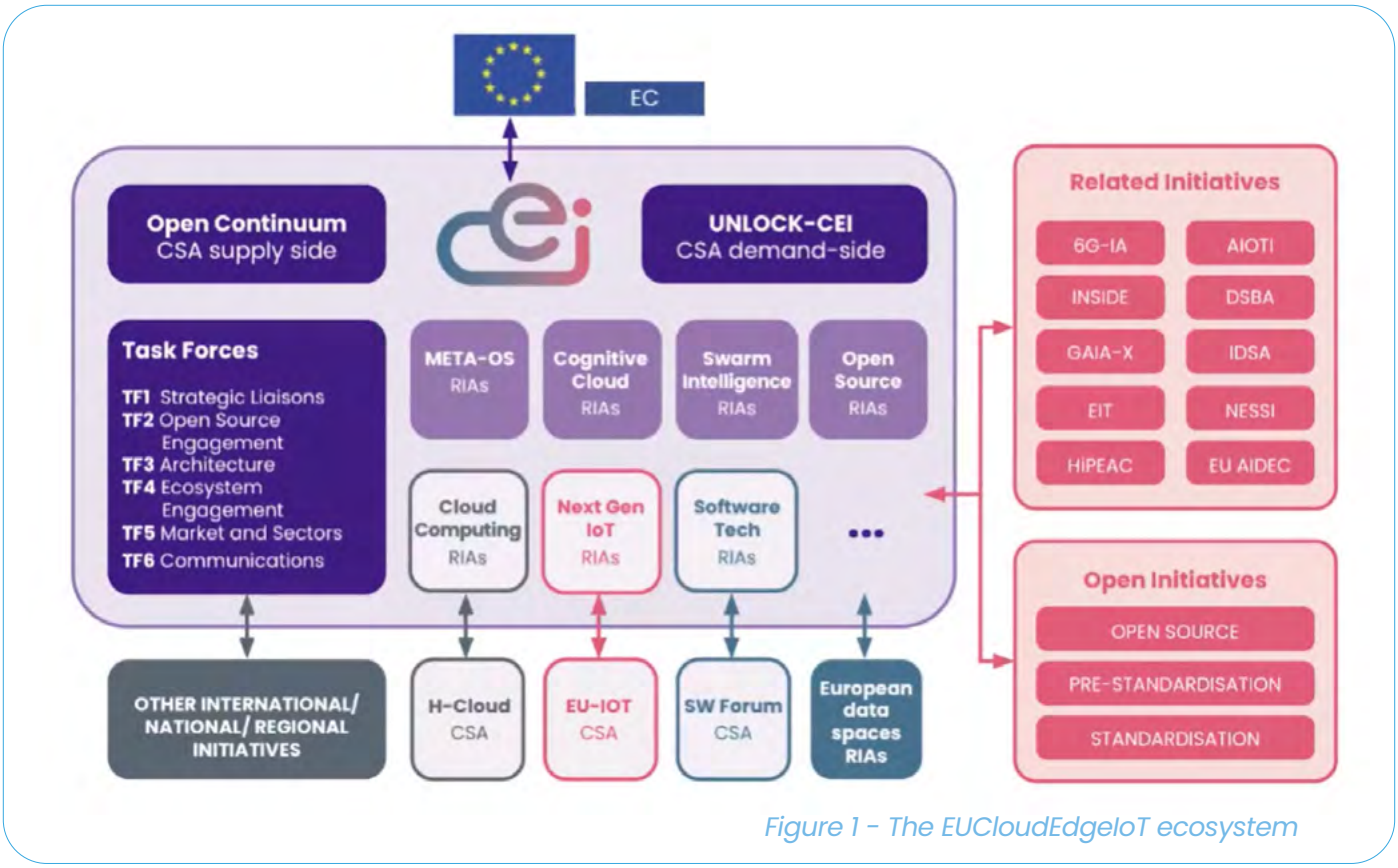


Figure 1 - The EUCloudEdgeIoT ecosystem

Join the EUCloudEdgeIoT Taks Forces

More specifically, UNLOCK-CEI aims to investigate the demand landscape of Cloud-Edge-IoT in Europe in terms of spending, trends, forecasts, value chain configuration and service requirements to define eventually the business opportunities for European players. According to IDC's forecast, the total Edge spending in the EU will reach 75 billion dollars by 2026. However, this paradigm shift is difficult to accomplish despite the massive investment, as industrial sectors still struggle to fill the demand-supply gap. This requires better understanding the demand landscape and generating insights on main drivers, challenges, business opportunities and market scenarios that could orient suppliers.

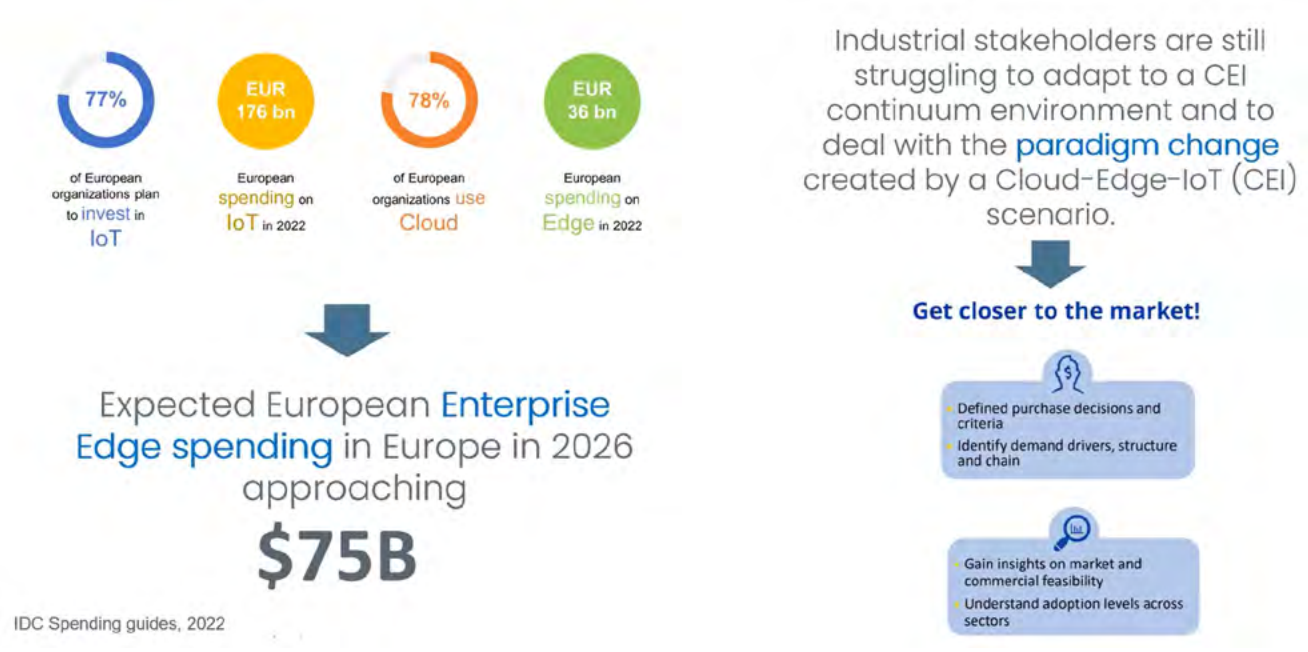


Figure 2 - Cloud-Edge-IoT market key figures

Hence, UNLOCK-CEI's main ambition is to facilitate and accelerate the deployment of the Cloud-to-Edge-IoT computing continuum in Europe by promoting knowledge about these technologies. The project aims to accomplish this by focusing on the demand-side drivers and challenges to identify technology-driven innovation and business opportunities driving demand value chains. This agenda can be resumed into five main objectives: 1) CEI landscape assessment; 2) Market scenarios definition; 3) CEI Industry Constituency building; 4) Coordination and interaction with the supply side (EU-funded RIA projects); 5) Awareness and Impact generation.

The project started about one year ago, interacting with the demand and supply side through different approaches. First, this was accomplished through a survey of the CEI demand landscape with over 700 interviews with European industrial players across five verticals. Other actions included deep dive interviews with value chain players, five online value chain adopters workshops with sector-specific actors, supply-side portfolio mapping, use-case harmonisation and webinars disseminating project results and market insights.

Discover UNLOCK-CEI preliminary results

Danilo Ardagna (Politecnico di Milano) introduced the AI-SPRINT project, whose main aim is to develop AI-enabled innovative applications through edge-enabled mechanisms. While the worldwide market for Cloud, Edge and IoT solutions is expanding, new challenges are posed by the heterogeneity of solutions that can impair their performance.

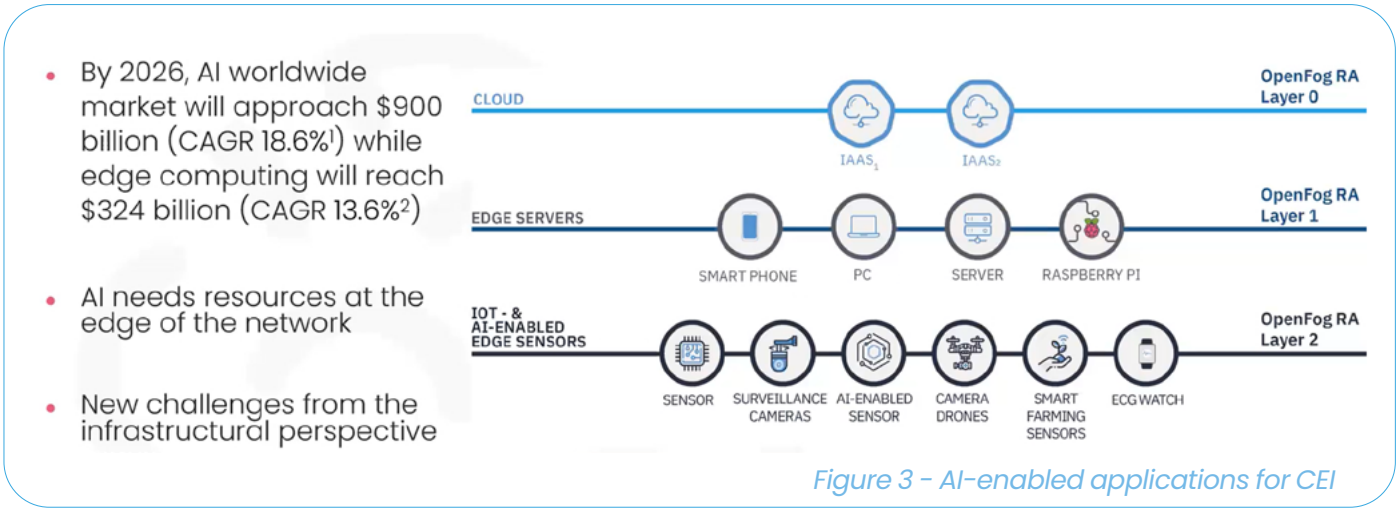


Figure 3 - AI-enabled applications for CEI

AI-SPRINT aims to design simplified programming models for AI designers and edge administrators, where application programmers can automatically deploy their codes and automatically fine-tune them in terms of performance and accuracy, leveraging this machinery model to identify costs. Other applications include developing tools that allow running applications through continuous deployment and tee-secure boot & patch management, ensuring privacy preservation, application reconfiguration and scheduling for accelerator devices. Finally, the project also aims at developing highly specialised building blocks, such as federated learning for distributed training and privacy-enabled applications.

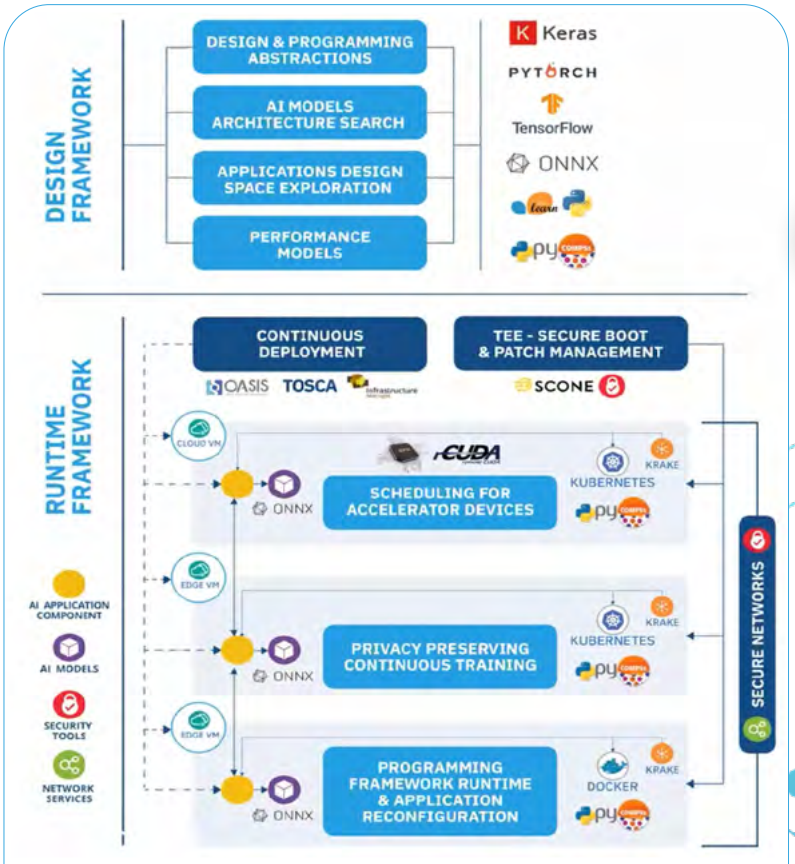


Figure 4 - AI-SPRINT Framework

The project is already in its final stages and currently looking for external users to test its applications through the Alliance programme.

Join the AI-SPRINT Alliance

John Gole (IDC) provided insights into the European demand market of CEI technologies. First, he highlighted that it is impossible to disentangle these technologies from spending statistics. Indeed, although these technologies have experienced different trends, they are also closely related. Clearly, cloud technologies have been around the longest, although they are still not completely mature, as witnessed by growing spending (so far 110 bn). IoT is also in a mid-maturity stage, although its total spending is higher (176 bn) given the many potential applications and devices this technology enables (sensors, software applications, middleware AI, etc.). Finally, Edge is a sort of newcomer, with 36 billion euros invested so far, although investments are also growing steadily.

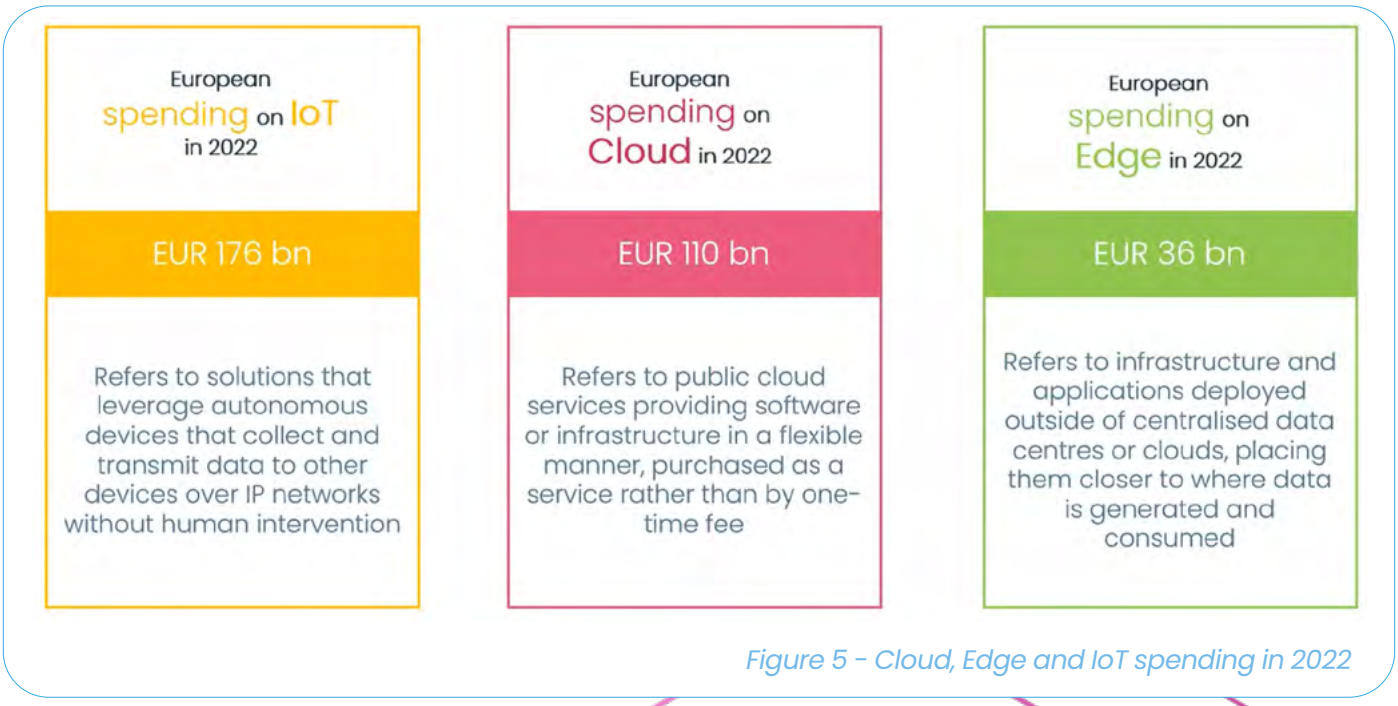


Figure 5 - Cloud, Edge and IoT spending in 2022

In the context of UNLOCK-CEI, a survey was conducted to understand the technologies' state of adoption. Cloud is the leading technology already being used by 70 per cent of total companies (although only 36 per cent is using it extensively). However, the survey demonstrates that many organisations plan to invest in all three technologies relatively homogeneously.

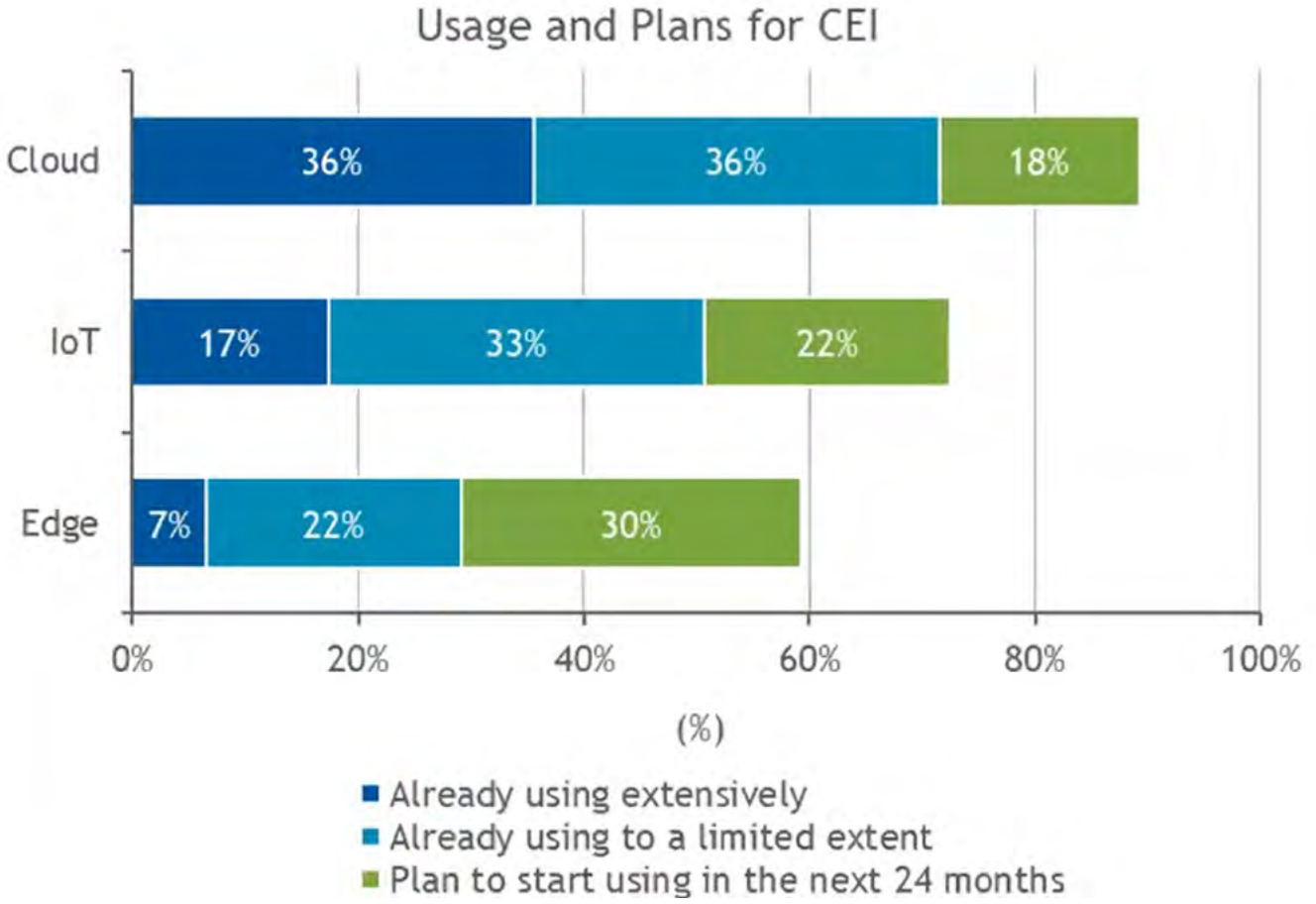


Figure 6 - Cloud, Edge and IoT usage and plans

More in detail into IoT, three primary industries, agriculture, energy and healthcare, have been considered. Healthcare is currently the leading industry, with 34 per cent already adopting it, followed by energy with 18 per cent, while agriculture significantly lags with only three per cent. However, looking at prospective adoption, all three sectors seem to even out, as all companies are willing to invest in solutions enabled by these technologies.

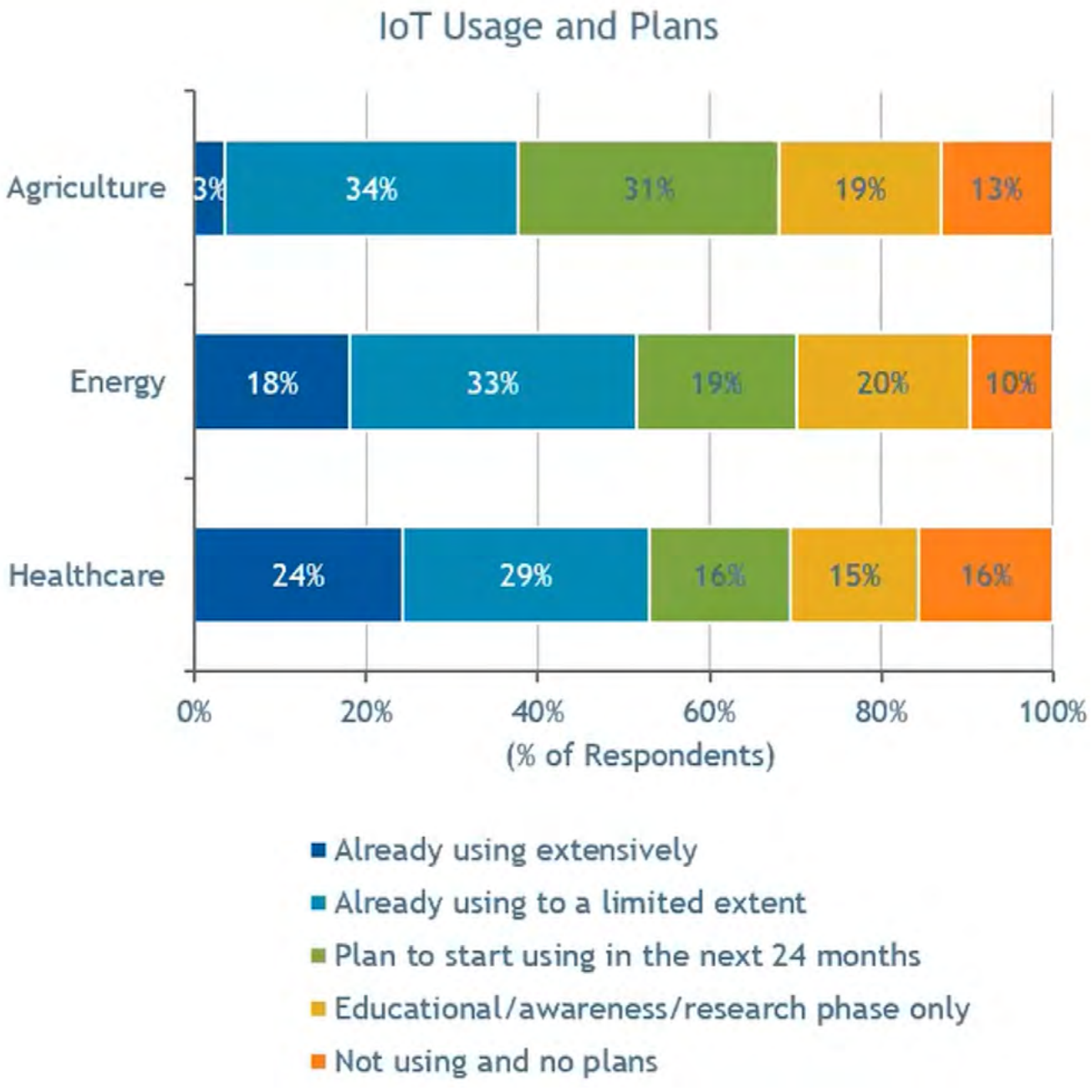


Figure 7 - IoT usage and plans in selected industries

What changes between sectors is the scale of adoption, where there seems to be room for significant improvement, considering that most industrial sectors still deploy these technologies at a small scale (60 per cent in the case of agriculture). In terms of large-scale adoption, healthcare leads with 47 per cent of industries already investing at a large scale, while energy and agriculture score 30 and 11 per cent, respectively.

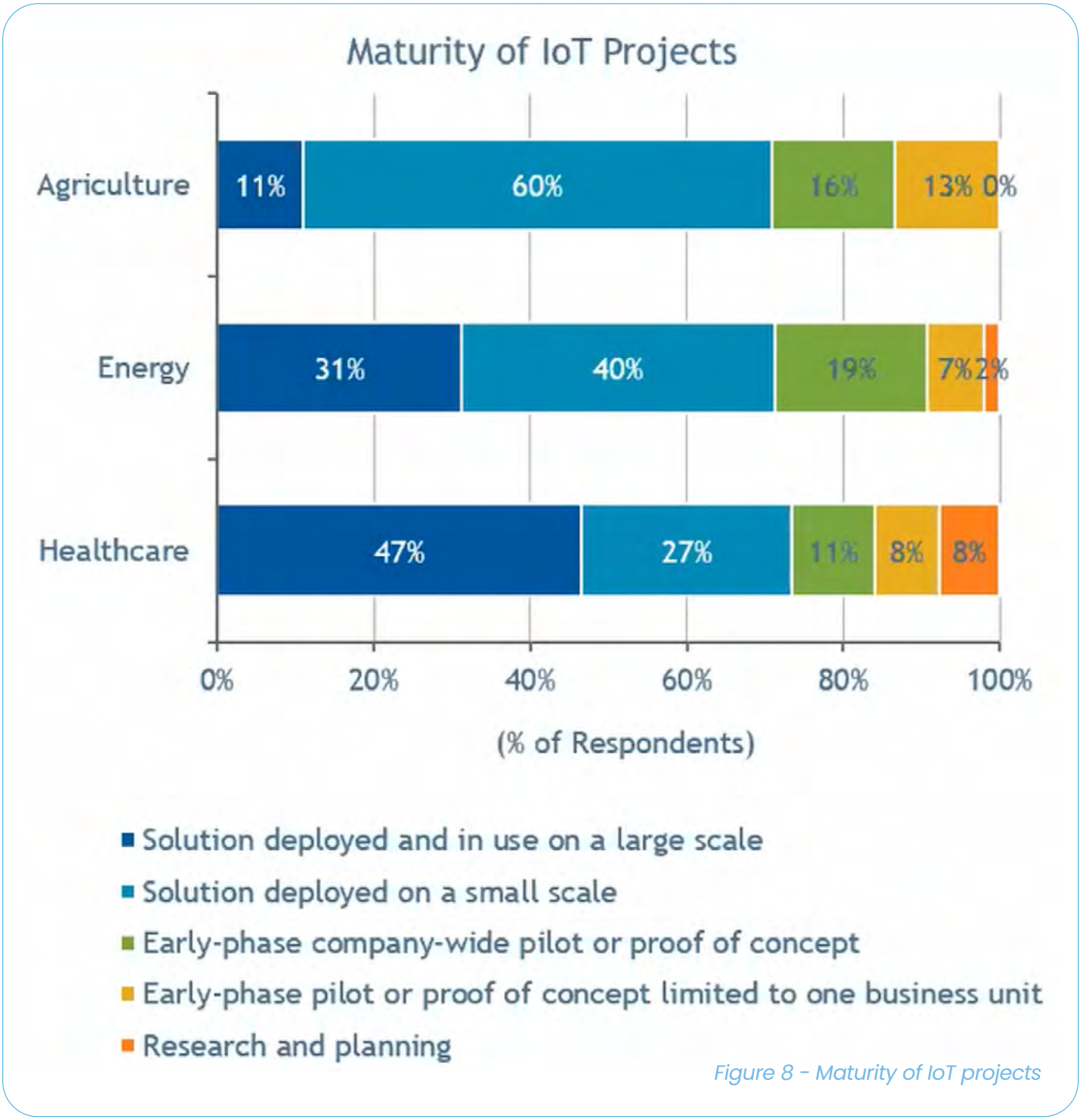


Figure 8 - Maturity of IoT projects

In terms of use cases, the most common use case for energy, agriculture and manufacturing is employee safety monitoring, followed by other trending use cases in asset location tracking, video surveillance and virtual inspection. Perhaps more interestingly, actual usage and planned usage even over time across sectors, demonstrating the will to invest in these technologies. In several widespread use cases, such as employee safety monitoring, these applications will most likely be edge-enabled, given the quick data analysis and response rates needed by such use cases. Other use cases requiring edge technologies or at least a mix of Cloud and edge computing applications include remote network management, process automation and optimisation, video security and surveillance, automated vehicles and AI-enabled diagnosis and treatment.

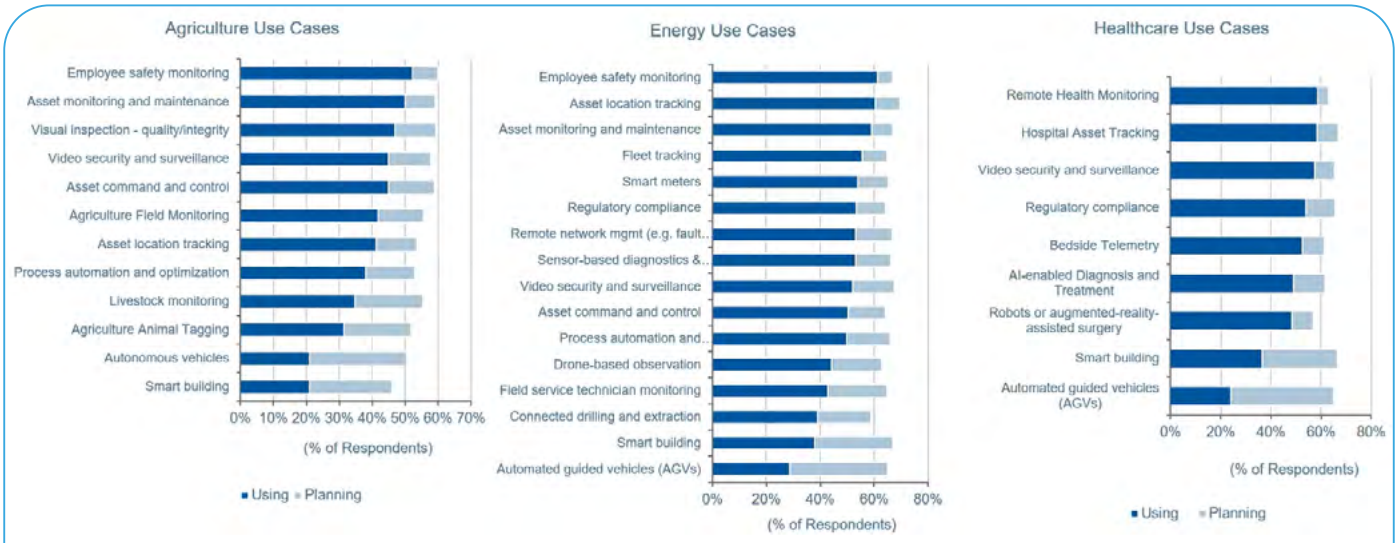


Figure 9 - Main use cases

In terms of edge utilisation, organisations seem more prone to adopt these technologies when they do not want the application to escape their control, do not rely on network connectivity, cut cloud costs, require ultra-low latency, more reliability, AI automation or aggregate IoT sensors. In terms of edge usage plans, healthcare is once again the main sector showing the capability to adopt these technologies due to its concentrated facilities and skilled workforce able to handle them locally. However, even open-field agriculture is supposed to adopt an increasing amount of edge technologies for precision applications and machinery automation. The will to embrace the telco network edge is instead a common trend emerging across all the surveyed industrial sectors.

Despite perspective spending seeming to be at a high level in these fields, several challenges remain hindering the deployment of these technologies. The main common denominators are deployment complexity and low market maturity due to lack of refinement and easy repeatability packaging of use cases, increasing costs, skills demand and security risks. However, there are also other common factors, such as overall deployment costs, security concerns, lack of skills, lack of interoperability mechanisms, and many others.

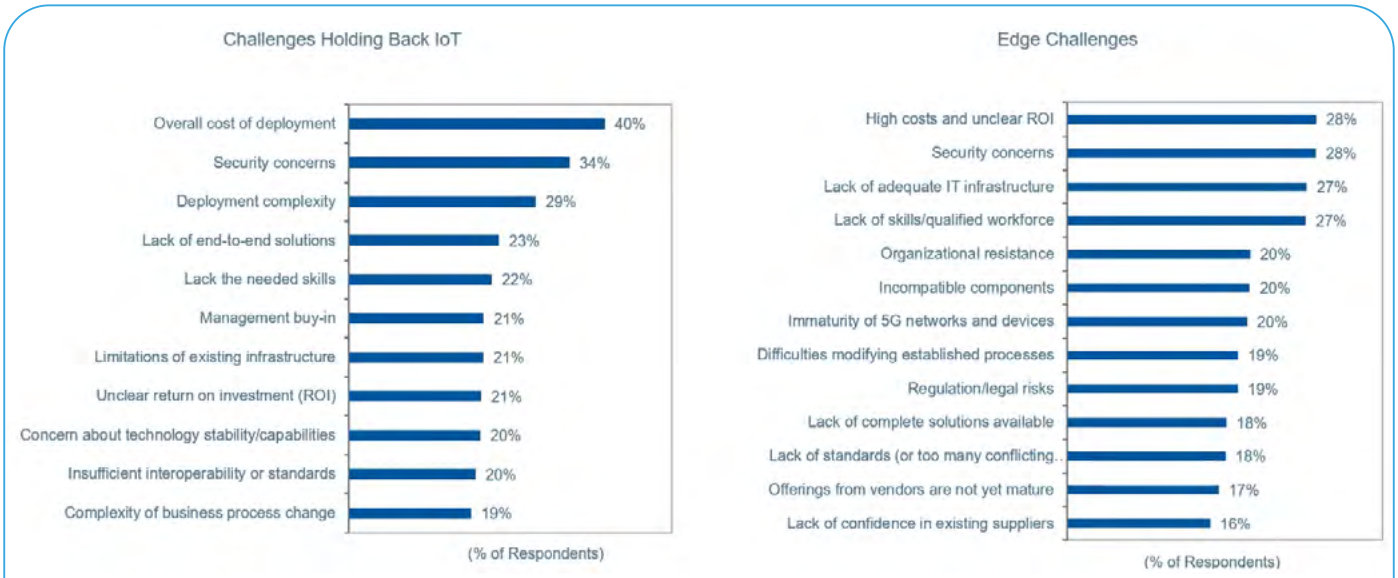


Figure 10 - IoT and Edge Adoption Challenges

Explore the Cloud-Edge-IoT market trends

Ewa Zborowska and **Jack Vernon** (IDC) provided a closer look and market analysis for AI on the computing continuum.

Zborowska started discussing the familiarity rate of CEI technologies, with cloud applications leading the polls, with 77 per cent of firms declaring it as a standard technology and already using it to a large extent (almost 80 per cent). However, despite the familiarity, Cloud is part of a large application of these companies' IT environments, which are increasingly looking for a hybrid and balanced approach to computing continuum applications to find the best application for each use case.

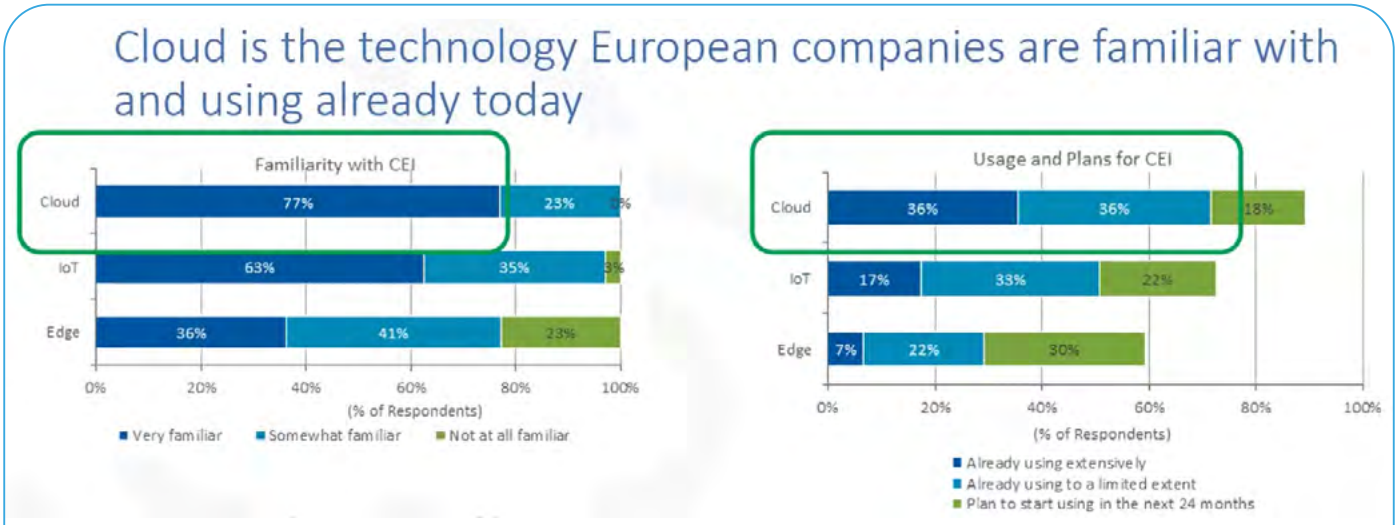


Figure 11 - Familiarity with Cloud Technology

Hence, the reason why in general, cloud adoption is quite widespread across different industrial sectors, although their adoption still heavily depends on resources owned and business demands. Specifically, the larger the companies, the more resources to build cloud-deployment competencies and capabilities.

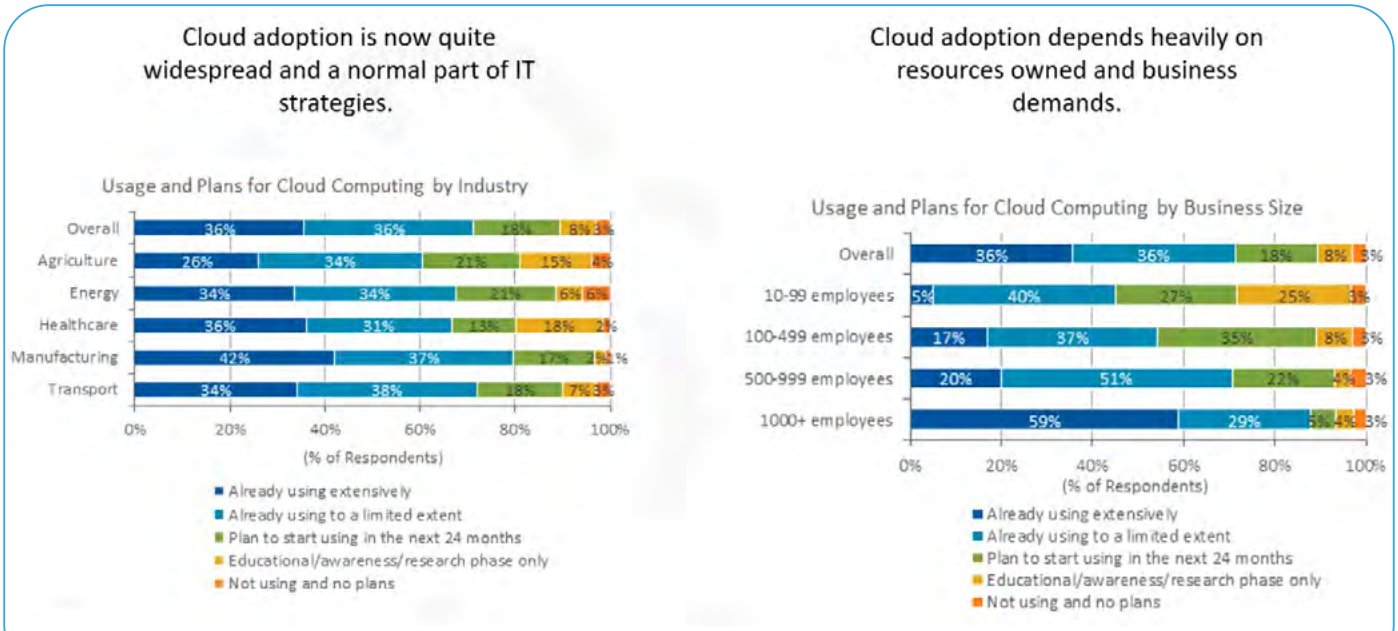


Figure 12 - Cloud Usage and Plans by industry and business size

However, not all companies have successfully leveraged cloud technologies, with only 43% of the companies analysed declaring their experience as successful. This is explained because, despite cloud application maturity, many companies tend to miss some essential adoption requirements. Specifically, they cannot set specific goals or plan activities properly, lack expertise, and cannot build an ecosystem. Additionally, they also often face integration challenges they cannot fulfil.

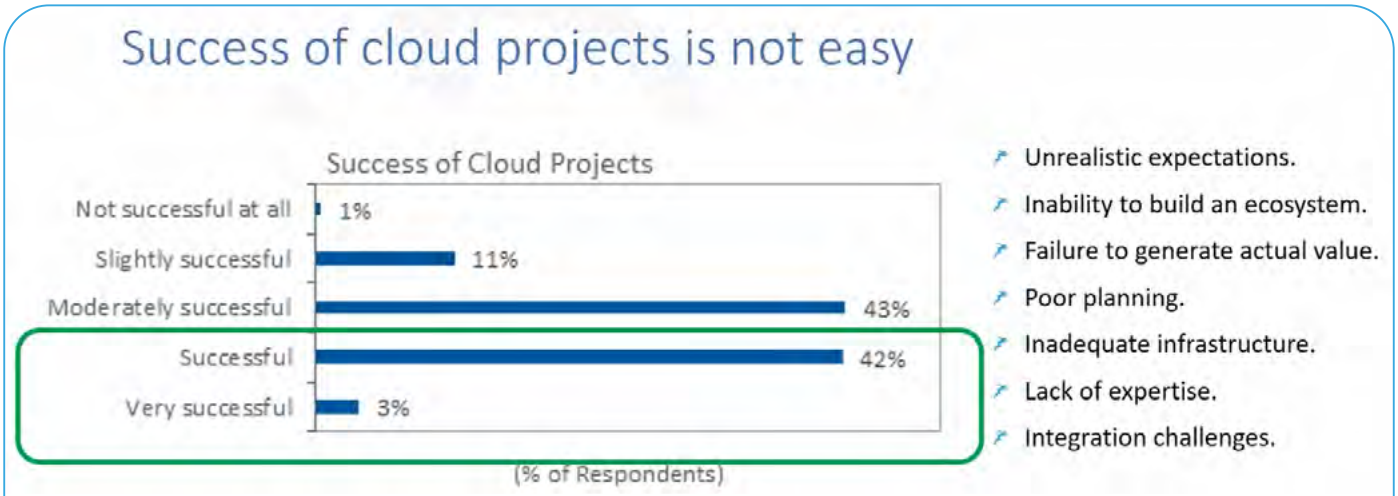


Figure 13- Success rate of Cloud projects

Cloud is nonetheless becoming an enabler of new IoT-based use cases. These include data storage and processing to generate vast amounts of data that are difficult to store and manage directly on-premise. Second, device management allows the Cloud to provide a centralised location for managing IoT devices, allowing for easier monitoring and control mechanisms. Third, cloud computing allows for remote access to IoT devices located in hard-to-reach locations, thus enabling businesses to perform these tasks from anywhere with an internet connection. Finally, cloud platforms offer powerful analytics and machine learning capabilities which can be used to extract insights from the data generated by IoT devices.

Cloud is also becoming the main platform for AI deployment. In the sole five-year span from 2021 and 2026, it is expected that more than 70 per cent of AI platforms will be delivered in service cloud models, with a total spending of over 36 billion dollars by 2026. Specifically, while on-premise deployment is estimated to grow at a Compound Annual Growth Rate (CAGR) of 17 per cent (from $6.6 bln in 2021 to 14.6bin 2026), public cloud deployment will increase by over 44 per cent, jumping from $5.9bln to 37.4.

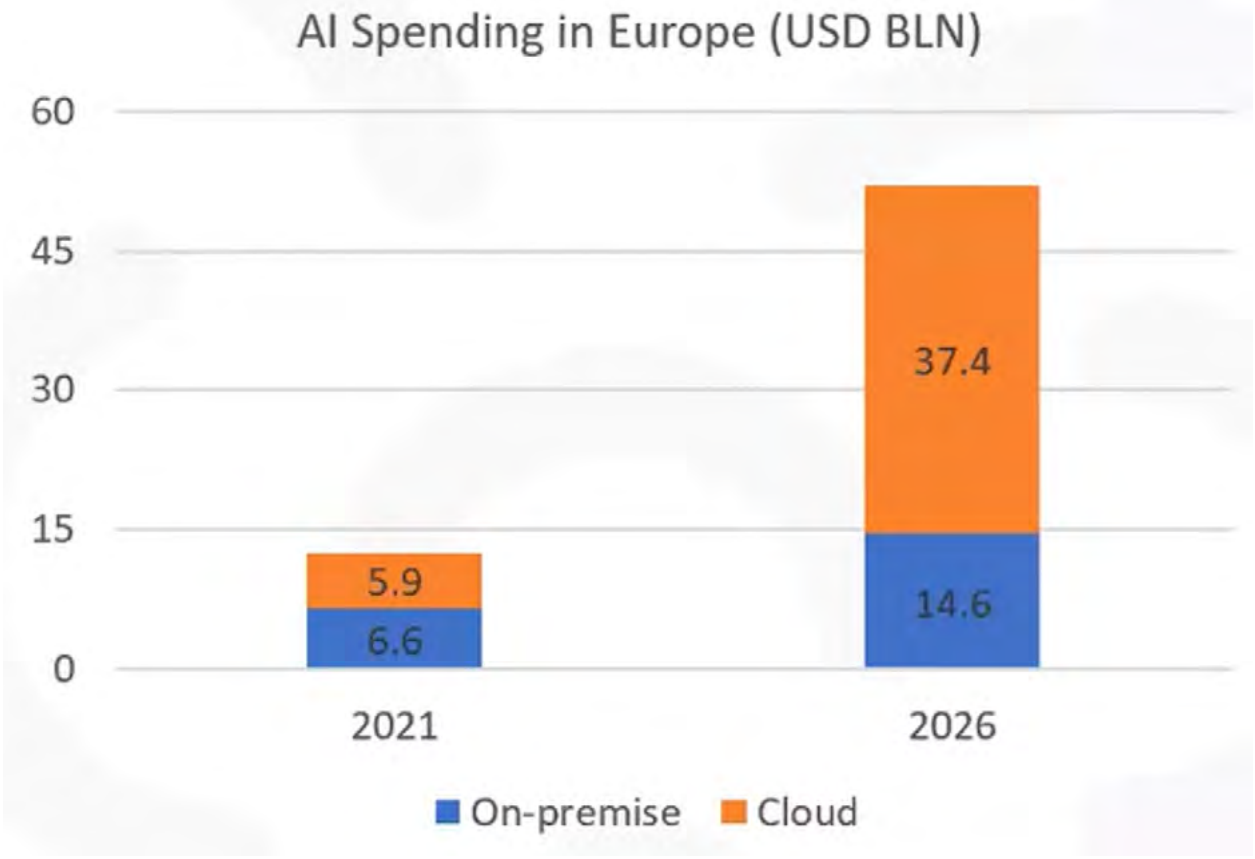


Figure 14 - AI spending in Europe

Such massive AI leverage will generate new business models, allow data monetisation, enable process automation to optimise processes and ensure a shorter time to market in product development. Other essential aspects are also the possibility to propel the construction of collaborative ecosystems with common values for customers and guarantee human capital growth, ensuring companies' growth despite potential ethical issues such as data management and lack of trust.

Although companies tend to differ from each other, AI will enable the growth of more use-cases-focused customers, looking for specific solutions that allow companies to deliver predefined use cases and create business models built from internal and specialised datasets in use cases-related particular domains.

Not surprisingly, the projected most popular cloud-based AI use cases include:

1. IT optimisation;
2. Quality management investigation and recommendation systems;
3. Diagnosis and treatment and government intelligence systems (e.g. defence, terrorism, investigation);
4. Expert shopping advisor and product recommendations.

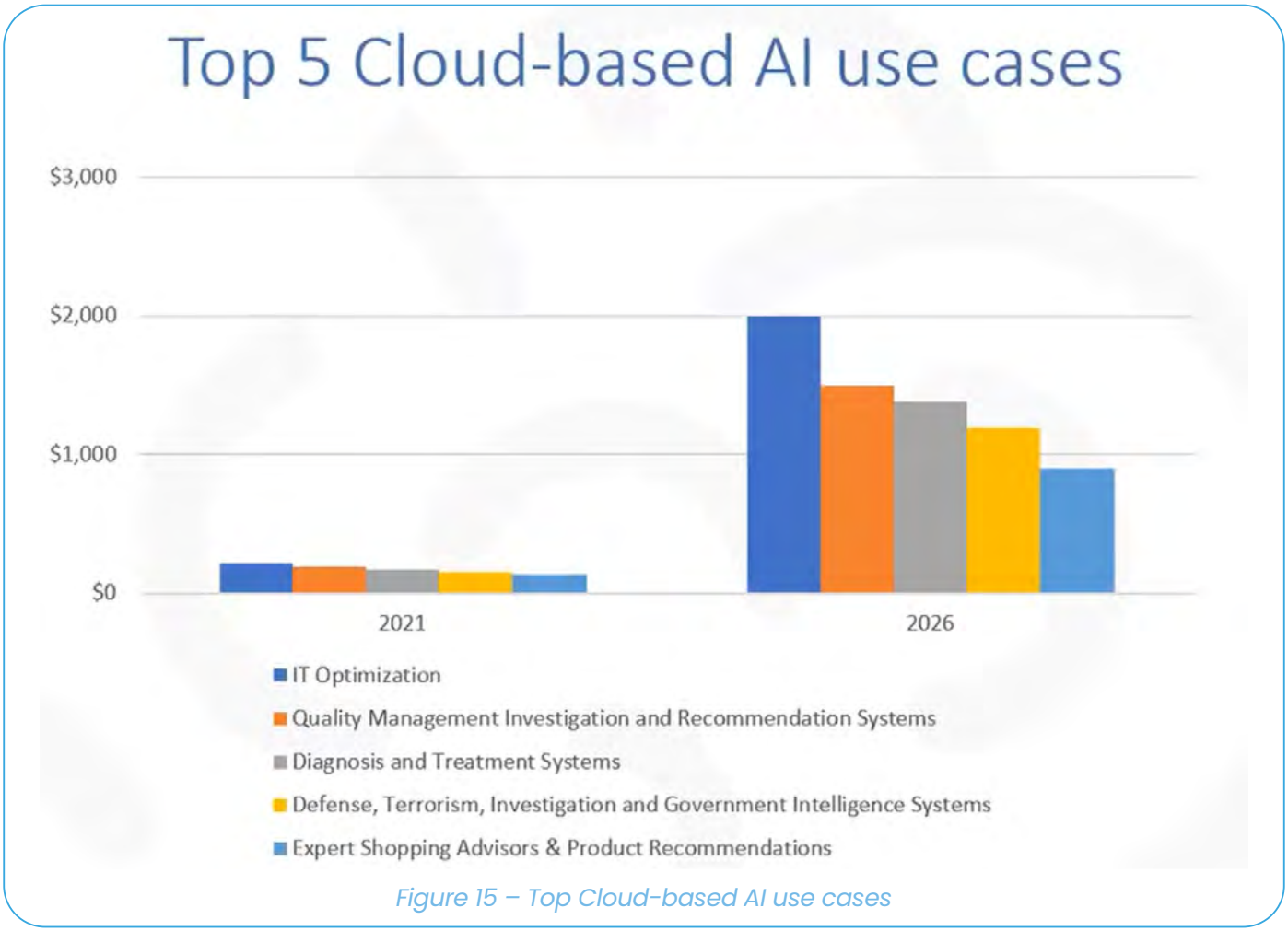


Figure 15 – Top Cloud-based AI use cases

Vernon deepened **Zborowska**'s insights into AI-cloud relationship management. Specifically, he touched upon AutoML's definition and evolution to prompt market advancement. These are particularly interesting because they have expanded their capabilities beyond machine learning to domains such as data processing, MLOps, trustworthy AI, data visualisation, management, and governance. Moreover, these capabilities are often offered as part of their standalone offering.

What drives demand for these AI capabilities is the will to build custom AI machine-learning modules (34 per cent of cases) and the passive consumption of AI-based features embedded in existing standalone software tools (over 56 per cent of cases).

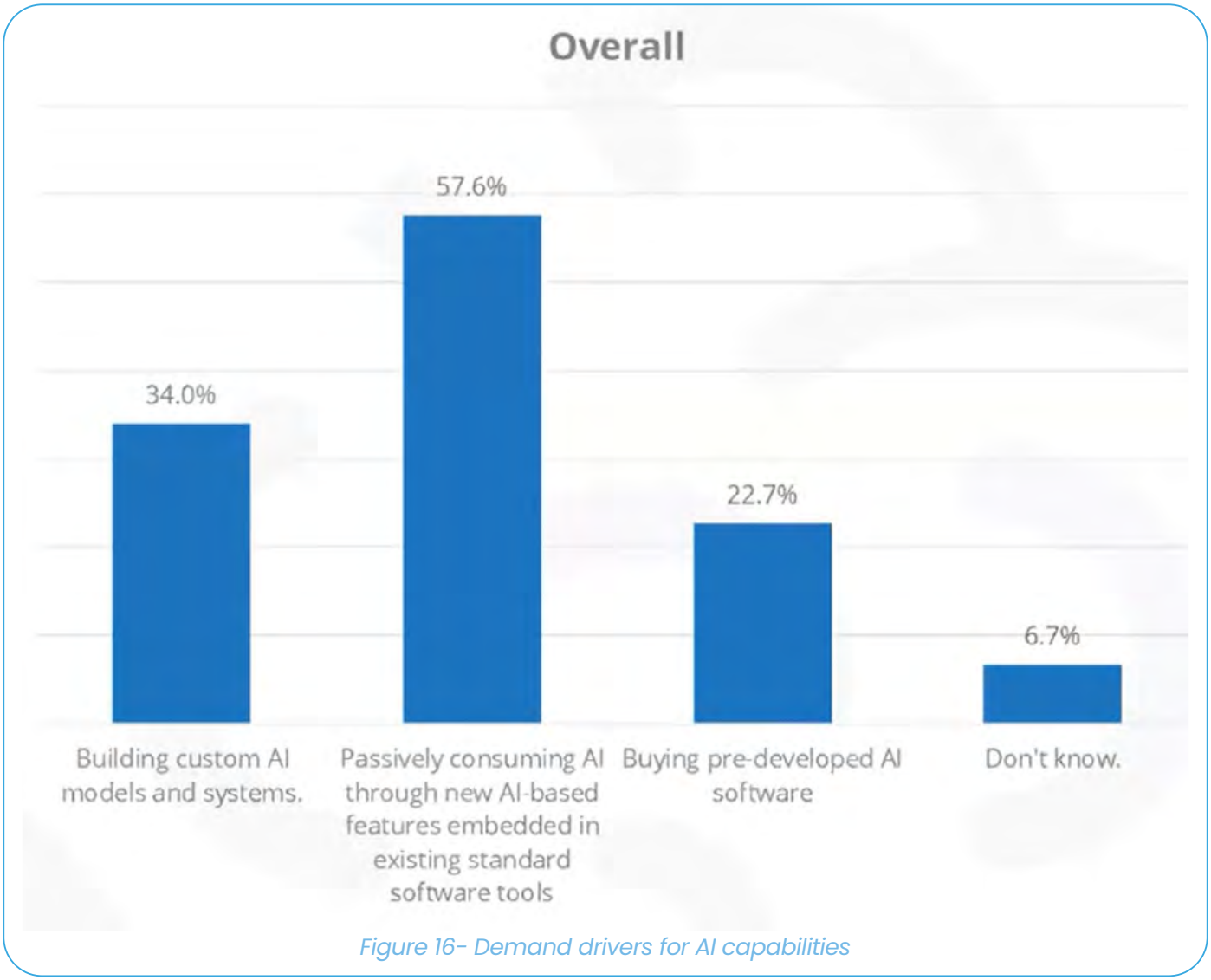


Figure 16- Demand drivers for AI capabilities

In terms of the main industrial actors driving the services, these are mainly infrastructure, manufacturing, distribution and service, as well as public sector and financial services

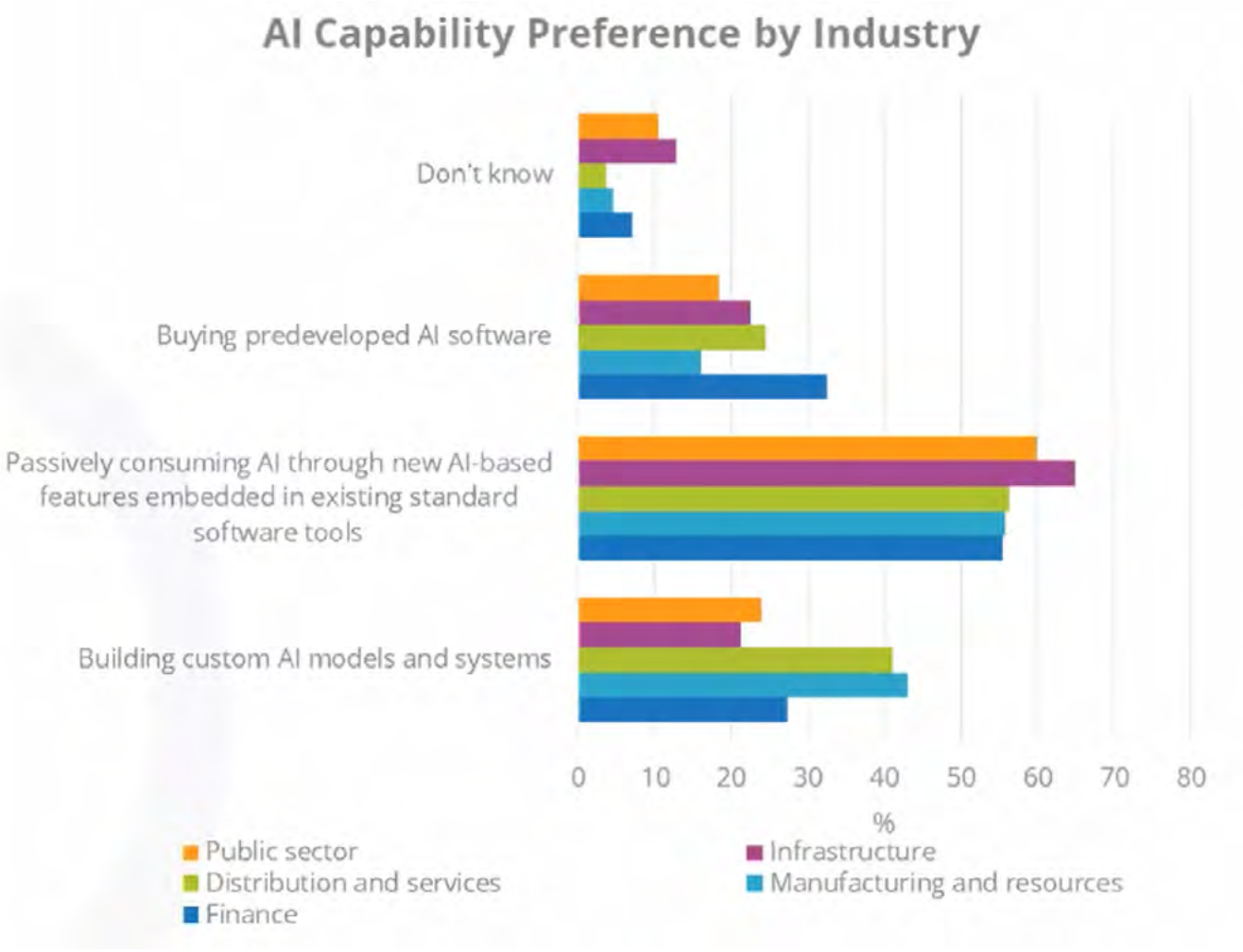


Figure 17 - AI capability preference by industry

In life-cycle software markets, significant growth is projected in public and on-premise Cloud, generating revenues of over 22 billion dollars by 2026, with particularly substantial indexes in the Americas.

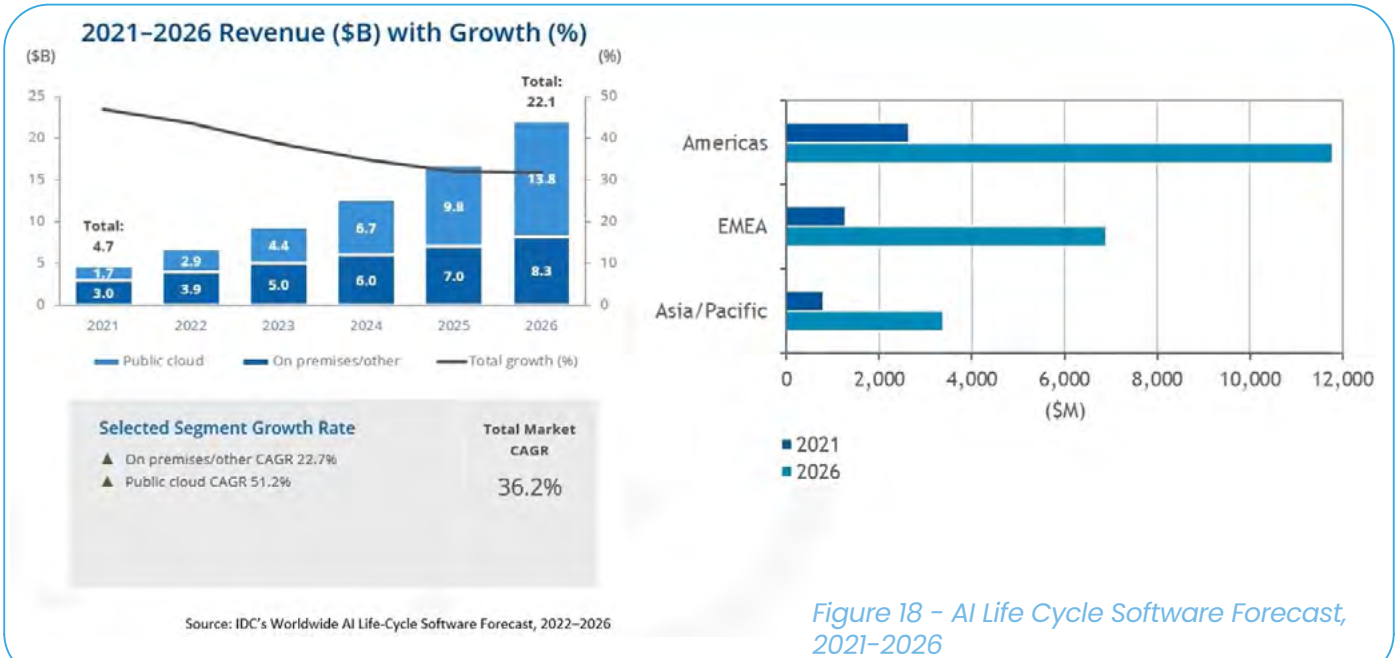


Figure 18 - AI Life Cycle Software Forecast, 2021-2026

The development of AI is also pushing the selling of life-cycle tools, both spreading in coded and lo/no code approaches, through a mix of commercial and open-source technology. In this context, startups remain prominent, especially in the low-code market segment (e.g. Dataiku, Peak.ai, Monolith). In addition, cloud vendors and providers have invested heavily in this space and are represented across multiple categories, encouraging cloud uptake by charging extremely low rates to users, as well as open source tools associated with established data science tooling.

In this context, technology opportunities include industry specialisation, transformer models, and optimised backend performance configuration of compute to drive runtime performance. There are also competitive challenges, such as figuring out alternative approaches related to the direct purchase of models from platforms and transformers leveraging optimal pieces of code (generative coding tools) instead of dedicated AI Life-cycle systems to cut costs.

Another interesting technology is federated learning which allows AI model training to take place collaboratively across a network of nodes, with all data remaining at each node and not needing to be exchanged. For example, a node could be a mobile device, a server system, a data centre, a hospital, an organisation or even a car.

Only updated learnings to each node are shared. Once shared, the update is immediately aggregated with updates from other nodes to improve the shared model and redistribute back out across the network of nodes. All the training data remains at the original endpoint node device.

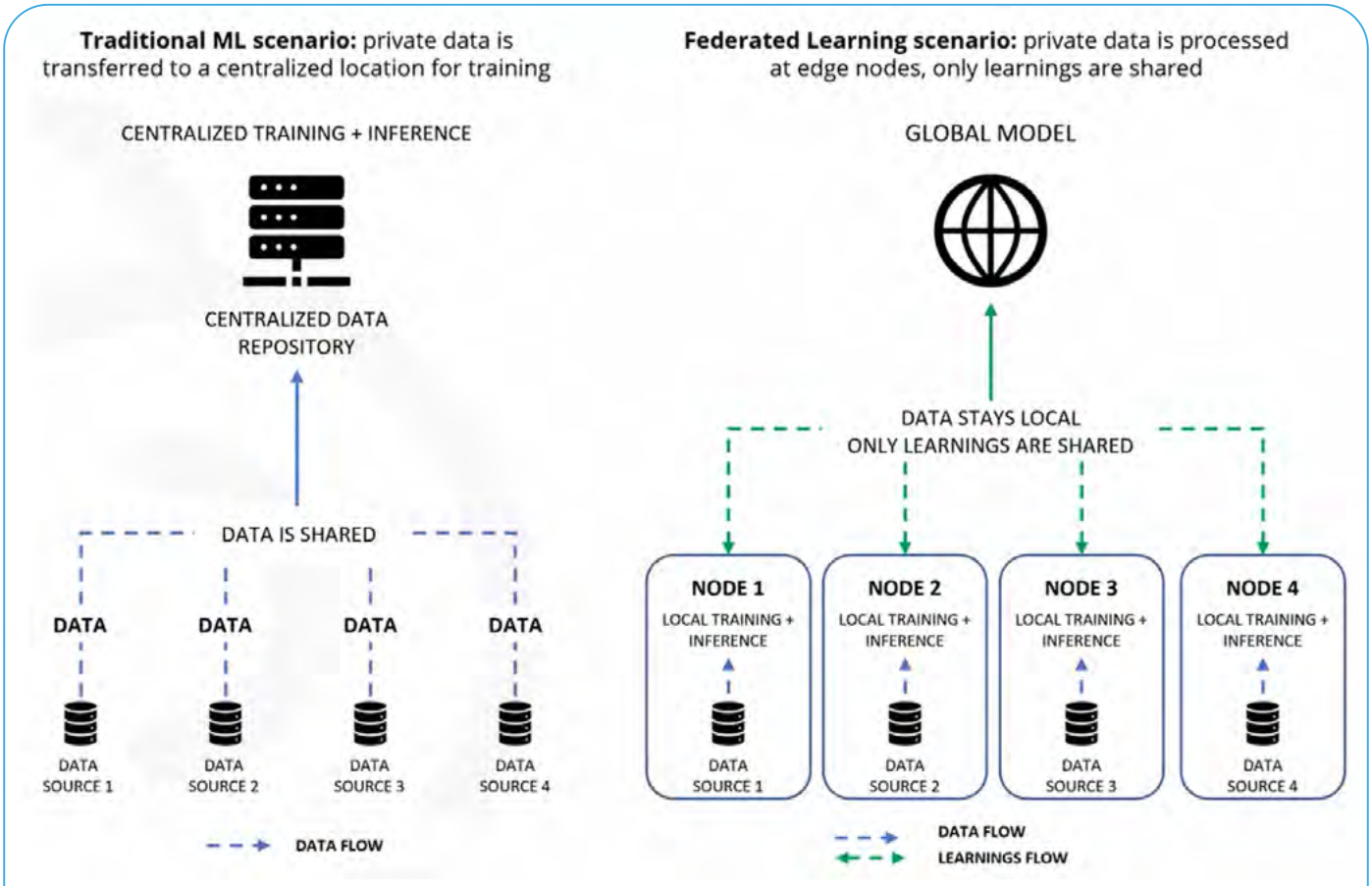


Figure 19 - Traditional Machine Learning vs Federated Learning

Thus, compared to a centralised data repository, with federated learning scenarios, data processing happens at a decentralised level, allowing sharing aggregates updates to the whole module, thus reducing the amount of information that needs to be shared across the network and keeping all the data at the edge environment. Therefore, the main benefits of federated learning include the possibilities for multiple parties to collaborate without sharing data, better security as the attack surface is reduced, less energy consumption and greater operational efficiency and flexibility.

One example where this technology has already been used in the market is the Google Keyboard prediction tool, a language model based on federated learning, where all personalised use information is not sent back to aggregate data centres but updated locally.

Different federated learning deployment paradigms include heterogeneous data mechanisms allowing two organisations to collaborate in developing machine-learning modules without sharing data. In particular, vertical federated learning allows one to work with the same users. In contrast, transfer learning allows one to analyse a small sample of common users and identify data samples for surveys and other operations.

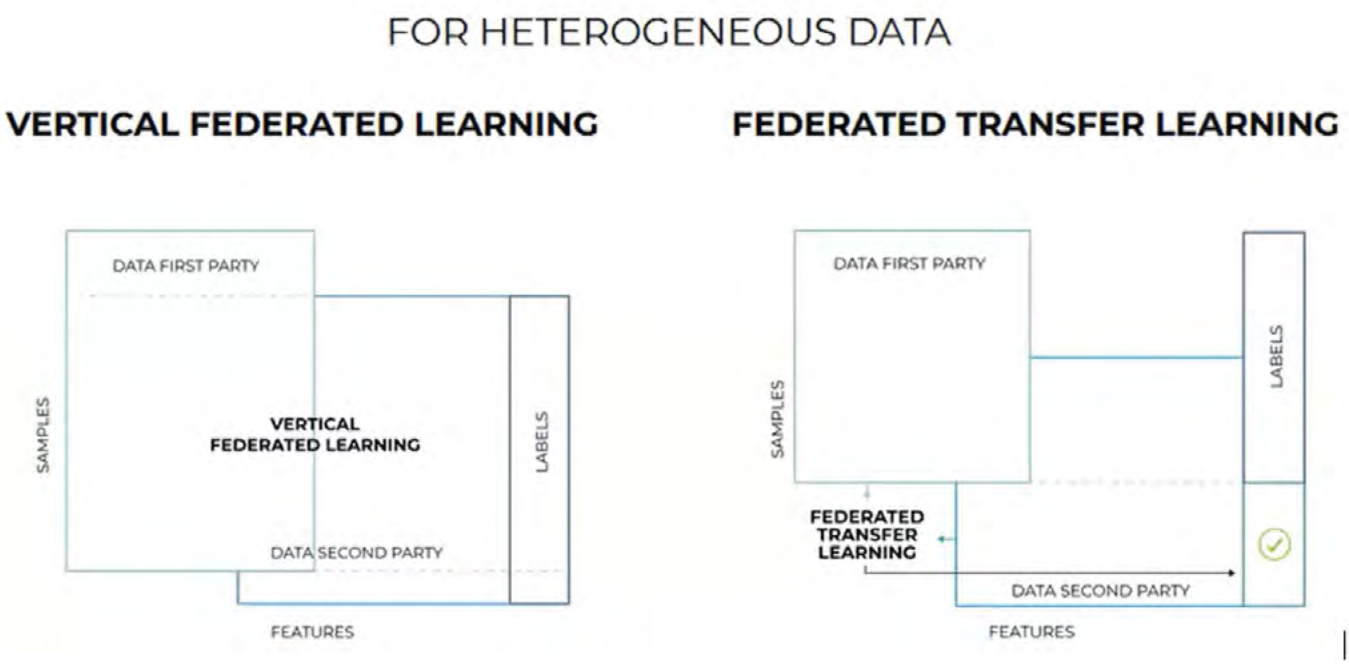


Figure 20 - Federated learning for heterogeneous data

In terms of vendors, there is a mix of open source and fully commercial technology efforts leading the charge in developing Federated Learning Platforms and a market engagement where startup and cloud vendors are prominent. Regarding future technological opportunities, federated learning offers greater regulations and the possibility of creating an open market for multi-party data collaboration and exchange. However, in terms of competitive challenges, scaled commercial deployments of the technology outside large tech companies remain limited as federated learning still needs to show its viability for a broader range of organisations.

A Q&A session followed, moderated by **John Favaro** (Trust-IT).

Question #1

How can EUCloudEdgeIoT Task Forces be joined, and who can join them?

Golboo Pourabdollahian (IDC) replied that all task forces are designed for the community's Research and Innovation Action projects. Task force leaders can be emailed in case of interest. Also, because the first task force is about strategic liaisons, different organisations that could be relevant for collaboration with the EUCloudEdgeIoT initiative would be welcome to join.

Question #2

Are new sensor technologies expanding the scope of possible cloud-edge IoT applications?

Danilo Ardagna (Politecnico di Milano) replied that processing sensors already embed data processing engines or partition the network across multiple layers of the computing continuum to select the best deployment to that specific instance and ensure optimised performance.

Question #3

Is it true that introducing AI in the Computing Continuum will also require new software engineering tools and processes?

Danilo Ardagna (Politecnico di Milano) remarked that designing AI is challenging from the software engineering perspective because the networks are task-specific and new frameworks evolve quickly. This makes it difficult for new developers to cope with this complexity and devise interoperability mechanisms. Thus, new solutions to these challenging complexities are needed in the computing continuum at the technological and market levels.

Question #4

Which topics should be prioritised in the next period?

John Gole (IDC) provided the IoT case, a technology which cannot be considered as just one market but a combination of different markets and assets interacting with one another. Given the many fundamental tools composing IoT technologies, much work is still needed to homogenise these applications across the CEI domain, creating horizontal categories and vertical solution areas. For example, regarding 5G networks – a crucial enabler for CEI-based use cases – creating a unified system to leverage these applications at private and public levels is vital.

Question #5

Are Large companies going to dominate the market? How can people acquire the needed skills to overcome these challenges?

Ewa Zborowska (IDC) remarked that it might be harder for SMEs to build internal capabilities to leverage these applications. Thus, it is crucial to find the appropriate partners – providers, researchers and vendors – to overcome these gaps. Providers can also offer training programmes for clients to leverage these applications. However, even in these instances, it is essential to guarantee knowledge transfer and access to these technologies, building internal competencies and capabilities.

Question #6

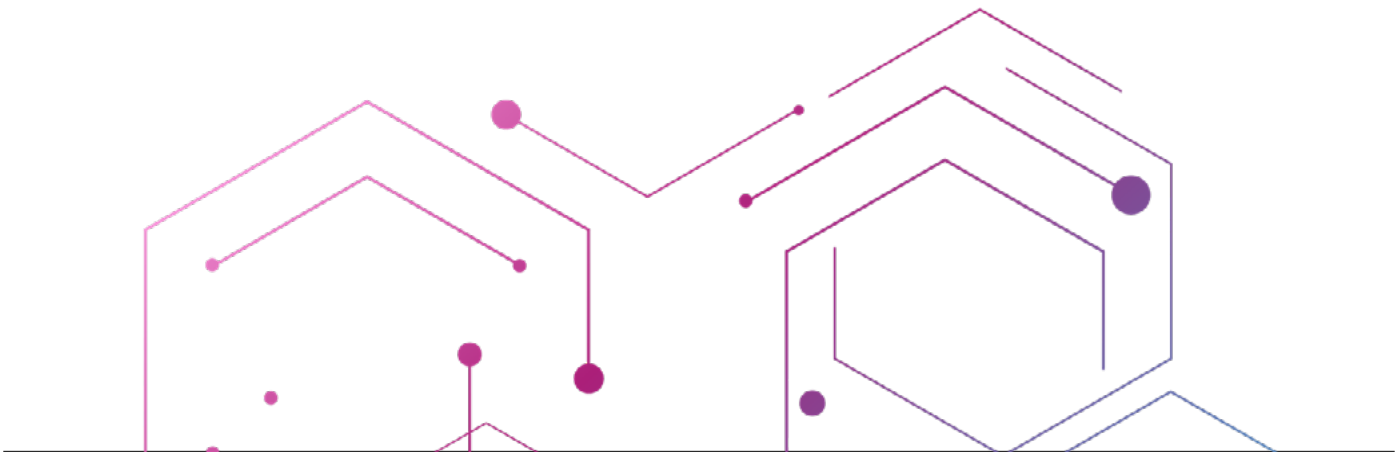
Is there already a market for federated learning or just a set of expectations?

Jack Vernon (IDC) responded that, especially in the healthcare context, building accurate applications requires leveraging these applications to access such a magnitude of data otherwise inaccessible. However, it is not possible to define it as a market yet, as players are just starting to leverage these technologies off the ground.

Question #7

Is heterogeneous data a new format fed into training/ML models or converted into a common scheme?

Jack Vernon (IDC) replied that federated learning technology providers situate data at customer node endpoints to guarantee that the models live at the Edge and that no exchanges or conversions into a new schema are needed.



Browse AI-Sprint and EUCloudEdgeIoT websites to know more about the projects

AI SPRINT
AI IN SECURE PRIVACY-PRESERVING COMPUTING CONTINUUM

EUCloudEdgeIoT.eu