Midagon Point of View

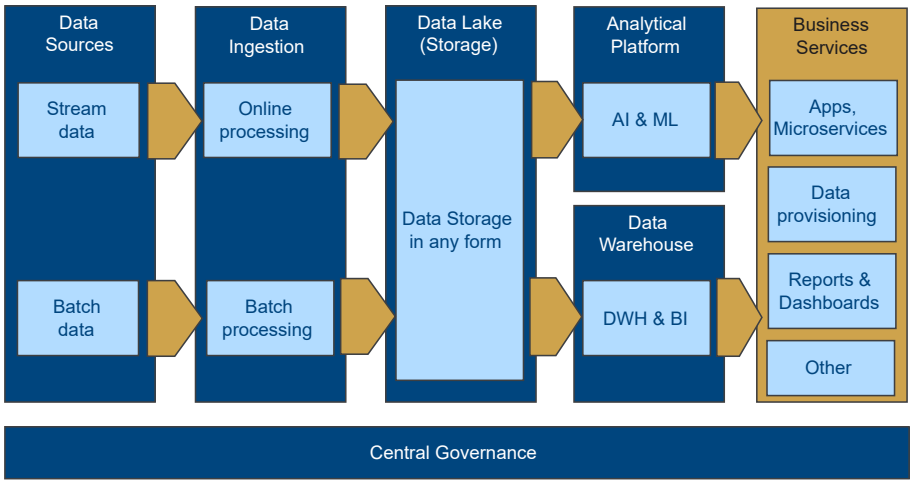
Enabling data-driven business with a Data Platform

In this Midagon Point of View, we focus on the Data Platform concept itself, and how it can improve data-based decision-making and lead to better new business services. The Data Platform is an essential part of the modern IT owned technology stack, providing opportunities for better business performance.

The utilisation of organisational business data to enable data-driven decisions has been a bit of a buzzword topic for a few decades now. First, Data Warehouses enabling Business Intelligence reporting with structured data marts became increasingly popular in the 1990s and 2000s, as data storage capacity and data computing power were still issues at this time. Then, in the early 2010s, "Big Data" enabled new business insights by analysing all an organisation's business data, becoming a key element on the wish lists of many businesses and IT organisations. Yet the amount of data vs. on-premise computing capacity and the availability of ready-made tools to analyze large amounts of data remained a challenge at this time. Only when Cloud service technologies such as Amazon Web Services, Azure and Google cloud became mainstream at end of 2010's did storage, computing and data analysis capacities with ready-made tools become readily available for many mid-size businesses. Now, in the 2020s, more and more organisations have harnessed this newfound power to enable data-driven business decision-making and utilize data for new online business services through the Data Platform concept.

A well-architected Cloud Data Platform solution with relevant governance can provide a single source for all of an organisation's business data. This in turn may enable a scalable environment for concurrency data workloads, with online data pipelines, business intelligence, and analytics with Artificial Intelligence (AI) and Machine Learning (ML). For users, this should create a seamless experience where an organisation's data is available in a single place, even if it is physically stored in multiple clouds and regions. A Data Platform also streamlines the development of different data application services with online business data, creating possibilities for new revenue streams and the utilisation of an organization's entire business data instantly and securely.

The Data Platform concept will be explained on a high level at the end of the article, but the focus of this Point of View is on raising some seldomly mentioned use cases, discussing the implementation topics and considering typical issues raised throughout the journey.



Considerations for implementing the Data Platform in an existing business environment

To get data to benefit an organisation, it is vital that the data itself is viewed as an integral part of the organisation's Strategic Planning. A Data Strategy with clear objectives creates an organisational commitment to achieving transformation through data utilization and allows the business to update current data solutions.

This fundamental thinking regarding data utilisation and the implementation of needed technologies and processes is enabled by good hybrid business and technical skills. If an organisation has a CDO (Chief Digital Officer or Chief Data Officer) role, this should be a catalyst for using data in strategy planning, shaping business processes towards data-driven business decision-making, and implementing good data governance to keep data valid as well as the processes that handle it.

The Data Platform concept is rarely implemented from scratch, as organisations tend to already have different kinds of data storage facilities in place, such as multi-Cloud environments and data warehouses for Business Intelligence (BI). When planning a Data Platform concept implementation vision, both business goals and technology solutions should be created with careful consideration of how the existing environment will be utilised. Multi-Cloud environments are supported by most Data Platform technologies, and there might even be on-premises data storage facilities that organisations seek to keep onsite for a variety of reasons, such as regulatory requirements. Hybrid or multi-cloud environments are technically possible (MS Azure for example), but the governance of the Data Platform in this case will be more complex and expensive compared to having a single data storage solution and set of tools.

Another aspect to consider would be what types of relevant data should be pushed onto the Data Platform. On the one hand, data is generally an under-utilised asset meaning sound advice would normally be to upload anything that may prove useful. On the other hand, we want to avoid excessive data waste, data complexity, and additional data lifecycle management requirements. A thorough case-by-case analysis is therefore necessary.

Since Data Platforms will become more extensive and complex over time, the big question becomes how to balance business benefits and solid platform development. How can you even get started on this journey? The Data platform usually can't be justified by any single business use case, yet platform development without solid business benefits will quickly fail.

Imagining what the Data Platform can do is sometimes difficult. We call this "flying car syndrome," as everyone thinks of different things when flying cars are brought up. Tackling this problem normally requires someone to build an initial version, making it easier to then develop further ideas and tackle additional issues.

The same approach has been successfully used with Data Platforms. The first use cases tend to bring quick business benefits while only using limited capabilities from the platform. This helps us to build confidence and further finance development. The journey should then be about balancing platform enhancements with new business use cases.

New opportunity: The Data Platform as an extension of ERP

The tradition has been to have point solutions for different data management needs. It is thus a concern that the Data Platform is considered too narrowly as a "new reporting solution." Yes, there are many beneficial reporting use cases which can be conveniently solved by Data Platforms. But there is also much more on offer. One of the big opportunities here is to use the Data Platform as a functional extension of your core systems.

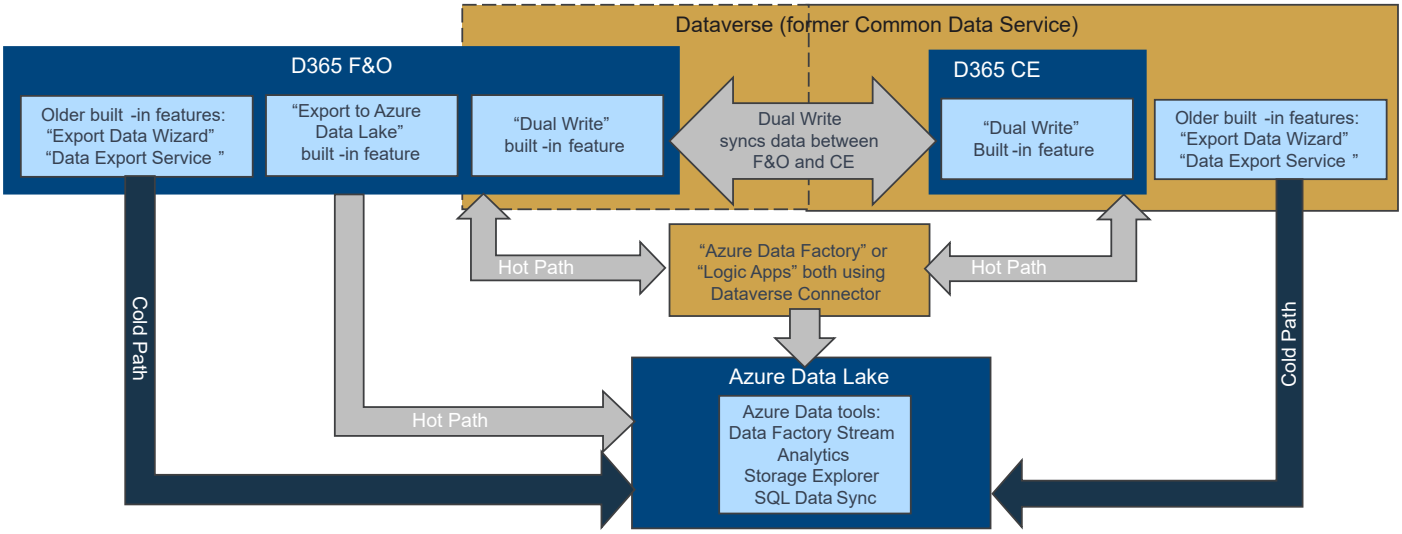
Since ERP's main purpose is to manage and integrate various business processes, such as finance, human resources, and supply chain management, ERP strategies can sometimes struggle to handle large amounts of data and provide real-time insights.

ERP systems usually rely on batch processing, which is a rather slow way to generate datasets to be shown in an online service like a mobile app or a web portal based on relevant requests.

This is where a Data Platform providing a one storage solution for all an organisation's data comes in, as it can be used as an extension of an ERP system to improve response times to milliseconds for online business services. In this scenario, the Data Platform will serve as a "data backend" for online services.

Modern cloud-based ERPs leave less room for specific tailoring, as this will quickly create a significant lifetime cost in evergreen maintenance in which you are forced to upgrade the ERP versions multiple times every year. One way to avoid the need for ERP tailoring is to move some of the non-standard functions to a near real-time ERP data replica on the Data Platform.

Methods for moving an organisation's data onto a Data Platform from cloud-based ERPs are now better than earlier versions. For example, in Microsoft Dynamics 365 Cloud ERP, data was first extracted and replicated as BYOD (Bring Your Own Data) SQL tables before then being transferred to a third location such as a Data Platform.



Microsoft has recently published new and more efficient ways to move data from Cloud D365 Finance & Operations (F&O) to a Data Lake:

- D365 F&O built-in technology to replace BYOD is an "Export to Azure Data Lake add-in" that connects D365 F&O tables (up to 350 of them) to a designated Data Lake. There is the possibility to select tables one by one or to select "Entities", a ready-made group of tables. With an "Enable near real-time data changes" feature, systems keep data in sync between D365 F&O and Azure Data Lake tables within a few minutes. This is quite an improvement on earlier BYOD data extraction.
- Azure Data Factory or LogicApps, both using the same Dataverse Connector (the earlier D365 Online Connector), are also BYOD dependent and may be used to move data near-online from D365 F&O to an Azure Data Lake using data change triggers. The difference between the two is that the D365 Connector for Azure Data Lake fulfils more complex data and transformation requirements.

Being able to have (near) online ERP data on the Data Platform together with a full D365 entity model also enables direct (BI) reporting from the replicated D365 F&O data model with more common user operated tools like PowerBI. Future trends will most likely give way to ready-made ERPs enabling basic reporting from the Data Platform.

A Data Platform makes data management and governance flaws more visible

Since Data Platforms typically bring data together from various systems, it makes all the data management shortcomings more visible. Typically, datasets do not match, there is a lack of sufficient master data and there are data protection challenges. All of this calls for better data governance.

Implementing a Data Platform governance layer can be a significant challenge, especially if data is flowing into the platform from multiple organisations in a group structure. A major issue is the potential lack of business ownership. With data coming in from multiple sources, it can be difficult to determine who is responsible for the overall management and governance of the platform. Naturally, Data Platform governance has a strong dependency on organisations' overall data management governance and master data management.

A lack of clear ownership can lead to confusion about roles and responsibilities and ultimately to reduced clarity as to the big picture of the Data Platform, since governance should cover both business and technology aspects. Without a single owner orchestrating such governance, it may be difficult to create consistent procedures and policies, leading to potential compliance issues like data privacy issues for example.

At Midagon, we feel that existing data governance models solve one of the burning issues, data ownership. But they totally miss the other key challenge of being able to match data with the business, and vice versa. This problem will not be solved by governance teams or committees, but rather by hybrid professionals who understand both the

business and the data. For example, they can immediately answer the question "does this sales figure include discounts and claims returns?" Or when there is something wrong with the figures, they can trace the source systems and root causes. Developing such hybrid skills is relatively slow and cannot be bought from outside. It also cannot be left to only a few individuals, as the competence will disappear if these individuals resign.

If governance layered implementation is not seen as an important priority, it may become a significant expense later on when it remains unconsidered and unstaffed in the Data Platform's implementation phase. The problem will then be compounded over the longer term as the overall functional complexity grows and it becomes more difficult to understand. In larger corporations, the most natural model is called data mesh. This is when data is classified into business-specific and group-common. The governance model is then built accordingly. You can read more about it here: https://www.datamesh-architecture.com/

Some optimal steps to building ideal Data Platform management in an organisation include:

1. Deploying a good overall data governance model for the whole organisation.
2. Appointing a Data Platform owner as the responsible authority regarding Data Platform decision-making.
3. Developing in-house hybrid professionals who know your business, your data, and their own role.

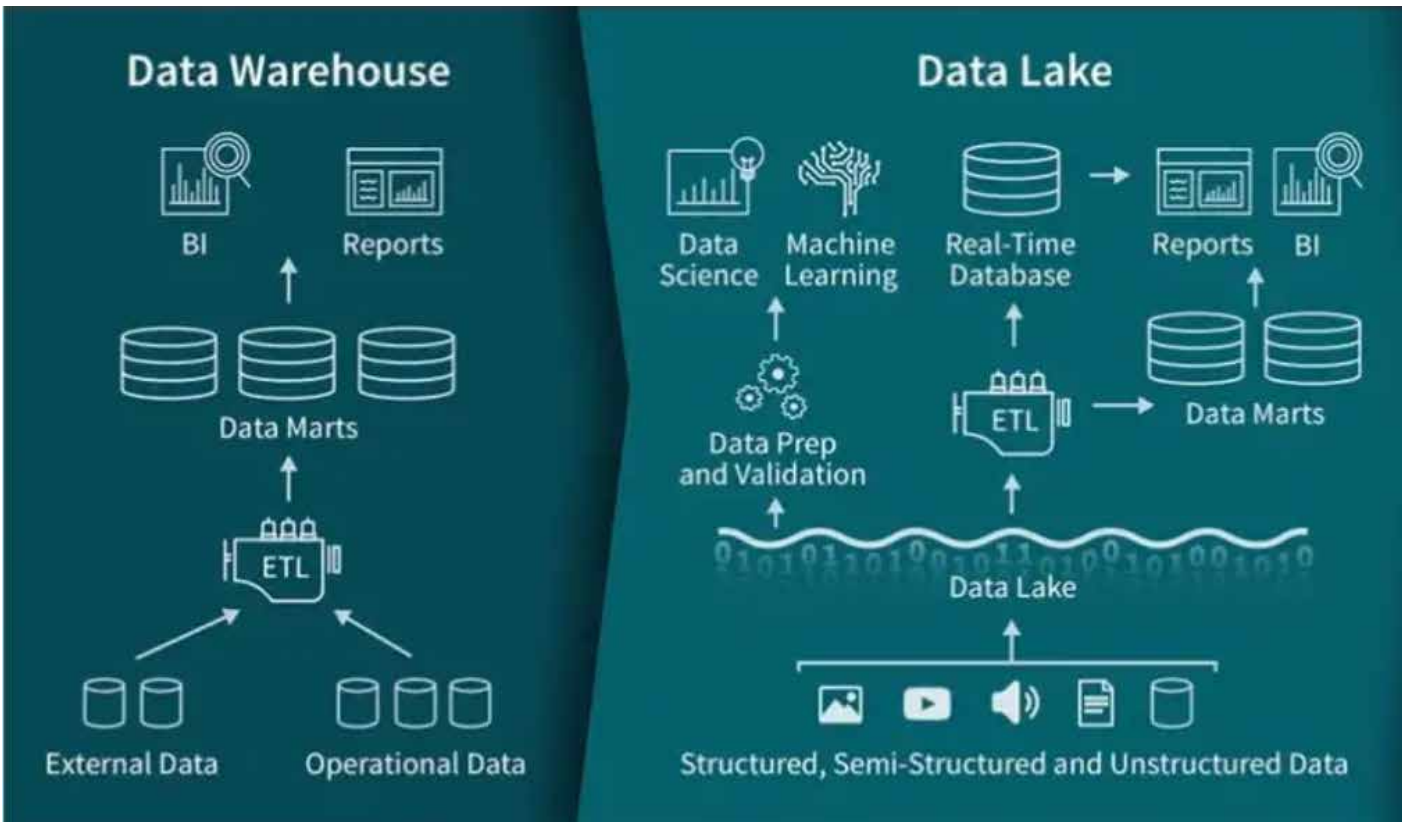
» Further reading

What is the Data Platform concept?

Main purpose of the Data Platform is to allow organisations to make better data-driven decisions and to enable more online business services by providing a central location for managing, storing, analysing, and providing their data.

The Data Platform concept is an integrated set of scalable (cloud) technologies and processes, where an organisations' business data is collected from various internal/external sources into a single data storage facility called a Data Lake. It is stored as raw data in any format (db, JSON, XML, PDF, video etc.) before then being processed for different business service purposes, like BI reporting and additional online business services.

The Data Platform is seen as a step up from Data Warehouses where data was collected from an organisation's IT systems and transferred into a very structured format in data marts through the ETL process for BI reporting. Data loads were usually done once a day. This is still considered a vital part of the Data Platform concept, but now with the addition of scalable cloud services. It also collects unstructured data and uses techniques like AI/ML, allowing for the development of many other online business services.



Source: Databricks blog – What is a lakehouse?

Business needs that the Data Platform should support through its collecting, storing and analyzing of large volumes of business data to arrive at fact-based decision-making include:

- **Enhanced business insights** – It may help to gain a better understanding of the business, e.g. from customers/ products, by providing tools for analysing data and identifying trends and patterns, e.g. for inventory management, marketing efforts, product/service development etc.
- **Faster real-time business agility** – It can enable businesses to respond more quickly to changing market conditions and customer needs by providing tools for analysing real-time data, e.g. for campaigns etc.
- **The identification of new business opportunities** – It may enable the exploration of new business models and revenue streams by providing solutions to existing problems and identifying new opportunities.
- **Real-time alerts** – It could make an alert by identifying different patterns in the data, e.g. fraud, predictive maintenance, identifying different kinds of risks etc.

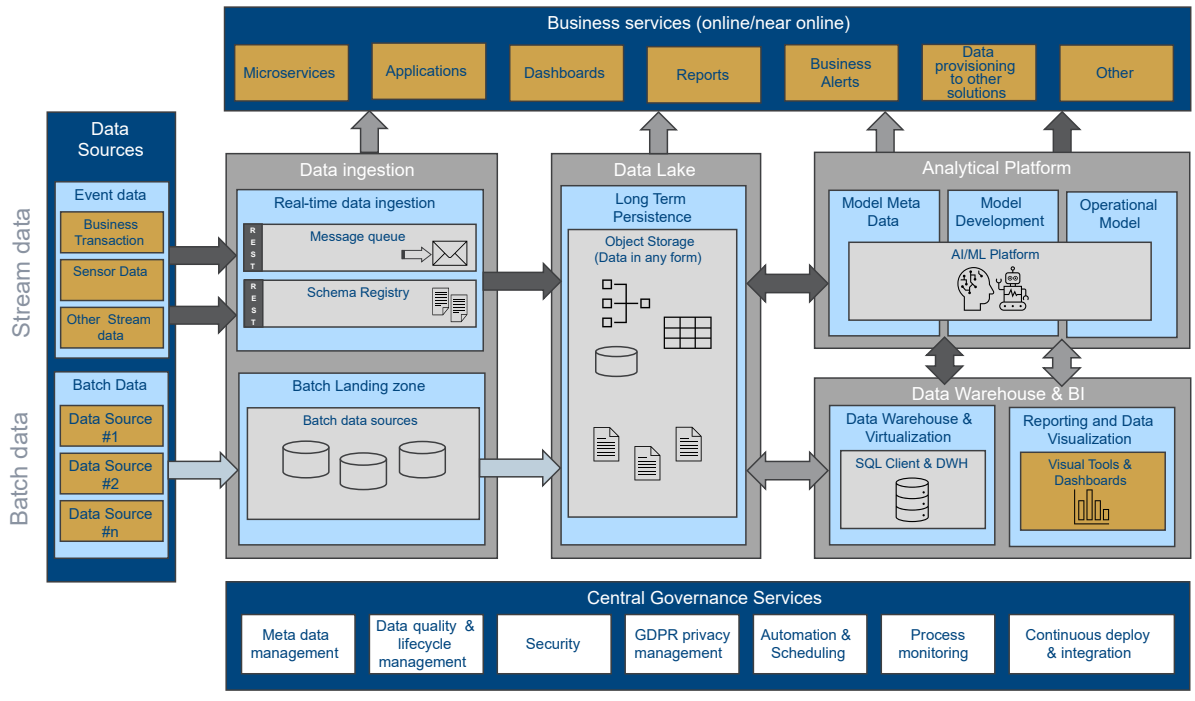
The Data Platform may also help to automate and streamline data-related tasks, such as data integration and data preparation, leading to improved efficiency that reflects in:

- **Improved data governance and security** – The Data Platform supports data governance by providing processes and controls to manage access to data and ensure compliance with relevant laws and regulations such as GDPR.
- **Increased data efficiency** – The Data Platform can help to automate and streamline data-related tasks, such as data integration and data preparation, leading to improved efficiency.
- **Reduced IT costs** – The Data Platform may also reduce costs by providing a central point of access for data that can be used by multiple systems and by automating and streamlining data-related tasks e.g. moving the needed tailoring from the ERP to the Data Platform for reporting and analysis.

Data Platform components

Data Platform components include (also see the picture below):

- Data sources – data from organisations' internal solutions or external services as batch or stream data.
- Data ingestion – the intake of data to the Data Platform from different sources as batch message intakes (for example ERP invoicing data) and real-time stream intakes (for example IoT sensor data).
- Data Lake – storage for raw data collected in any data format (e.g. XML, JSON, DB, PDF, etc.).
- Data Warehouse – data processed and optimised in a structured form for BI and other Business services.
- Analytical platform – data processed through Artificial Intelligence and Machine Learning to produce insights for Business services.
- Business Services – an outcome of utilising data (e.g. BI reports, dashboards, data provisioning, data services / microservices / apps, alerts etc.)
- Central governance – services for building, maintaining and running the Data Platform (e.g. meta-data management, data quality & lifecycle management, monitoring, automation, scheduling, GDPR compliance etc.).



The Data Platform's data processing pipelines are usually divided into two categories whereby the time of data processing from source to actual service is the dividing component. These two pipelines are explained below:

The Cold Path data pipeline (a.k.a. an offline pipeline):

- This refers to a data processing pipeline handling large amounts of historical data in batch mode, including data ingestion, processing, and analysis tasks that are performed on a needs basis for different not time-critical decisions.
- Internal and external data sources are usually batch data solutions, such as ERP, CRM, etc., and can sometimes be developed using more cost-effective technologies than Hot Path.
- A business use case example could be a retail company that stores customers purchasing data and creates targeted marketing campaigns. It could also for be example store performance analysis and making decisions to improve store efficiency and profitability.

The Hot Path data pipeline (a.k.a. an online pipeline):

- This refers to a data processing pipeline handling high volumes of reliable data in real-time or near-real-time, including data ingestion, processing, and analysis tasks that need to be performed quickly in order to enable real-time decision making or other data-driven activities.
- Internal and external data sources are usually stream data providing solutions, such as IoT sensors, GPS tracking etc., and typically require more resources and advanced technologies to support the high volume and real-time processing requirements like only memory-based processing. The implementation costs are therefore quite high compared to Cold Path technologies.
- A business use case example could be a taxi company that collects their vehicle locations, weather and traffic conditions from various sources. They can then use this data to optimise routes and the picking of passengers against orders online while also providing mobile app services to passengers.

Data Platform technology providers

If Data Platform technology is roughly divided into four parts, including Cloud service providers, Cloud Database Management providers, Business intelligence and AI/ML providers, the leading providers according to Gartner Magic Quadrant™ are as follows:

The Cloud Infrastructure and Platform Services (CIPS) 2022 Leaders are predictably Microsoft, Amazon Web Services and Google Cloud, offering access to infrastructure computing, networking, and storage resources, alongside various other services.

The Cloud Database Management Systems 2022 Leaders, seeking to fully supply provider-managed public or private cloud software systems that manage data on cloud storage, are the three CIPS Leaders accompanied by many more vendors, such as Oracle (utilizing AWS cloud) and Snowflake and Databricks (utilizing any of three CIPS providers).

The Analytics and Business Intelligence Platforms 2022 Leaders, seeking to analyse, explore, share, and manage data and allow users to uncover and visualise insights, are Microsoft, Salesforce (acquired Tableau) and Qlik.

The Cloud AI Developer Services 2022 Leaders are Microsoft, AWS, Google and IBM. Here, AI is defined as applying advanced analysis and logic-based techniques, including machine learning (ML), to interpret events, support and automate decisions and take action. Cloud AI developer services (CAIDS) are cloud-hosted or containerized services that enable development teams and business users who are not data science experts to use artificial intelligence (AI) models via APIs, software development kits (SDKs) or applications.