International Journal of Computer Trends and Technology
ISSN: 2231–2803 / https://doi.org/10.14445/22312803/IJCTT-V72I3P104

Volume 72 Issue 3, 27-32, March 2024

Original Article

Transforming Enterprise Resource Planning Data Migration through Artificial Intelligence

Indrajit Roy Chowdhury[1], Gunjan Goswami[2]

[1]*Technology, Nielsen, Georgia, USA.*
[2]*Program Management, Honeywell, Georgia, USA.*

[1]*Corresponding Author : roy2k5@gmail.com*

Received: 19 January 2024 Revised: 25 February 2024 Accepted: 10 March 2024 Published: 28 March 2024

Abstract *- This article investigated the role of cloud-based Artificial Intelligence (AI) to enhance data migration within enterprise resource planning (ERP) transformation. It highlights the integration of advanced AI tools, including both machine language models and large language models, to improve data migration processes in ERP implementations. The study details how AI facilitates the transformation of both structured and unstructured data, augments data quality assessments, and streamlines mapping and transformation logic. Additionally, it addresses the automation of testing and quality checks by AI during the transformation and loading phases. We performed an in-depth analysis of the necessary technical architecture for AI integration with major standard ERP systems and discussed security and privacy concerns in cloud-driven data migrations. Given the recent developments in standard ERP packages, which now enable easy integration with the cloud, this research also explores ERP professional communities' perception regarding AI's application for data migration purposes.*

Keywords *- ERP, Data Migration, Data Conversion, Artificial Intelligence, Cloud, SAP.*

1. Introduction

By 2030, the Enterprise Resource Planning (ERP) software market is forecasted to increase from USD 71 billion in 2023 to USD 187 billion, growing at a compound annual growth rate of 14.8% [1] (Fortune Business Insights, published by Yahoo Finance 2023). Data migration, a pivotal element in ERP projects, accounts for a substantial portion of the effort involved in such transformations. If data migration accounts for 20% to 25% of the total ERP transformation effort, it will translate to an estimated monetary impact of $37 billion to $46 billion by 2030. ERP data migration includes a series of intricate steps: extracting data from various sources, cleansing it to enhance quality and consistency, adapting it to fit new system requirements, and importing it into the destination ERP system, all while maintaining data integrity, performing comprehensive verification, and testing. The adoption of Artificial Intelligence (AI) in this context can be groundbreaking. The increasing integration of AI and Machine Learning (ML) in this domain is revolutionizing the process, significantly improving efficiency and accuracy.

2. Background

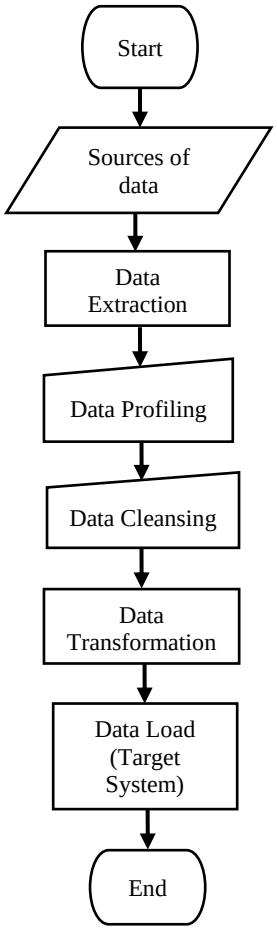
In this section, we discuss the importance, current processes, and challenges associated with the data migration process.

2.1. Importance of data migration in ERP

The role of data migration in enterprise resource planning (ERP) system implementation is critical. It involves the transfer of vital business data from existing systems to the new ERP platform, a process central for maintaining data accuracy and integrity, which are key elements for the ERP system's operational effectiveness and strategic decision-making capabilities. Data discrepancies might introduce significant operational challenges and result in erroneous decisions. The migration process is pivotal for ensuring uninterrupted business continuity because any disruption can cause substantial operational downtime and associated costs. Data migration directly impacts data management quality, including accuracy, elements, accessibility, and performance [2] (Zamzami, Fatani and Zammarah, 2011). Successful data migration is critical for leveraging an ERP system's full potential, including its advanced analytics and automation capabilities, which hinge on the quality and completeness of data. Compliance with data regulations and security during migration is also imperative to prevent data breaches and legal ramifications. This process profoundly affects user confidence and training, with the system's data reliability influencing user adoption. Successful data migration reduces the time and costs related to ERP's implementation, preventing financial losses due to delays or inaccuracies. Moreover, efficient data

migration paves the way for the ERP system's future growth and adaptability, aligning with evolving business needs.

2.2. Contemporary ERP Data Migration Processes



Current data migration tools interface with various source systems, encompassing a range from legacy systems and flat files to databases and other ERP systems. They adeptly manage both structured and unstructured source data. The initial step in the data migration process involves profiling data to evaluate its quality and identifying issues such as inconsistencies, missing values, and duplications. Subsequently, we extract data from the source system into the data migration tool's environment for thorough cleansing, transformation, and standardization. This process is pivotal to ensure the data's accuracy and format compatibility with the target system. The migration tools are versatile and capable of connecting to both traditional and cloud-based ERP systems. Following transformation, it efficiently loads data into a target system in the correct format and can manage large data volumes. It also offers robust reporting features for documentation and auditing, along with comprehensive error handling, debugging, and recovery functionalities. Although these tools are effective for current migration needs, their limitations should be acknowledged.

2.3. Challenges in Traditional Data Migration Systems

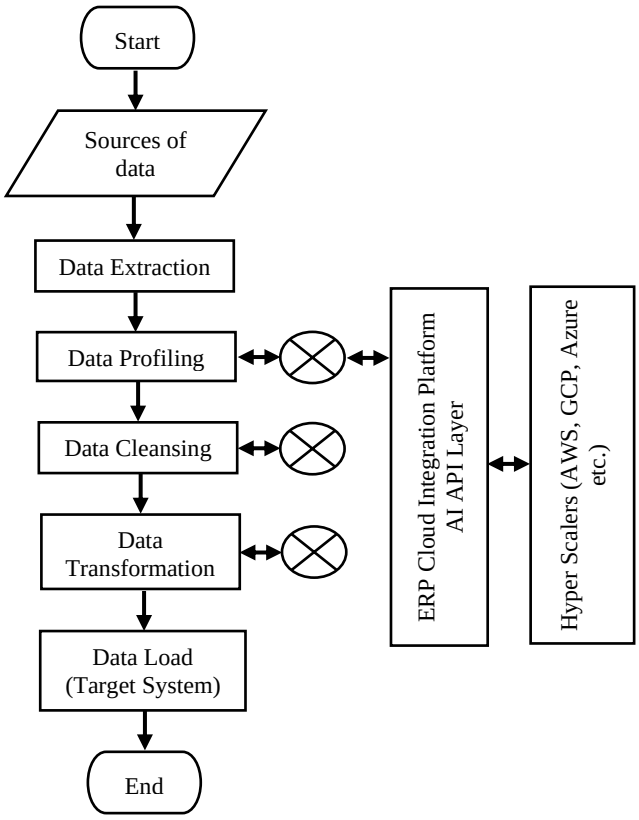
ERP data conversion is a major challenge due to the high complexity and large volumes of data involved, making it a challenging task for software practitioners [3] (Asgar, Akour and King, 2012). Traditional data migration tools play pivotal roles in diverse data integration, transformation, and migration tasks. Despite their strengths, these tools exhibit notable limitations, particularly when juxtaposed with newer AI and ML-enhanced tools. Traditional tools often require extensive manual intervention for data mapping and cleansing, which are both time-consuming and susceptible to human error, especially in complex data landscapes. By contrast, AI-driven tools automate a considerable portion of this process, learning from patterns and offering intelligent suggestions. Ensuring data quality, a crucial aspect of data migration, required substantial manual effort when using non-AI tools. However, AI tools can automatically perform quality checks, identify inconsistencies, and suggest corrections. A significant limitation of non-AI tools is their difficulty in handling unstructured data, such as text, images, or videos. These non-AI tools are primarily designed for structured and semi-structured data, whereas AI-enhanced tools leverage natural language processing and other advanced techniques for analyzing and processing unstructured data. Moreover, the limited learning capability of traditional tools presents a challenge. Unlike AI tools, which learn from past migrations and improve over time, each new migration with traditional tools can be as labor-intensive as the previous ones. Adapting to new or diverse data sources presents more challenges with traditional tools, often necessitating manual adjustments or custom coding. By contrast, AI-driven tools more readily adapt to new data types and sources.

3. Research Problem

In this study, we explored the potential of utilizing cloud-based AI systems to improve data migration processes in ERP systems. Historically, ERP systems have exhibited limited in-process AI capabilities. However, recent advancements have enabled these systems to connect with cloud infrastructures, thereby harnessing the capability of cloud-based AI. This study investigated the extent to which cloud AI can be integrated into ERP systems to optimize and streamline data migration processes, thereby reflecting a significant evolution in ERP technology.

3.1. Enhancing ERP Data Migration with Cloud AI Capabilities

Cloud-based AI can enhance the migration of legacy data and ensure security compliance in cloud ERP implementation [4] (Hustad, Sørheller, Jørgensen, Vassilakopoulou, 2021). Cloud-based AI significantly enhances data migration in ERP, offering diverse benefits that streamline and improve the process. It automates essential steps, such as data extraction, cleansing, and transformation, increasing efficiency and reducing the need for manual intervention.

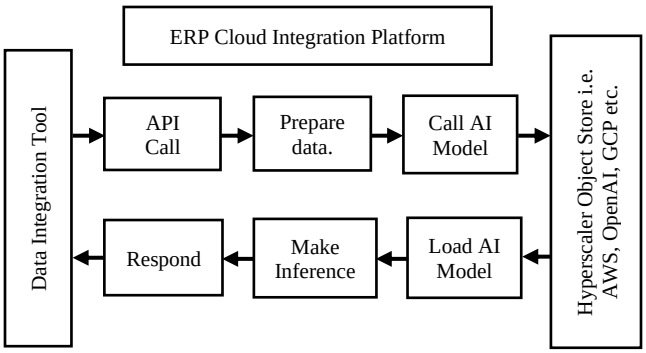


The precision with which AI identifies and corrects data errors ensures a higher degree of accuracy than manual processes. Given the large volumes of data involved in ERP migration, the scalability of cloud-based AI solutions is particularly advantageous. These solutions can be cost-effective because they eliminate the requirement for large upfront investments in hardware and software, operating instead on a pay-as-you-go basis. AI's contribution to data quality involves standardization, deduplication, and validation that lead to more reliable analytics in new ERP systems. The continuous learning capability of AI systems allows for ongoing improvements in efficiency and accuracy, which can be applied to future data-intensive projects. Additionally, AI's predictive analytics are crucial for preventing risks during migration. Cloud-based AI transforms ERP data migration, making it faster, more accurate, and cost-effective while improving data quality and yielding beneficial business insights.

3.2. Connecting ERP to cloud based AI Models

Contemporary market-leading ERP packages incorporate a technology platform that facilitates seamless integration with various cloud services. This integration enables us to use cloud-based applications, thereby addressing the dynamic business needs of organizations. Critical components central to these platforms include application development, AI, data management, and integration. AI and ML are pivotal services within this framework, offering enhancements to applications through intelligent features. They can automate processes and extract insights from data. Additionally, AI's integration significantly streamlines and improves the efficiency and accuracy of data migration while simultaneously reducing manual efforts. Notable examples of such technology platforms include the SAP business technology platform and Oracle cloud infrastructure.

3.3. How ERP Systems Interface with Cloud-Based AI



ERP platforms are equipped with an AI API layer that not only enhances connectivity with internal applications but also enables integration with external cloud-based data and applications. This layer is pivotal for facilitating AI-driven business services, which are monitored through a dedicated separate service. The platform provides a robust infrastructure that supports the productization of AI-based applications and services, allowing for the cost-efficient training and deployment of AI models at scale. These models, particularly suitable for data migration applications, can be developed using open-source machine learning frameworks.

Open-source frameworks are important in this context, offering free and accessible software tools for ML algorithm development. These frameworks encourage community collaboration and knowledge transfer, fostering innovation. The service extends to full lifecycle management support, including content deployment guided by the GitOps principle. This principle utilizes Git's version control, collaboration, and history tracking features to manage and automate applications and infrastructure's deployment effectively.

The platform leverages open-source, container-native workflow engines, such as Argo Workflows, to orchestrate parallel jobs in Kubernetes environments. In the Argo Project, this feature streamlines complex data processing and ML pipelines, allowing the creation of intricate workflows that include dependencies, branching logic, and error handling.

The platform supports Serving, a cloud-native, Kubernetes-based platform that allows the efficient deployment, monitoring, and scaling of ML models in a serverless environment. Integral to the KubeFlow ecosystem,

KFServing streamlines ML workflows on Kubernetes. It simplifies the serving of models using popular frameworks, such as TensorFlow, PyTorch, and scikit-learn, and offers various features, including canary rollouts, autoscaling, and support for multiple frameworks. These attributes significantly improve production-ready models' management, facilitating easier experimentation and faster updates, thereby augmenting the MLOps pipeline.

Such a platform offers a service to manage AI models on an enterprise scale. This service provides a responsive interface to oversee the entire lifecycle of AI models, capturing and analyzing metrics generated by supported AI runtimes. With a focus on the full lifecycle management and operations of AI processes, it provides an overview of all metrics and artifacts, enabling the analysis and evaluation of critical productive AI KPIs. Notable examples of such services include SAP AI Launchpad and Oracle Data Science, provided by SAP and Oracle, respectively, which exemplify their practical applications in the business sector.

3.4. Training an ML Model

ML models are trained using appropriate learning algorithms and training data, which are used to learn and test data to verify correct functioning [5] (Bergen, Chen, Li, 2019). To effectively train an ML model, a dataset should be first supplied to the AI service. This dataset should be stored in a hyperscaler object store, where appropriate access rights are granted to the AI service. The hyperscaler object store represents an optimal solution for storing datasets and models owing to its scalability and container-based pipeline management. The AI artifacts must be methodically organized into two categories: AI core instances and resource groups. The core instances encompass executables, whereas the resource group included training datasets. These executables, along with input datasets, are integral for formulating a training configuration. This configuration is pivotal for training an ML model. A model training process involves the extraction and analysis of results. Upon successful training, when the model demonstrates satisfactory metrics, it is stored in the connected hyperscale object-store. Concurrently, this model is automatically registered within the ERP cloud integration layer, rendering it ready for deployment.

3.5 Deploying the Model

Deploying ML models on Kubernetes using Kube Flow is easy and can be performed as pipelines, with performance, limitations, and features analyzed [6] (Zhou, Yu, Ding, 2020). Upon the completion of its training, the ML model enters the deployment phase, where it is primed to analyze new datasets and generate predictions, a process commonly referred to as model serving or inferencing. Kubernetes clusters offer a flexible environment and exhibit features (e.g., auto-scaling) and the ability to scale to zero, thereby augmenting operational efficiency. The deployment of the model within an ERP cloud platform involves the development of web applications capable of transmitting data. These applications function by exposing an endpoint on the Internet, and their scalability is efficiently managed via the Kubernetes infrastructure. The web application is tasked with processing incoming data requests and retrieving the trained model from the hyperscale object-store. Once a model generates a prediction, the application packages this prediction into a response, which is then utilized by the custom service. This system accommodates both batch processing and ad-hoc calls made through HTTP requests, providing versatile access to models stored in hyperscale object stores. Such a deployment strategy ensures that this model is not only readily accessible but also capable of managing a diverse range of data processing requirements in real time.

3.6. Calling Large Language Models

Modern ERP platforms can interface with Large Language Models (LLMs), a feature that significantly enhances their functionality, particularly in processing unstructured datasets for data migration. These ERP solutions have predefined codes, facilitating the generation of Docker images and templates. Docker images, serving as containers, encapsulate specific applications, reference libraries, datasets, and configurations essential for deploying software solutions. Utilizing these Docker images and templates, NLP capabilities can be integrated into ERP software, thereby expanding its range of applications. A prime example is the SAP Content Package service from SAP AI, which enables integration with Microsoft Azure OpenAI. This platform uses advanced LLMs, such as GPT 3.5 and GPT 4, providing a framework for constructing large language-based applications tailored for migrating unstructured data. The availability of standard packages (e.g., SAP Content Package) offers numerous advantages. It streamlines the incorporation of LLM functionalities into ERP applications through cloud platforms, such as SAP Business Technology Platform (BTP) or Oracle Cloud Interface (OCI). Following the integration and training of LLMs, they become instrumental in analyzing, cleansing, and formatting source data. This preparation is critical for efficient data loading, achieved through strategic prompt engineering.

3.7. Confidentiality and Security of Corporate Data in Cloud Environments

Maintaining corporate data's security and privacy in a cloud environment is a complex process that requires a comprehensive approach. It encompasses various strategies and best practices. A fundamental aspect of this approach is the shared responsibility model, which delineates security obligations between a cloud provider and a client. In this model, the cloud provider manages the cloud infrastructure's security, whereas the client is accountable for protecting the data stored in the cloud. It is imperative to select cloud providers that not only demonstrate strong security measures but are also compliant with security standards, such as GDPR, ISO 27001, and HIPAA. Data encryption is crucial for

safeguarding data both at rest and in transit. Utilizing robust encryption protocols and managing encryption keys securely, preferably via a key management service, is essential. Attribute-based encryption (ABE) is particularly effective in cloud computing for encrypting data. ABE ensures the dual encryption of data, providing added security and sustaining decentralization in multi-authority environments [7] (Cui, Deng, Li, 2017). Privacy and anonymization services, such as K-anonymity, L-diversity, and anatomy algorithms, are crucial for transforming original datasets into modified forms. This transformation enables the preservation of privacy while minimizing information loss, ensuring that published and shared data remain meaningful and private.

Regarding data storage security, employing a group digital signature is a significant cryptographic method that enhances cloud data security. Group digital signature is used to identify anonymization and secure data storage in private clouds, making cloud service delivery more efficient, productive, and agile [8] (Govinda, Ravitheja, 2012). Finally, implementing strict access controls based on the least privileged principle is vital. Access to sensitive data should be limited to authorized personnel, and their access should be confined to what is necessary for their role. The proposed key management protocol for data sharing in cloud computing based on bilinear pairing provides robust access control, ensuring the security and privacy of data stored in the cloud [9] (Zhang, Cui, Mu, 2019). In summary, safeguarding corporate data in cloud environments hinges on a comprehensive approach encompassing shared responsibility, advanced encryption like ABE, privacy-focused algorithms, group digital signatures, and stringent access controls. This multifaceted strategy, underlined by collaborative security efforts and adherence to strict data access protocols, ensures robust protection of sensitive corporate data, complying with prevailing standards and adapting to technological advancements in cloud computing.

4. Results and Discussions

We conducted a study to determine the perceived impact of AI on ERP data migration processes, as assessed by experienced ERP practitioners. The investigation was structured using various critical questions posed during an online survey, which was hosted on a reputable technology research company's website, accessible exclusively to professionals with significant expertise in their respective fields. The survey aimed to address the following queries:

1. The extent to which AI is expected to influence ERP-based data migration processes.
2. Preferences between embedding AI directly within an ERP-based data migration tool versus leveraging it from a cloud-based platform.
3. Areas within ERP data migration wherein AI can have the most substantial impact, including the automation of data cleansing, deduplication, normalization, intelligent data profiling through classification and categorization, and predictive data mapping.
4. The necessity for AI-based data migration systems to be capable of adapting and learning from previous migration experiences.

The responses to this survey provided intriguing insights into the professional community's perspectives on the evolving role of AI in ERP data migration. The study provides varying opinions on how AI integration should be approached, the specific functions where AI could be most beneficial, and the adaptive learning capabilities of AI-driven data migration systems.

- AI's role in transforming ERP data migration processes was affirmed by most respondents, with 87% agreeing on its significant impact. However, this perception varied across industries, with professionals in finance, banking, and insurance being notably more skeptical compared with those in the software industry. Smaller organizations (51 to 200 employees) also showed more reluctance compared with larger companies with over 10,000 employees.
- When asked about the preferred mode of AI integration into ERP-based data migration tools, 85% of respondents advocated for leveraging AI from a cloud-based platform when compared with directly being embedded within the system. This preference reflects the perceived benefits of cloud platforms, including flexibility and scalability.
- When identifying where AI would be most impactful in ERP-based data migration, 59% of participants saw intelligent data profiling as the primary area, followed by automated data cleansing (21%) and predictive data mapping (18%). This suggests a high value placed on AI's ability to classify and categorize data effectively.
- The final question regarding AI's capability to adapt and learn from past migrations received varied responses, with 72% being in favor. Notably, respondents from different industries and organizational roles exhibited diverse opinions, indicating differing levels of acceptance and skepticism toward AI's adaptability in data migration.

The survey results reveal an optimistic outlook on AI's role in ERP-based data migration, highlighting its potential in data profiling and predictive mapping. However, perceptions are influenced by industry-specific challenges and organizational size, reflecting the varying degrees of AI readiness and exposure across sectors.

5. Conclusion

Given the ERP market's substantial growth and the crucial role of data migration in ERP projects, the increasing incorporation of AI and ML is revolutionizing this domain, significantly enhancing efficiency and accuracy. The

importance of data migration in ERP is further emphasized by its critical role in sustaining data integrity and accuracy, which is essential for effective operations and strategic decision-making. The survey conducted among ERP practitioners sheds light on the effect of AI in this process, revealing that a majority believe that AI will significantly influence ERP data migration. However, perceptions differ across industries and organizational sizes, with some sectors and smaller entities expressing more skepticism about AI's effectiveness compared with larger organizations. The preference for leveraging AI from a cloud-based platform, as indicated by most respondents, reflects the advantages of cloud platforms in their scalability, flexibility, and cost-effectiveness. The survey also indicates that the most impactful areas for AI in ERP data migration are intelligent data profiling, predictive data mapping, and automated data cleansing. The notion that AI-based data migration systems should adapt and learn from past experiences is prevalent; opinions vary based on industry and job roles. In conclusion, the study underscores the growing optimism regarding AI's role in ERP-based data migration, highlighting its potential for significant contributions in areas such as data profiling and predictive mapping. However, the views are affected by factors such as industry type and organizational size, reflecting varying levels of readiness and exposure to AI technologies.

References

[1] Global ERP Software Market Size, Share and Growth Analysis [2030] | 14.8% CAGR, Yahoo Finance, 2023. [Online]. Available: https://finance.yahoo.com/news/global-erp-software-market-size-123000998.html

[2] Ikhlas Fuad Zamzami, Hanan Abdullah A. Fatani, and Nuha Abdullah H. Zammarah, "Data Migration Challenges: The Impact of Data Quality — Case Study of University Putra Malaysia UPM," *2011 International Conference on Research and Innovation in Information Systems*, pp. 1-5, 2011. [CrossRef] [Google Scholar] [Publisher Link]

[3] Talukdar S. Asgar, Mohammed Akour, and Tariq M. King, "Applying Test-First and Parallel Processing Techniques to ERP Data Conversion," *2012 Ninth International Conference on Information Technology - New Generations*, pp. 269-274, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[4] Eli Hustad et al., "Moving Enterprise Resource Planning (ERP) Systems to the Cloud: The Challenge of Infrastructural Embeddedness," *International Journal of Information Systems and Project Management*, vol. 8, no. 1, pp. 5-20, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[5] Karianne J. Bergen, Ting Chen, and Zefeng Li, "Preface to the Focus Section on Machine Learning in Seismology," *Seismological Research Letters*, vol. 90, no. 2A, pp. 477-480, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[6] Yue Zhou, Yue Yu, and Bo Ding, "Towards MLOps: A Case Study of ML Pipeline Platform," *2020 International Conference on Artificial Intelligence and Computer Engineering (ICAICE)*, pp. 494-500, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[7] Hui Cui, et al., "Attribute-Based Storage Supporting Secure Deduplication of Encrypted Data in Cloud," *IEEE Transactions on Big Data*, vol. 5, no. 3, pp. 330-342, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[8] Kannayaram Govinda, and Perla Ravitheja, "Identity Anonymization and Secure Data Storage Using Group Signature in Private Cloud," *Proceedings of the International Conference on Advances in Computing, Communications and Informatics*, pp. 129-132, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[9] Leyou Zhang, Yilei Cui, and Yi Mu, "Improving Security and Privacy Attribute Based Data Sharing in Cloud Computing," *IEEE Systems Journal*, vol. 14, no. 1, pp. 387-397, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[10] Get Actionable Advice and Clear Direction Within 24 Hours, Gartner Peer Community. [Online]. Available: https://www.gartner.com/peer-community

[11] The Future of Learning at SAP, SAP Learning Hub. [Online]. Available: https://learninghub.sap.com