

Management and technology challenges of AI-enabled application projects.

BY RUA-HUAN TSAIH, HSIN-LU CHANG,
CHIH-CHUN HSU, AND DAVID C. YEN

The AI Tech-Stack Model

PRESENTLY, ENTERPRISES HAVE implemented advanced artificial intelligence (AI) technologies to support business process automation (BPA), provide valuable data insights, and facilitate employee and customer engagement.⁷ However, developing and deploying new AI-enabled applications poses some management

and technology challenges.^{3,5,12,15} Management challenges include identifying appropriate business use cases for AI-enabled applications, lack of expertise in applying advanced AI technologies, and insufficient funding. Concerning technology challenges, organizations continuously encounter obsolete, incumbent information technology (IT)/information systems (IS) facilities; difficulty and complexity integrating new AI projects into existing IT/IS processes; immature and underdeveloped AI infrastructure; inadequate data quantity and poor-quality learning requirements; growing security problems/threats; and inefficient data preprocessing assistance.

Furthermore, major cloud service vendors (for example, Amazon, Google, and Microsoft) and third-party vendors

» key insights

- Although AlaaS offerings allow companies to take advantage of AI without investing massive resources from scratch, numerous issues have emerged to hinder the development of desired AI systems.
- The proposed modularized, loosely coupled, seven-layered AI tech-stack model can leverage AlaaS offerings to help organizations resolve management and technology challenges. The seven layers—from bottom to top—are AI infrastructure, AI platform, AI framework, AI algorithm, AI data pipeline, AI services, and AI solution layers.
- A synthesized IT/IS framework is proposed to address the implementation issues of a desired AI system. The framework incorporates three capabilities (connection, incumbent IT/IS, and AI) that impact the development/deployment of a desired AI system and underlying in-house/outourcing decisions.

Table 1. List of abbreviations.

Abbr.	Definition
AI	Artificial Intelligence
AIaaS	AI as a Service
API	Application Programming Interface
AutoML	Automated Machine Learning
CI/CD/CT	Continuous Integration/Continuous Delivery/Continuous Testing
CMM	Capability Maturity Model
CRM	Customer Relationship Management
DaaS	Data as a Service
DataOps	Data Operations
DBP	Digital Business Platform
DevOps	Development Operations
ERP	Enterprise Resource Planning
FPGA	Field-Programmable Gate Arrays
GPU	Graphics Processing Unit
IaaS	Infrastructure as a Service
IEP	Intelligent Engagement Platform
IT/IS	Information Technology/Information System
KM	Knowledge Management
MLOps	Machine Learning Operations
NN	Neural Networks
NLP	Natural Language Processing
OA	Office Automation
PaaS	Platform as a Service
SaaS	Software as a Service
SCM	Supply-Chain Management
SLA	Service-Level Agreement
STRS	Smart Tourism Recommendation System
TPU	Tensor Processing Unit

Table 2. The assessment of AIaaS offerings from current major cloud service and third-party service vendors.

AIaaS Providers	Strengths ^{4,15}	Weaknesses ^{4,15}	Benefits the Proposed Model Should Provide
AWS ^{1,2,4,13,14,15,20}	<ul style="list-style-type: none"> ► Strong cloud offerings ► Innovation leader ► Engineering support 	<ul style="list-style-type: none"> ► Interoperability issues ► Lack of open community support 	<ul style="list-style-type: none"> ► Modularization can improve interoperability among cloud vendors and encourage open source community support. ► The unambiguous conceptualization of the layered technology components enables executives to define clear SLAs. ► The layered architecture simplifies implementation in a specific layer, efficiently reducing resilience and complexity concerns. ► The proposed decomposition concept mitigates vendor lock-in and proprietary concerns. ► The loosely coupled layered model helps organizations integrate the desired SaaS solution with other AIaaS offerings to achieve competitive advantage.
Google ^{1,4,13,15,20}	<ul style="list-style-type: none"> ► Innovation velocity in AI backbone ► Strong open community support 	<ul style="list-style-type: none"> ► Low extensibility of DataOps ► Lack of post-sales support 	
Microsoft ^{1,4,13,15,17,20}	<ul style="list-style-type: none"> ► Leader in enterprise MLOps ► Broad IS offerings ► Strong enterprise relationships 	<ul style="list-style-type: none"> ► Commercial complexity ► Resilience issues 	
IBM ^{1,4,13,15}	<ul style="list-style-type: none"> ► Strong IT resource support ► Watson AI offerings ► AI governance 	<ul style="list-style-type: none"> ► Legacy concerns ► High product-bundling costs ► Weak AutoML offerings 	
Salesforce ⁹	<ul style="list-style-type: none"> ► Strong SaaS solutions in CRM ► Leader in CRM APIs 	<ul style="list-style-type: none"> ► No public cloud services ► Limited AIaaS offerings ► Low extensibility of AutoML 	
SenseTime ²⁴	<ul style="list-style-type: none"> ► Strong SaaS solutions in surveillance ► Leader in surveillance APIs 	<ul style="list-style-type: none"> ► No public cloud services ► Limited AIaaS offerings ► Low extensibility of AutoML 	

(for instance, Salesforce and SenseTime) have stepped up efforts as major players in the AI-as-a-service (AIaaS) race by integrating cloud services with AI core components (for example, enormous amounts of data, advanced learning algorithms, and powerful computing hardware).⁴ Although AIaaS offerings allow companies to leverage AI power without investing massive resources from scratch,⁸ numerous issues have emerged to hinder the development of desired AI systems. For example, current AI offerings are recognized as a fully bundled package, offering less interoperability between different vendors and causing vendor lock-in and proprietary concerns. In addition, the tightly coupled components of different layers limit the extension of new functionality and inhibit developers' flexibility and adaptability when choosing suitable AI components for practical implementation. Moreover, when a vendor bundles several AI offerings into one package, reliability becomes questionable since it is challenging to define a transparent service-level agreement (SLA) for each AI offering. Furthermore, bundled AI offerings are perceived as tightly controlled systems that inhibit the open source community's support and raise the lock-in cost, which increases potential incompatibility and introduces future migration costs among different vendors.

To address these issues, the mixed integration method²⁰ was used to integrate cloud stacks with AI core components to propose the seven-layered AI tech-stack model. The seven layers, from bottom to top, are AI infrastructure, AI platform, AI framework, AI algorithm, AI data pipeline, AI services, and AI solution. The AI infrastructure layer is closest to the machine, while the AI solution layer is closest to the end user. The proposed model aims to address the following two issues in developing and deploying AI-enabled applications:

- Leveraging AIaaS offerings to help organizations resolve management and technology challenges

- Integrating the desired AI system with incumbent IT/IS facilities and synthesizing digital business platforms (DBPs) into viable IT/IS frameworks.


The remainder of this article is organized as follows. We review AIaaS offerings and benefits provided from the proposed model, detail the AI tech-stack model, and cover the synthesized IT/IS framework. Then, we provide case studies, discussing the deployment of the proposed model into smart tourism applications. The article concludes by covering the academic/practical implications of this study and discussing future work. To enhance readability, all abbreviations used in the main text are provided in Table 1.

AIaaS Offerings


Major cloud and third-party service vendors have begun integrating AI core components into cloud services to enable organizations to develop, train, deploy, and manage AI-enabled applications in the cloud. These services, known as AIaaS, aim to make AI adoption accessible and affordable to any organization, regardless of the organization's size, level of technological advancement, or funding.¹⁷

The strengths and weaknesses of AIaaS offerings from some of the major providers^{1,6,10,11,13,14,16,19,24} are summarized in Table 2. The assessment of current AIaaS offerings shows that major AIaaS providers develop and fine-tune their existing cloud services to offer additional AI capabilities in specific—but not all seven—AI tech-stack layers. For example, IBM excels in providing AI data-stack solutions and development tools (for instance, IBM®Watson AI) to serve data scientists and/or engineers.¹⁰ Microsoft emphasizes machine-learning operations (MLOps) and lifecycle-management platforms that include data preparation and feature engineering functions.¹⁶ Other cloud infrastructure-as-a-service (IaaS) providers, such as Amazon and Google, provide AI infrastructure, AI platforms, and AI data-pipeline services.^{19,21} Furthermore, cloud software-as-a-service (SaaS) vendors such as Salesforce and SenseTime provide application programming interface (API) functions to customers seeking to leverage their big data analytics solutions.¹³

Table 2 also highlights the proposed model's associated benefits, which justify the need for the AI tech-stack model as a guideline to help organi-



Although AIaaS offerings allow companies to take advantage of AI power without investing massive resources from scratch, numerous issues have emerged to hinder the development of desired AI systems.



zations efficiently and cost-effectively develop and deploy new AI-enabled applications. Checking current academic studies and/or practitioner development, no such model exists. Consequently, this study is timely and significant. Specifically, the proposed model offers a layered architecture to help executives understand which layer might best fit a specific vendor's AIaaS offerings and, equally important, differentiate among multiple vendors' offerings to realize competitive advantages. The layered architecture also enables executives to reduce/decompose meta-level management and technology challenges to the layer perspective so that interoperability and reliability are further enhanced. This is simply because when using the decomposition concept, each vendor's AIaaS offering corresponds to one layer's functionality; thus, a change of vendor will not impact a change of vendors in other layers through the modularization architecture. Furthermore, IT managers can replace existing AI functions with better alternatives without interfering with the working functions of other layers. For example, a company may implement its own AI-enabled applications on Amazon AWS or Microsoft Azure. Switching to a different AI infrastructure may only raise management and technology issues for that specific AI infrastructure layer without affecting the layers above it.

The AI Tech-Stack Model

To define each specific layer in the AI tech-stack model, the following principles were adopted:

1. Categorize similar functions into the same layer to enable function changes within a layer without affecting other layers.
2. Create a boundary at a point where the service description can be concise and the number of interactions across boundaries is minimized.
3. Decompose layers for handling AI jobs that are manifestly unique in task description or skill requirements.
4. Develop a boundary at a point where industry solutions are available and have proven useful.

It is essential to combine IaaS and the AI accelerators to provide the hardware capacity needed to handle big data and AI-enabled applications in the

Figure. The synthesized IT/IS framework for AI-enabled application projects.

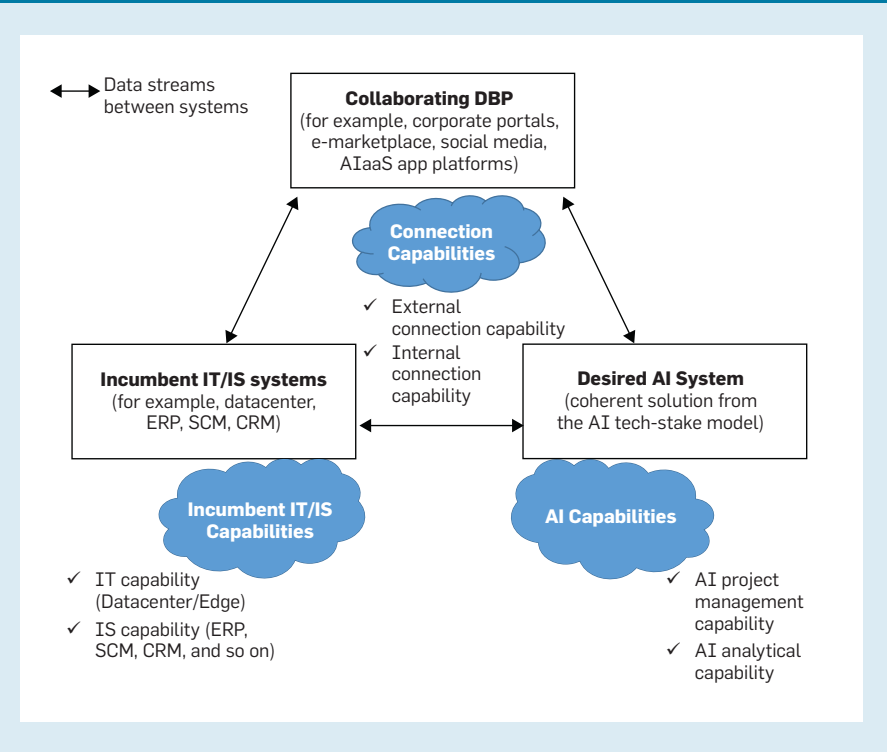


Table 3. AI tech-stack model.

AI Tech Stack		Definitions	Functionalities
SaaS + AI-embedded functions	AI Solution 7	AI-enabled solutions to address business applications in a specific domain.	A variety of capabilities to deliver business intelligence in a specific function domain (for example, customer service) or industry domain (for example, banking).
	AI Service 6	Well-defined AI-enabled services that can be fulfilled by API function call request.	General-purpose API services such as recommendation, ranking, intelligent search, language translation, and object detection.
DaaS + DataOps	AI Data Pipeline 5	Data lifecycle-management platform to generate curated data and feature engineering for AI/ML model training.	DaaS, DataOps platform for data pipeline automation, orchestration, visualization and feature operations, and more.
AI software modules	AI Algorithm 4	Training methods by which the AI system conducts its learning task.	Supervised, unsupervised, and reinforcement learning.
	AI Framework 3	AI-related frameworks for defining AI/ML architectures and invoking algorithms, library, and accelerator drive on the hardware.	Tensor computing, AI libraries (for example, Torch, TensorFlow), prebuilding AI models (for example, neural network models), and so on.
PaaS + MLOps	AI Platform 2	A unified interface that facilitates communication between software and hardware throughout the ML lifecycle.	PaaS, AI/ML lifecycle-management platform (for example, MLOps, IEP) to build, train, test, evaluate, serve, and monitor AI-enabled applications.
IaaS + Accelerator	AI Infrastructure 1	Hardware units that orchestrate and coordinate computations among accelerators.	IaaS, AI accelerators for highly parallel operations (for example, GPU, TPU, FPGA), tools for monitoring, clustering, and so on.

upper layers. Thus, the AI infrastructure layer is identified as the lowest layer in the AI tech stack. To facilitate communication between the hardware and application software throughout the ML lifecycle, the AI platform layer sits atop the AI infrastructure layer to provide a unified user interface on which MLOps engineers can collaborate across the ML lifecycle. This includes building, training, evaluating, deploying, and monitoring AI models, while platform-as-a-service (PaaS) providers supply operating systems, databases, middleware, and other services to host the user’s application. Running on the AI platform are AI-specific software modules, which refer to the tools and algorithms that allow the system to exhibit intelligence. The tools that invoke executable algorithms on the hardware through the AI platform were identified as the AI framework layer, which sits atop the AI platform layer. This layer provides pre-built AI models, accelerator drivers, and support libraries to help data scientists and AI developers build and deploy AI models faster and easier.

AI algorithms are a set of training methods by which the AI model conducts its learning task. They form the AI model’s core capability, through which the model can “learn” the data and produce outputs when given new inputs. Thus, the AI algorithm layer sits atop the AI framework layer. An AI algorithm cannot function without data. Namely, an AI model is doomed to failure if the data is poorly structured, inaccurate, and inconsistent. Therefore, the need for a dedicated layer for data processing and management was identified, leading to the establishment of the AI data pipeline layer, which rests atop the AI algorithm layer. Finally, AI-enabled applications are offered by SaaS providers or in-house software engineers to embed AI-specific functions in their business applications. Some AI-enabled applications, such as facial authentication, image recognition, and natural language understanding, support specific AI tasks in a general domain. These AI building-block services, which end users can consume through APIs, comprise the AI service layer, which sits atop the AI data-pipeline layer. The remaining set of AI-enabled applications are those delivering broader

capabilities to address a larger scope of business problems about a specific industry's or company's domains, such as intelligent customer solutions for banking. Consequently, we consider these AI-enabled business solutions as comprising the AI solution layer—the tech stack's highest layer.

Table 3 illustrates the resulting AI tech-stack model with seven layers. A more detailed illustration of each layer, starting from the AI infrastructure layer at the bottom and ending at the AI solution layer at the top, is provided in the following paragraphs.

AI infrastructure layer. This layer integrates all IaaS components,²³ accelerators, such as graphics processing units (GPUs), tensor processing units (TPUs), and field-programmable gate arrays (FPGAs); and additional services, such as monitoring, clustering, and billing tools, to provide the hardware infrastructure environment for essential computation, storage, and network communication required by AI-enabled applications.

AI platform layer. This layer integrates the PaaS components (for example, operating systems, programming-language execution environment, database, web server, and so on), MLOps, and intelligent engagement platform (IEP) for performing complete AI-enabled application lifecycle management. Specifically, it coordinates model building, evaluation, and

monitoring to ensure the execution of AI applications' continuous integration (CI), continuous delivery (CD), and continuous testing (CT).

AI framework layer. This layer contains all AI-related frameworks to accelerate the process of AI-enabled application development and deployment, including tensor computing with strong acceleration via GPUs; the automatic differentiation system, pre-building AI algorithm libraries, such as Torch and TensorFlow; and pre-building AI models, such as neural networks (NNs).

AI algorithm layer. This layer provides many well-defined open source and/or self-customized sets of AI algorithms (for example, supervised, unsupervised, and reinforcement learning) to help perform problem-solving and decision-making tasks.

AI data pipeline layer. This layer includes data as a service (DaaS) and the DataOps platform, which integrates various data architectures to facilitate the end-to-end data lifecycle. This layer also provides data-preprocessing functionalities and feature-engineering capabilities for managing internal and external data sources, manipulating various stationary and nonstationary data types, and handling batched and real-time data access.

AI service layer. This layer contains many ready-to-use, general-purpose APIs for AI-enabled services, such as

image processing and nature language processing (NLP). The API transfers either the information (for example, insights derived from ML technologies) or raw data (for instance, source data for deriving insights through ML technologies) that can be executed with existing IT applications and/or enterprise systems—for example, enterprise resource planning (ERP), customer relationship management (CRM), and supply-chain management (SCM)—to deliver solutions in the upper level.

AI solution layer. This layer contains AI-enabled solutions to address problems in a specific business domain. With the AI solution layer, business analysts help domain users deliver broad AI capabilities in different companies or industry domains.

The proposed AI tech-stack model can be deemed as a conceptual framework and does not map to specific systems. However, clearly characterizing the services provided by each layer, the model facilitates vendor interoperability because vendors can follow the same rules and requirements to develop products and services at one specific layer, acquire services from the layer below, and expose a consistent service interface to the layer above. It also means that users do not need to rely on a single vendor to provide all products and services, from the physical layer to the application

Table 4. AI tech-stack interoperability matrix.^{2,22}

Layers/ Providers	AlaaS	Third-Party	Open Source
AI Solution	Amazon Personalize, Google DeepMind, IBM Watson, Alibaba Cloud Intelligence Brain, Tencent Video and Media AI, and so on.	Oracle AI, Salesforce Einstein, SenseTime, Appier, UiPath RPA, Blue Prism RPA, Plantier, Affirm, Teladoc Health, Uber, Airbnb, and so on.	Acumos AI, Mycroft.ai, Rasa Open Source, Tesseract OCR, and so on.
AI Services	Amazon Lex, Rekognition API, Google Translation, NPL API, Microsoft Teams, Azure Bot API, Alibaba Open API, Tencent Video API, and so on.	Oracle Forecasting and Anomaly Detection API, Salesforce Tableau CRM, SenseTime facial recognition API, Bloomberg API, Reuters API, PlaceIQ API, and so on.	Foursquare Places API, Data.gov open data API, Berkeley DeepDrive API, Weather API, and so on.
AI Data Pipeline	Amazon SageMaker Data Wrangler, Amazon SageMaker Feature Store, IBM business analytics, Google Looker, Google Dataflow, Microsoft Azure DataOps, and so on.	SAS, DataRobot, DataKitchen, DataBricks, Dataiku, Hopworks, Snowflake, Tableau, Uber's Michelangelo, and so on.	Apache Airflow, Prefect, FEAST, KNIME, and so on.
AI Algorithm	Customized algorithm.	Customized algorithm.	ANN, CNN, RNN, GAN, OpenCV, OpenNN, and so on.
AI Framework	Google TensorFlow, Microsoft CNTK, Amazon Theano, and so on.	SenseTime Parrots, Salesforce Einstein, and so on.	Keras, Torch, Scikit-learn, Caffe2, JAX, Kubernetes, Apache MXNet, and so on.
AI Platform	Amazon SageMaker, Google Vertex, IBM Watson Studio, Microsoft Azure AutoML, and so on.	RapidMiner, H2O, TinyMLOps, TIBCO, Alteryx, Domino, C3.AI, and so on.	MLflow, DVC, Jupyter Lab, Kubeflow, and so on.
AI Infrastructure	Amazon, Google GCP, Microsoft Azure, Alibaba Cloud, Tencent Cloud, and so on.	H2O, Cloudera, and so on.	Apache Zookeeper, MESOS, ARGO, and so on


layer, thereby minimizing the threat of vendor lock-in.

As shown in Table 4, users can select services across layers from different vendors.^{2,22} Taking smart CRM as an example, companies can choose Amazon as the AI infrastructure provider; MLflow to get the open source MLOps offering in the AI platform layer; the open source framework (for instance, TensorFlow, Keras, and PyTorch) in the AI framework layer; some in-house, customized, pre-built algorithms in the AI algorithm layer; third-party Hopsworks solutions to get DataOps tools in the AI data pipeline layer; external NLP and recommendation APIs in the AI service layer; and some in-house smart CRM solutions extending from incumbent CRM with AI-enabled NLP and recommendation services in the AI solution layer. In summary, the proposed AI tech-stack model allows companies to have more informed choices among layers, choose AIaaS and third-party vendors based on SLA and pricing options, and leverage open source solutions to maintain flexibility.


Applying the AI Tech-Stack Model in Enterprises

To understand the implementation issues of a desired AI system, this study adopted a systemic perspective to derive the synthesized IT/IS framework as shown in the Figure. When corporate introduces advanced AI technologies into the organization to develop a desired AI system, it is required to integrate advanced AI technologies with the incumbent IT/IS systems and to leverage the collaborating DBP. The incumbent IT/IS systems refer to the firm's legacy IT/IS and the data center used for handling the firm's present activities, such as ERP, SCM, CRM, office automation (OA), and knowledge management (KM), to facilitate and manage fundamental corporate data streams—information flow, cash flow, commodities flow, and logistics activities. We propose that the synthesized IT/IS framework incorporates three capabilities that impact the development and deployment of a desired AI system and the underlying in-house/outourcing decisions.

Connection capabilities. Connection capabilities enable the incumbent IT/IS



The proposed AI tech-stack model can be deemed as a conceptual framework and does not map to specific systems.



system, the collaborating DBP, and the desired AI system to communicate with each other to extract internal data and obtain external data in a usable form for analytics and ML. Therefore, solutions provided in the AI data pipeline layer are expected to access internal and external data sources much easier and hence, broaden the scope of accessible data and more easily integrate with other affordable solutions. To this end, it is expected that companies with low connection capabilities prefer to outsource their AI data pipeline to AIaaS or third-party vendors to meet their data enrichment and employment needs. Organizations with robust capabilities tend to prefer an in-house option.

Incumbent IT/IS capabilities. Incumbent IT capabilities provide organizations with the supportive hardware infrastructure environment for essential computation, storage, communication, and security. It is expected that organizations with low incumbent IT capabilities tend to adopt the off-the-shelf AI infrastructure service. On the other hand, incumbent IS capabilities are related to organizations' ISSs, such as ERP, SCM, and CRM, and their accumulated data. It is expected that corporations with low incumbent IS capabilities have poor-quality or insufficient internal data/information for data analysis and thus tend to gather alternative data from a variety of external data sources through ready-to-use, general-purpose APIs provided by AIaaS (for example, Google analytics) or third-party vendors. Corporations with high IS capabilities usually prefer an in-house option.

AI capabilities. AI capabilities refer to (1) the AI analytical capabilities needed to excel at extracting insights and patterns from large datasets through the employment of AI algorithms, and (2) the AI project management capabilities to enable AI's CI/CD/CT across AI project lifecycles while encouraging innovation by using the unified AI framework tools. It is expected that corporates with low AI capabilities tend to outsource the platform, framework, and algorithm of AI to AIaaS and third-party vendors. These vendors also offer different types of pre-trained models and customized algorithms with which companies can perform data analyt-

ics through intuitive, no-code tools. Therefore, organizations can take advantage of derived insights without massive upfront investment in acquiring talent and resources. Corporations with high capabilities may desire an in-house option.

To further justify our propositions, we conducted an empirical investigation of four major tourism companies located at Taiwan. Tourism has been the industry most affected by the COVID-19 pandemic and has urgently sought digital transformation opportunities to boost business in the post-pandemic market by activating an AI-driven smart tourism strategy. We explained the proposed AI tech-stack model in an executive MBA class, and four companies expressed their interest in applying the model to their desired smart tourism recommendation system (STRS). In the next section, case-based research is conducted before the presentation of the key insights from our case analysis.

Case Study

The case companies, shown in Table 5, were Lion Travel (L), Colatour Trav-

el (C), Tripaa Travel (T), and Foru-Tek Travel (F). Companies L and C are conventional travel agencies that offer customers a large number of outbound, guided tours and thus have a large amount of historical transactional data in their ERP systems. Therefore, their desired STRSs focus on cross-selling and upselling by recommending new trip destinations to attract existing customers. On the other hand, companies T and F are startups providing premium and independent tours. Their customers are mostly independent travelers who are quite familiar with online booking tools and applications. Furthermore, company T focuses on inbound tourism, while company F emphasizes on outbound tourism to Japan. Their desired STRSs place a unique focus on understanding their target audiences and generating new sales.

Applying the AI tech-stack model to the STRS projects of these four companies, we list the identified and desired AI offerings of each layer in Table 6. Since the AI tech-stack model identifies the essential AI resources required in each individual layer,

the participating executives can easily pinpoint their specific concerns for each individual layer. Specifically, combined with the work-out context of the synthesized IT/IS framework in the , the AI tech-stack model allows executives to assess and evaluate their current status and desired capabilities at each layer. Therefore, to agilely develop and deploy the STRS, each company can use the evaluated results to choose either an in-house design of its own AI offering or various AIaaS offerings pertaining to each layer.

Table 7 summarizes the analysis results. As provided, all four companies tend to adopt the off-the-shelf AI infrastructure service because they have very limited incumbent IT capabilities to cope with the AI infrastructure requirement. Specifically, IT staff in companies L and C do not have the skills needed to use advanced AI hardware and software. Digitally savvy companies T and F have already outsourced products from public cloud vendors that they can tailor in their IT facilities and thus have no major problems migrating to the AI infrastructure through their current cloud

Table 5. Company profiles of interviewees.

Company Name (code name)	Company URL	Travel Type	Focus of Travel Services	Focus of Tourism Destination	Annual Number of Tourists	Interviewee Position
Lion Travel (L)	www.liontravel.com	Leisure and business	Escorted tours	Outbound, inbound, and domestic	3 million	VP
Colatour Travel (C)	www.colatour.com.tw	Leisure and business	Escorted tours	Outbound, inbound, and domestic	2 million	CEO
Tripbaa Travel (T)	www.tripbaa.com	Leisure	Independent tours	Domestic	0.2 million	CEO
Foru-Tek Travel (F)	www.foru-tek.com	Leisure	Independent tours	Domestic	1.4 million	CEO

Table 6. AIaaS offerings for STRS.

	Smart Tourism Recommendation System	Offerings from AIaaS and Third-Party Vendors
AI Solution	Smart tourism recommendation system	Not available.
AI Services	Smart tourism classification service, smart tourism ranking service, smart tourism recommendation service, and so on.	Amazon Personalize, Microsoft Azure Cognitive Search, Google Web search, SkyScanner API, Airline API (IDS), Hotel booking API, and so on.
AI Data Pipeline	Smart tourism data pipeline tools for external and internal data sources.	AWS SageMaker Feature store, IBM Data lake, Google data Curation, Snowflake, Splunk, Dataiku, DataRobot data platform solution (DataOps), and so on.
AI Algorithm	Algorithms for enhanced classification, regression, cluster, association, and so on for smart tourism.	Pre-built algorithms (ANN, SVM, DT) and customized recommendation algorithms, and so on.
AI Framework	TensorFlow+Keras, PyTorch, Parrot, and so on.	TensorFlow+Keras, PyTorch, Parrot, Caffe2, MxNet, SKLearn, and so on.
AI Platform	PaaS, AI/ML lifecycle management (CI/CD/CT), and so on.	SageMaker Studio, Google Cloud AutoML, IBM Watson Studio, Microsoft AI Tools, MLflow, and so on.
AI Infrastructure	IaaS, GPU, TPU, FPGA, edge computing, and so on.	AWS, Google Cloud, Microsoft Azure, Alibaba Cloud, H2O, and so on.

Table 7. Capability assessments and in-house and/or outsourcing decisions for case companies.

Capabilities	Capability/ Company	L	C	T	F
Connection capabilities	External connection	Low	Low	Moderate	Moderate
	Internal connection	Moderate	Moderate	Low	Low
Incumbent IT/IS capabilities	Incumbent IT	Low	Low	Low	Low
	Incumbent IS	High	Moderate	Low	Low
AI capabilities	AI project management	Low	Low	Low	Low
	AI analytical	Moderate	Moderate	Moderate	High
In-house and/or outsourcing decisions for case companies					
Smart Tourism Recommendation System	AI Solution	●	●	●	●
	AI Service	▲	▲	▲	▲
	AI Data Pipeline	☆	☆	★	★
	AI Algorithm	■	■	■	■
	AI Framework	■	■	■	■
	AI Platform	■	■	■	■
	AI Infrastructure	■	■	■	■
● In-house design		▲ Off-the-shelf offerings for information and data			
■ Off-the-shelf offerings		☆ Off-the-shelf offerings mainly for internal data			
▲ Off-the-shelf offerings for information		★ Off-the-shelf offerings mainly for external data			

service partners. Regarding the AI platform and AI framework layers, all four companies have low AI project management capabilities and thus tend to adopt the AI platform and framework service chosen from AIaaS providers.

In addition, companies L, C, and T need the ability to learn and understand user behavior to help them predict user preferences about specific products. This is a classic recommendation problem that does not require strong AI analytical capabilities and can be resolved via existing AI algorithms. Consequently, these three companies prefer to use currently available AI algorithms to meet their desired recommendation needs. However, company F's customers are younger and prefer an independent, customized, and exclusive Japan tour experience. It may not be suitable to offer these customers conventional recommendations. To meet such expectations, company F has developed some AI analytical capabilities to design its own algorithms for providing its unique customers with travel suggestions.

Related to the AI data-pipeline layer, all four companies have low or moderate connection capabilities. They are unable to develop their own

data-pipeline platforms so they tend to take the AIaaS offering. Specifically, companies L and C have developed relatively stronger internal connection capability because they focus on analyzing the historical travel records to recommend proper travel plans for customers. In contrast, companies T and F have fewer customers and prefer to focus on external connection capabilities, permitting them to analyze external travel-data queries and social media-associated records to recommend adequate travel plans for their customers.

Concerning the AI service layer, companies L and C have stronger incumbent IS capabilities than companies T and F. Because they have already collected good-quality historical data, their main purpose in adopting the off-the-shelf open API is to leverage information from collaborating DBPs to derive useful business insights. In contrast, companies T and F have low incumbent IS capabilities and much smaller internal databases; thus, they prefer to adopt the off-the-shelf open API to obtain external (for instance, structured and unstructured) data from many collaborating DBPs.

Finally, in the AI solution layer, all four companies seek to develop their own STRSs. Companies L and C already

possess a large, internal data source in their ERP systems, so their STRS is used to recommend new trip destinations for existing customers. In contrast, company T emphasizes inbound, independent tourism, while company F concentrates on outbound, independent tourism to/in Japan. Consequently, the STRS is designed to identify the travel needs of new customers.

In summary, the smart tourism case study demonstrates how the AI tech-stack model can be applied to a desired AI-enabled application. The evaluation logic can be used in other cases, such as smart CRM. For instance, combined with the synthesized IT/IS framework, the layered structure helps provide a visual description of what is going on regarding each AI-enabled application in smart CRM. This can help executives identify critical issues associated with incumbent IT/IS facilities and the specific layers with which they need to work as well as focus on the relevant capabilities to address the encountered business issues. Moreover, these executives decide between outsourcing and developing solutions in-house as well as the demand of recond AIaaS offerings in smart CRM to reduce specific vendor lock-in concerns.

Conclusion and Future Work

To resolve management and technology challenges in the development and deployment of AI-enabled applications, this study proposed an AI tech-stack model and a synthesized IT/IS framework that integrates the desired AI systems with incumbent IT/IS systems and collaborating DBPs. Three capabilities were proposed to impact the development/deployment of a desired enterprise AI system in terms of making in-house/outsourcing decisions. The study employed the STRS project to conduct four case studies to illustrate and justify how the AI tech-stack model can help executives address these aforementioned challenges.

The contribution of this study is multifold. From a technological perspective, the layered AI tech-stack model reduces the tightly coupled developments of emerging AI technologies and AIaaS offerings, thereby resolving the fully bundled, proprietary

issue and the vendor lock-in concern as well as offering developers the flexibility of choosing AI offerings for efficient implementation. The layered structure also allows management to easily define SLAs and apply fault-tolerance deployment in each layer. The AI team could have open source community support as well as apply up-to-date AIaaS offerings and MLOps/DataOps. The model of on-demand and pay-per-use price for AI resources can be further deployed to reduce lock-in and future migration costs among different vendors.

From a managerial perspective, the proposed model allows managers to examine existing AIaaS offerings in specific layers and provides a broad overview of both in-house development or outsourcing AI projects. The proposed model also provides an unambiguous conceptualization of the layered technology components. This conceptualization provides a shared reference/guideline for formulating assessments and obtaining insights across various organizational units, offering an avenue for joint assessments of current and required AI components by the involved stakeholders to develop and deploy the AI solutions. Moreover, managers can use this conceptualization to identify which layer to assess, evaluate current capabilities in different layers, and suggest the in-house development or outsource external vendors for implementing needed capabilities.

Finally, from an academic perspective, the proposed AI tech-stack model provides a foundation for further research in the areas of SLA, the capability maturity model (CMM), AI readiness, and AI-enabled application lifecycle management frameworks. For instance, the AutoML products (MLOps, DataOps and DevOps) provided through the AI Platform layer and/or the CI/CD/CT management functions offered in the AI Data Pipeline layer help manage the complicated AI pipeline and ensure the needed results in the enterprise. With a rich set of non-lock-in, loosely coupled, and flexible services provided by each layer, the AI tech-stack model can be used as a basis to build AI lifecycle management and address the constant model revisions needed in AI.

Solutions provided in the AI data pipeline layer are expected to access internal and external data sources much easier.

References

1. Bala, R. et al. Magic quadrant for cloud infrastructure and platform services. Gartner (July 27, 2021).
2. BasuMallick, C. Top 10 open source artificial intelligence software in 2021. Spiceworks (February 10, 2022); <https://bit.ly/3klAeZV>.
3. Brock, J.K.U. and Von Wangenheim, F. Demystifying AI: What digital transformation leaders can teach you about realistic artificial intelligence. *California Mgmt. Rev.* 61, 4 (2019), 110–134.
4. Brynjolfsson, E. and McAfee, A. What's driving the machine learning explosion? *Harvard Business Rev.* 18, 6 (2017), 3–11.
5. Chui, M. Artificial intelligence, the next digital frontier. *McKinsey and Company Global Institute* 47 (2017), 3–6.
6. Comparing machine learning as a service: Amazon, Microsoft Azure, Google Cloud AI, IBM Watson. Altexsoft. (April 25, 2021); <http://bit.ly/3CTIWFL>.
7. Davenport, T.H. and Ronanki, R. Artificial intelligence for the real world. *Harvard Business Rev.* 96, 1 (2018), 108–116.
8. Garrison, G., Kim, S., and Wakefield, R.L. Success factors for deploying cloud computing. *Commun. ACM* 55, 9 (Sept. 2012), 62–68.
9. He, X., Zhao, K., and Chu, X. AutoML: A survey of the state-of-the-art. *Knowledge-Based Systems* 212 (January 5, 2021), 106622.
10. IBM PowerAI Enterprise platform. Armlinsoft; <https://armlinsoft.net/ibm-powerai/>.
11. Krensky, P. et al. Magic quadrant for data science and machine learning platforms. Gartner (2021).
12. Loucks, J. Artificial intelligence: From expert-only to everywhere. Deloitte Insights (December 2018); <http://bit.ly/3GPSQbV>.
13. Machine learning lens. Amazon Web Services; <http://bit.ly/3WINKuw>.
14. Mazalon, L. A guide to 27+ Salesforce Einstein AI products and tools. *Salesforce Ben* (August 4, 2022); <http://bit.ly/3IPvjzV>.
15. Magoulas, R. and Swayer, S. AI adoption in the enterprise 2020. *O'Reilly* (March 18, 2020); <http://bit.ly/3w7Wbyg>.
16. Microsoft artificial intelligence: A platform for all information worker skill set levels. Microsoft U.S. Partner Team (May 1, 2018); <http://bit.ly/3IRybXO>.
17. Pandl, K.D. et al. Drivers and inhibitors for organizations' intention to adopt artificial intelligence as a service. In *Proceedings of the 54th Hawaii Intern. Conf. on System Sciences* (January 2021), 1769.
18. Sculley, D. et al. Hidden technical debt in machine learning systems. *Advances in Neural Information Processing Systems* 28 (January 2015), 2503–2511.
19. Sharma, A. AWS v/s Google v/s Azure: Who will win the cloud war? upGrad (Aug. 23, 2020); <http://bit.ly/3wd0AQG>.
20. Tillman, M.A. and Yen, D.C.C. SNA and OSI: Three strategies for interconnection. *Comm. ACM* 33, 2 (Feb. 1990), 214–224.
21. Top cloud computing platforms for machine learning. GeeksforGeeks (October 14, 2020). <http://bit.ly/3XyiOrm>.
22. Turck, M. Red hot: The 2021 machine learning, AI and data (MAD) landscape. Matt Turck (September 28, 2021); <https://mattturck.com/data2021/>.
23. Tsaih, R.H. et al. Challenges deploying complex technologies in a traditional organization. *Commun. ACM* 58, 8 (Aug. 2015), 70–75.
24. Xu, M. et al. A first look at deep learning apps on smartphones. *WWW '19: The World Wide Web Conference* (May 2019), 2125–2136.

Rua-Huan Tsaih is a professor in the Dept. of Management Information Systems, National Chengchi University, Taiwan.

Hsin-Lu Chang (hlchang@nccu.edu.tw) is a professor in the Dept. of Management Information Systems, National Chengchi University, Taiwan.

Chih-Chun Hsu is a Ph.D. candidate in the Dept. of Management Information Systems, National Chengchi University, Taiwan.

David C. Yen is a professor at the Jesse H. Jones School of Business, Texas Southern University, Houston, Texas, USA.

Copyright held by authors/owners.
Publication rights licensed to ACM.