



# Innovative Banking with Generative AI

Exploring Use Cases and  
Value Creation



# Contents

## Innovative Banking with Generative AI: Exploring Use Cases and Value Creation

Introduction: The Dawn of Generative AI in Banking	3
1. The Impact of Generative AI on Banking	4
2. Collaborative Models and Strategic Utilization of Generative AI in Banking	5
3. Identifying Use Cases for the Banking Industry	8
4. Evaluating the Business Impact of Generative AI (Post-Implementation)	11
5. Case Studies of Generative AI Utilization in Major Banks and Vendor Examples	13
6. Insights and Recommendations for Bank Executives	17



## Introduction

# The Dawn of Generative AI in Banking

Banking is a mature market built on trust and personal relationships, where face-to-face interactions remain crucial. However, advancements in internet and digital technologies have accelerated the shift towards digital banking. The COVID-19 pandemic, in particular, further propelled the adoption of digital banking services.

Within this evolving landscape, the rise of AI technology is having a profound impact on the banking industry. A significant number 77% of banking executives believe that extracting value from AI is key to competitive advantage.<sup>\*1</sup> Recent years have witnessed the emergence of generative AI, powered by large language models (LLMs), as a particularly promising development. Generative AI holds the potential to redefine operational efficiency, customer experience, and decision-making processes. In fact, approximately 67% of banking leaders anticipate that generative AI will fundamentally change the way they do business.<sup>\*2</sup>

Currently, the application of generative AI in banking is primarily focused on efficiency gains, such as chatbots and employee assistance, with limited adoption in core banking functions. This limited adoption is partly attributed to the industry's low tolerance for errors caused by AI hallucinations, as well as stringent regulations regarding privacy protection and concerns about data leaks.<sup>\*3</sup>

This paper aims to provide insights and recommendations for banking executives based on an investigation into the transformative potential of generative AI in banking, its current limitations and challenges, and the adoption patterns and practices of major banks.

\*1 EIU (June 09, 2020) "[Forging new frontiers: advanced technologies will revolutionize banking](#)"

\*2 McKinsey (December 2023) "[Capturing the full value of generative AI in banking](#)"

\*3 The Stack (July 15, 2024) "['We'll move slowly': Wells Fargo CEO reveals bank's cautious approach to Generative AI](#)"; The Banker (January 2024) "[The story of generative AI is 'cautiously optimistic'](#)"

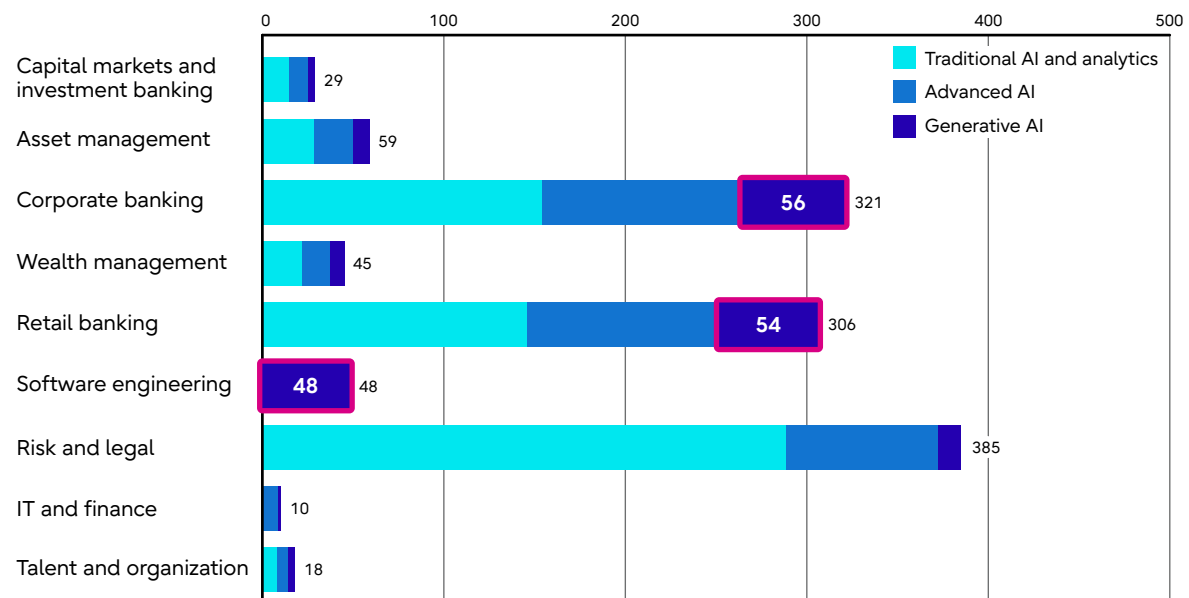
# 1. The Impact of Generative AI on Banking

This chapter delves into the transformative potential of generative AI within the banking industry. According to a 2024 Gartner survey of CEOs and senior business executives, AI emerged as the top transformative theme after digital, with a staggering 59% of respondents citing it as the next frontier. This represents a threefold increase from 2023, significantly surpassing the second-ranked digitalization (5%).<sup>\*4</sup>

The digital technologies mentioned in this survey primarily encompass the internet, smartphones, and cloud computing. Digital transformation (DX) has been a key focus in recent years, involving the use of these technologies to fundamentally revamp business models, processes, and culture. However, the industry is now shifting towards a new era of AI transformation, leveraging AI technologies such as machine learning, deep learning, natural language processing, and generative AI (LLMs). AI transformation similarly aims to revolutionize business models, processes, and culture through the application of these technologies. Notably, 61% of banking technology leaders believe that AI/machine learning including generative AI, will be crucial for achieving short-term business goals.<sup>\*5</sup>

As previously mentioned, banking leaders are highly optimistic about the transformative power and value creation potential of AI. Figure 1 illustrates the anticipated value creation across different banking segments and functions using three types of AI technologies. Generative AI is projected to unlock a potential value of \$200 billion to \$340 billion annually for banks, representing 9% to 15% of their operating profit. The impact of generative AI is expected to benefit all banking segments and functions. For instance, in the corporate banking segment, generative AI can be used to automate loan applications and risk assessments, while in retail banking, it can personalize customer interactions and provide tailored financial advice.

**Figure 1 Value Created by AI by Banking Industry Segment and Function (billions of dollars)**



Source: McKinsey (December 2023) "[Capturing the full value of generative AI in banking](#)"

\*4 Gartner (June 2024) "[How Your CEO Is Thinking About AI](#)"

\*5 KPMG (March 2024) "[Frontiers in Finance](#)"

The diverse spectrum of AI technologies, as depicted in Figure 1, comprises traditional AI, advanced AI, and generative AI, each generating different levels of value. Traditional AI primarily focuses on data analysis using rule-based systems and basic algorithms. Advanced AI leverages neural networks to employ more complex and sophisticated algorithms, enabling the resolution of intricate problems.<sup>\*6</sup> In contrast, generative AI utilizes even more complex neural networks to build models capable of creating new content. While these technologies have progressively revolutionized various aspects of banking operations, generative AI holds the potential for a dramatic industry-wide transformation.

Although traditional AI has significantly contributed to enhancing efficiency, customer experience, and decision-making in banking, it faces limitations when handling unstructured data, natural language comprehension, and complex contextual analysis.<sup>\*7</sup> Generative AI addresses these limitations by providing a more flexible and human-like AI system capable of understanding and generating natural language text while considering context. In banking, this translates to streamlining customer support, automating report generation, conducting sentiment analysis on unstructured text data, and even generating personalized financial advice based on customer interactions and preferences.

While generative AI has the potential to dramatically improve the efficiency of banking business models, its ability to enhance customer experience and drive growth through differentiation is arguably its most compelling aspect. Accenture estimates that early adopters of generative AI will experience a 20% to 30% productivity increase, a 6% rise in revenue growth rate, and a 3% increase in return on equity over the next three years.<sup>\*8</sup>

## 2. Collaborative Models and Strategic Utilization of Generative AI in Banking

The banking industry has actively embraced digital technologies, evolving from online banking services leveraging the internet to mobile banking solutions for smartphones, and real-time interactions using cloud technology. Witnessing the rapid rise of generative AI, the banking sector is now swiftly and cautiously exploring how to integrate this new disruptive technology into their business operations.

### Collaborative Models of Humans and Generative AI: Chatbots, Copilots, and Agents

Generative AI, centered around large language models (LLMs), is the focus of this paper. It envisions a society where humans and generative AI collaborate, enhancing the productivity and innovative capabilities of bank employees. Generative AI can play various roles, from back-office and middle-office functions to front-office roles such as tellers, asset management advisors, relationship managers, call center agents, and even middle managers and executives.

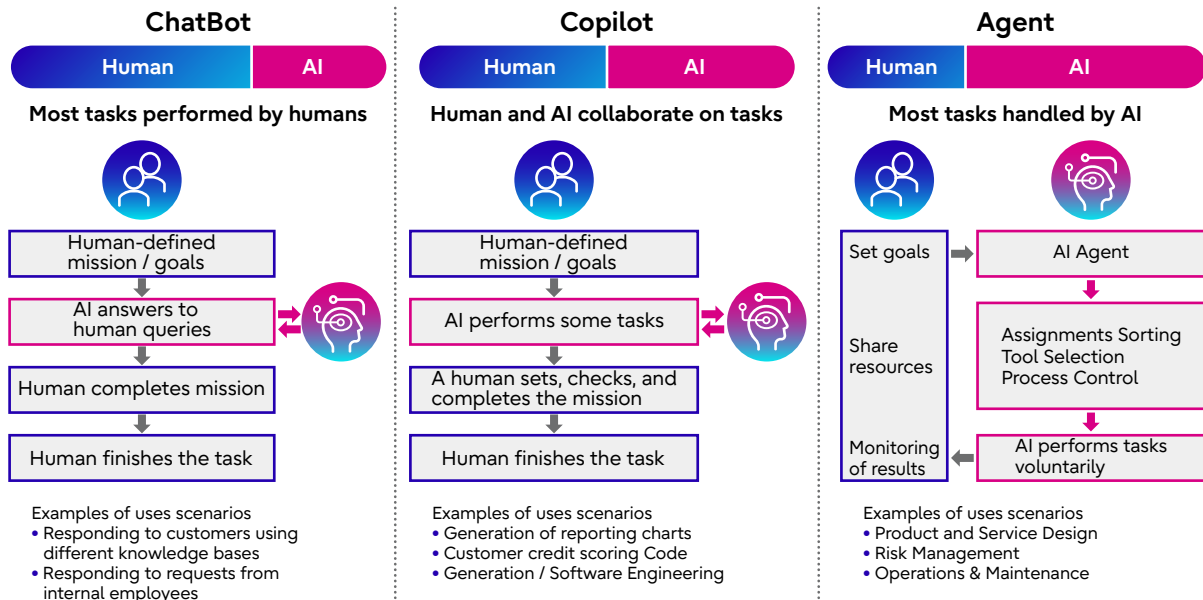
\*6 AI is generally discussed in terms of a dichotomy between traditional AI, including advanced AI, and generative AI.

\*7 Nadejda Alkhalidi (November 2023) "[Generative AI vs. AI: choosing the right technology to move your business forward](#)"

\*8 Accenture (January 2024) "[The age of AI: Banking's new reality](#)"

Figure 2 illustrates the conceptual model of human and machine (generative AI) collaboration.<sup>\*9</sup> The division of roles between humans and machines (generative AI) varies based on factors such as the maturity of the technology, risk aversion potential, and cost/performance considerations.

**Figure 2 Three models of how humans and generative AI work together**



Source: Compiled by the author from KPMG in China (June 2024) and other sources.

## Utilization Methods of Generative AI: Function Expansion, Content Integration, Software Transformation, and Product/Experience Differentiation

Generative AI can be applied in various scenarios within the language and data-intensive banking industry. From the perspective of creating business value, its utilization methods can be classified into four categories:

### (1) Expansion of Existing/New Solutions

This involves leveraging commercial solutions integrated with generative AI or incorporating it into in-house developed solutions. Notable examples include Microsoft 365 and Salesforce's CRM assistant (Einstein). Applications like smart chatbots, language translation, document editing, and search/summarization functions can enhance productivity, efficiency, and employee satisfaction within banks.

### (2) Content Integration

This method involves summarizing and analyzing vast amounts of information held by banks to derive insights. It includes transcribing and summarizing customer call recordings, extracting information from RAGs<sup>\*10</sup> and research reports, anomaly detection, and customer sentiment analysis. This contributes to improved productivity and risk management in banks.

\*9 KPMG in China (June 2024) "[A Preliminary Study on the Application Value of Large Models and Generative AI in the Banking](#)", etc.

\*10 RAG (Retrieval Augmented Generation): A mechanism for acquiring and using knowledge outside of learning in LLM

### **(3) Software Transformation**

Generative AI can convert text and natural language into code in various programming languages. Code assistants, using natural language prompts to update legacy code to new languages or enable low-code/no-code business solutions, are promising tools. This enhances corporate productivity and software developer satisfaction.

### **(4) Product/Experience Differentiation**

This involves innovating and differentiating products, marketing, and customer interactions using generative AI. It is expected to significantly contribute to bank revenues by improving customer satisfaction, retention rates, and share of customer spending. Specific examples include developing products and services tailored to individual customers, creating customer experiences combined with behavioral economics, and offering customized pricing.

## **Challenges and Limitations of Utilizing Generative AI in Banking**

While banking leaders are captivated by the potential of generative AI, they approach its adoption with caution. The following points highlight the reasons and barriers:

### **(1) Data-Related Issues**

LLMs, the core of generative AI, are trained on vast amounts of unlabeled data, leading to issues such as data bias, ownership, and confidentiality protection. Banks need to filter out inappropriate data and content.

### **(2) Explainability and Reliability of Results**

Generative AI relies on neural networks with tens of billions to hundreds of billions of parameters, making it difficult to explain how specific answers are generated. Additionally, it may produce different responses to the same prompt, making it challenging for users to assess the accuracy and reliability of the output. In banking, errors can lead to significant losses and reputational damage, necessitating caution.\*<sup>11</sup>

### **(3) Security and Regulatory Uncertainty**

As generative AI evolves, there are concerns that fraudsters will also advance their techniques using generative AI. Moreover, due to the recent emergence of generative AI, regulations and rules are not yet well-established. Banks, subject to stringent consumer protection and system regulations, need to balance regulatory risks.

Sooner or later, banks will need to redesign their risk and model governance frameworks and devise new governance measures in response to the adoption of generative AI.

\*<sup>11</sup> Forbes (April 22, 2024) "[Wells Fargo Senior EVP Ather Williams On Digital And AI In Banking](#)"

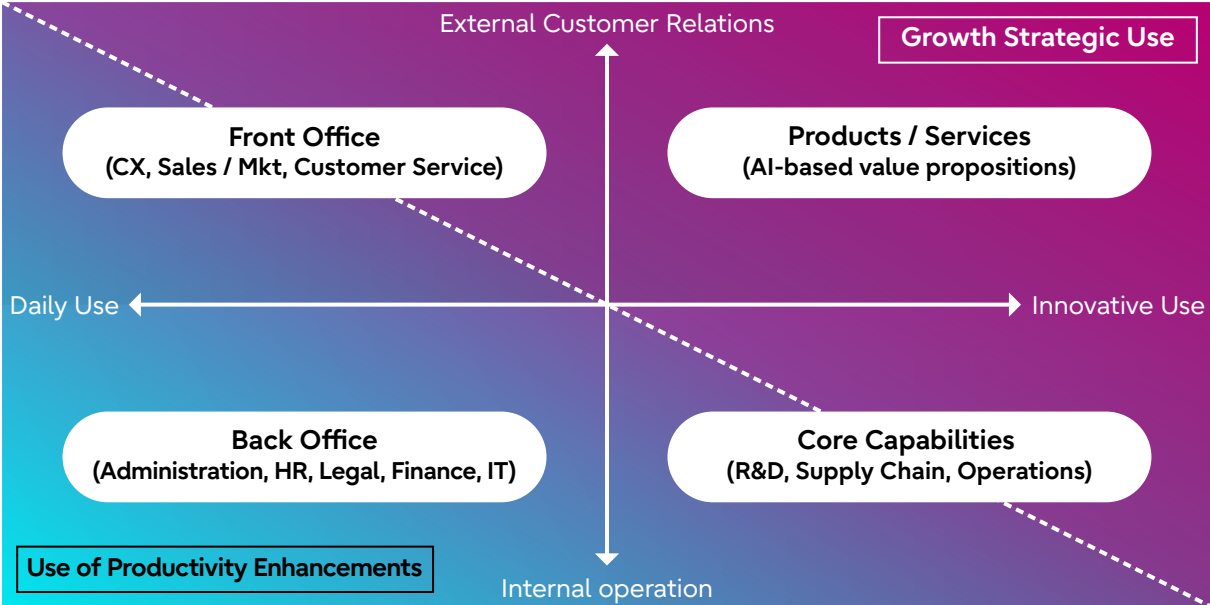
# 3. Identifying Use Cases for the Banking Industry

While thousands of potential AI use cases in banking operations are often cited, several hundred promising use cases with relatively high business impact have been identified through various research studies and the practices of major banks worldwide.<sup>\*12</sup> These promising use cases span the entire banking supply chain, including enterprise applications, risk and compliance, operations (maintenance and management), technology and data, customer engagement, sales and marketing, and products/services. However, it is impossible for any bank to consider hundreds of use cases at once. The major challenge is to identify which of the promising use cases should be prioritized for adoption.

## Strategic Plans for Implementing Generative AI and Framework for Selecting Use Cases

To fully harness the potential of generative AI, banking leaders must see this disruptive technology not only as a means for cost reduction and productivity enhancement but also as a transformative tool for growth. However, specific strategic plans for adopting generative AI must balance risk mitigation, including the reliability of generative AI and regulatory uncertainties. Figure 3 categorizes AI utilization in business operations into three axes: "Daily Use" vs. "Innovative Use," "Internal operation" vs. "External Customer Relations," and adds a third perspective: "Use of Productivity Enhancements" vs. "Growth Strategic Use."

Figure 3 Business AI Application Scenarios



Source: Prepared by author based on McKinsey (2024), et al.

\*12 Accenture (January 2024) "[The age of AI: Banking's new reality](#)"



Based on these three axes, companies' generative AI utilization can be classified into three types:

### (1) Productivity-Focused Companies

These companies primarily use generative AI for automating front-office and back-office operations to pursue productivity. They proceed cautiously when using generative AI for core functions, mainly growth strategy functions.

### (2) Companies Avoiding Customer-Facing Use

These companies focus on internal use of generative AI and avoid using it in customer-facing front-office operations.

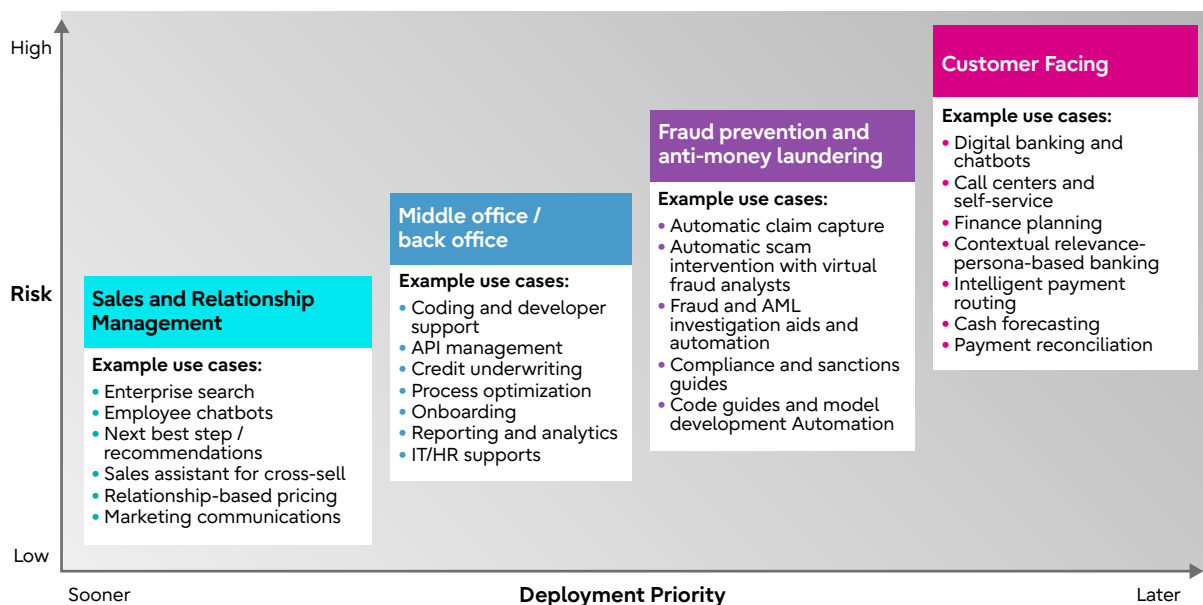
### (3) AI-First/AI-Native Companies

These companies aim to integrate generative AI across the entire value chain, broadly pursuing AI opportunities throughout the organization.

## Identifying Priority Use Cases for Your Company

As of August 2023, a survey by KPMG revealed that 76% of banking executives plan to prioritize the use of generative AI for fraud prevention, 68% for compliance and risk management, and 62% for customer service and personalization.<sup>\*13</sup> In contrast, many banks tend to start with simple, low risk use cases and gradually move to more complex ones as the technology matures (see Figure 4). Most banks are cautiously adopting generative AI, primarily conducting internal experiments in the middle and back offices, and building PoCs (proof of concepts) or pilots for customer-facing use cases. This cautious approach is likely due to the potential significant impact on decision-making for both customers and the bank itself, even if some tasks, such as preprocessing, are considered low risk. Benchmarking best practices from leading companies is crucial.

**Figure 4 Correlation between Use Case Selection Priority and Risk, as Seen from Bank Practices**



Source: Prepared by author based on Datos (December 2023), etc.

\*13 KPMG (March 2024) "[Frontiers in Finance](#)"

When identifying the optimal use cases for their organization, banks need to simultaneously evaluate technical feasibility and business impact.

Table 1 presents a streamlined set of evaluation criteria for identifying use cases, focusing on both business impact and technology feasibility. Companies may assign weights and scores to these criteria, prioritizing use cases that demonstrate strong performance across both dimensions.

**Table 1 Criteria for Determining Business Impact and Technical Feasibility**

	Category	Criteria
<b>Business Impact</b>	Value creation	Can the value be quantified? Is the performance incremental or stepwise?
	Alignment with strategy	How well does this align with or support the company's key strategic goals?
	Adoption difficulty	Are end-users enthusiastic about adopting the solution? Is their demand for more features and performance?
	Business readiness	Is it the right time to implement this solution, considering ongoing transformations and other projects?
<b>Technical Feasibility</b>	Data readiness	Is the data readily available, or does it need to be created or synthesized? Are there special considerations for handling confidential data?
	Solution readiness	Does the solution require proven technology or new technology?
	Scalability	Can the proposed business model sustain an increase in the number of users and cloud instances?
	Reusability	Can the solution components be reused for other use cases?

Source: Created by the author based on McKinsey (May 2024) "[Moving past gen AI's honeymoon phase: Seven hard truths for CIOs to get from pilot to scale](#)"

## 4. Evaluating the Business Impact of Generative AI (Post-Implementation)

Turning the potential of generative AI into tangible business value is far more challenging than conducting proof of concepts (PoCs) or pilot projects that assess technical feasibility and business impact. One reason for the cautious expansion of generative AI in the banking sector is not only the aforementioned risk considerations but also concerns about its cost-effectiveness and revenue contribution. For instance, generative AI is a high-cost solution, and there are doubts about its ability to deliver the expected cost-effectiveness and return on investment (ROI), raising concerns about a potential generative AI bubble.<sup>\*14</sup> Consequently, discussions are emerging about entering the second wave of generative AI in 2024, transitioning from the honeymoon phase to scaling up from pilots.<sup>\*15</sup> Some banks are looking for easier and cheaper ways to deploy generative AI.<sup>\*16</sup>

### Groups of Use Cases and Return on Investment

However, the time to create net value and turn ROI positive varies depending on the implementation cost, complexity, and model-related risks of the use cases. Gartner categorizes generative AI use cases into three types, explaining their approach characteristics and the expected time to value creation.<sup>\*17</sup> Similarly, McKinsey classifies generative AI use cases into three approaches: 'Take' (rapid access to commercial services), 'Shaper' (customizing on top of external foundational models), and 'Maker' (building proprietary models from scratch), providing insights into each.<sup>\*18</sup> The author also focuses on the choice of LLM as the foundation of the application and summarizes three options (using a commercial LLM (via API), customizing an existing open source LLM, or developing the LLM in-house).<sup>\*19</sup>

#### (1) Use Cases with Rapid Success

By leveraging commercial applications, these use cases are easy to start, test, and purchase, focusing on potential productivity improvements with a short time to value realization. However, productivity improvements alone may not differentiate from competitors, potentially reducing the source of differentiation over time. This corresponds to McKinsey's "Take" approach.

#### (2) Differentiating Use Cases Creating Competitive Advantage

These use cases leverage generative AI to differentiate using proprietary data, creating competitive advantages. The time to value realization is moderate, offering differentiation and some competitive edge. However, cost predictability is challenging, and self-management of potential risks is required. This corresponds to McKinsey's "Shaper" approach.

<sup>\*14</sup> Goldman Sachs (June 25, 2024) "[Gen AI: too much spend, too little benefit?](#)"

<sup>\*15</sup> McKinsey (May 2024) "[Moving past gen AI's honeymoon phase: Seven hard truths for CIOs to get from pilot to scale](#)"

<sup>\*16</sup> Isabelle Bousquette (July 17, 2024) "[These AI Models Are Pretty Mid. That's Why Companies Love Them: Companies are looking for simpler and cheaper ways to deploy artificial intelligence](#)"

<sup>\*17</sup> Gartner (August 2023) "[Take This View to Assess ROI for Generative AI](#)"

<sup>\*18</sup> McKinsey (July 2023) "[Technology's generational moment with generative AI: A CIO and CTO guide](#)"

<sup>\*19</sup> Jianmin Jin (February 2024) "[Leveraging the LLM: Strategy from Model Selection to Optimization –Insight for top management–](#)"

### **(3) Transformative Use Cases**

These use cases focus on developing new generative AI models or tools that have the potential to transform business models and markets. They are the most ambitious, characterized by high costs, complexity, and risks, with a long time to value realization. Initially, they aim for strategic benefits that are difficult to quantify. This corresponds to McKinsey's "Maker" approach.

However, the analytical methods of Gartner and McKinsey can be considered use case-level approaches. Companies, especially large enterprises, have multiple business categories and need to address multiple use cases when leveraging generative AI. While evaluating value creation at the individual use case level is necessary, a framework for assessing the impact at the overall corporate level is also required.

## **KPIs for Evaluating Use Case Success**

The journey of value creation through generative AI includes setting business goals, selecting technology and assessing risks, selecting and implementing use cases, monitoring operations, evaluating success, and scaling up. To evaluate the success of generative AI (post-implementation), it is necessary to conduct both qualitative and quantitative assessments using pre-established KPIs. While the KPIs set by individual banks will vary according to their specific management goals and strategies, the following items are generally emphasized:<sup>\*20</sup>

### **(1) Business Growth KPIs**

These include price increases, cross-selling opportunities, demand forecasting, and monetization of new assets.

### **(2) Customer Success KPIs**

These include retention metrics, customer satisfaction metrics, and share of customer spending.

### **(3) Cost-Effectiveness/Productivity KPIs**

These include cost reduction, employee productivity, and asset optimization.

<sup>\*20</sup> Gartner (2024) "[Building a Value-Driving AI Strategy for Your Business](#)"

## 5. Case Studies of Generative AI Utilization in Major Banks and Vendor Examples

As previously mentioned, given the significant potential impact of generative AI on banking operations, bank executives are starting to learn about and integrate generative AI technologies to enhance operational efficiency and improve customer experiences. Meanwhile, vendors supporting enterprise adoption of generative AI are striving to innovate and design architectures to address challenges hindering customer adoption. Below, we introduce case studies from major banks and vendors worldwide to examine their actual implementations.

### Case Studies from Major Banks

#### (1) Wells Fargo: Aiming to Become a “Digital-First” Bank

Wells Fargo is a global banking group with \$115.3 billion in revenue, around 70 million customer accounts, and 226,000 employees. In 2021, it declared its goal to become a “digital-first” bank and is an advanced bank in AI utilization with around 30 million mobile banking users.

Wells Fargo has actively deployed traditional AI, implementing numerous use cases such as fraud detection and credit assessment. Regarding generative AI, the bank prioritizes internal productivity improvements in the short term, considering the accuracy achieved and the tolerance for risks like hallucinations. However, it moves very cautiously in areas affecting consumers to ensure a thorough understanding of the impact. Nonetheless, the bank is committed to investing in technology validation (PoC).<sup>\*21</sup> Based on internal PoC results, Wells Fargo plans to swiftly implement generative AI, considering factors like hallucination prevention, RAG technology evolution, and regulatory guidelines.

As an example of generative AI implementation, Wells Fargo introduced the generative AI virtual assistant “Fargo” (powered by Google’s PaLM2 LLM) in March 2023.<sup>\*22</sup> Its features include handling everyday inquiries, checking credit scores, opening accounts, stopping payments, and reporting fraud. Wells Fargo announced that within a year of its introduction, Fargo had reached 15 million users and handled 117 million interactions. Additionally, in October 2023, the bank introduced the financial management solution “LifeSync,” leveraging LLMs. This solution is expected to expand to all 70 million customers, providing insights and financial advice through Fargo to help users achieve their goals more effectively. Furthermore, Wells Fargo is adopting open-source LLMs, including Meta’s Llama 2 model, for internal applications such as content generation for reports and drafts. The bank is also actively considering automating customer service with solutions like automated call centers.

<sup>\*21</sup> [Wells Fargo \(WFC\) Q2 2024 Earnings Call Transcript](#) (Jul 12, 2024)

<sup>\*22</sup> AIX (July 24, 2024) “[Case Study: Strategic AI Implementation at Wells Fargo](#)”,

Tovie AI (February 19, 2024) “[Generative AI in Banking: Real-Life Applications Worldwide in 2024](#)”

## **(2) HSBC: Exercising “Cautious Optimism” with Generative AI**

HSBC is a global banking group with \$134.9 billion in revenue, around 41 million customer accounts, and 221,000 employees. The bank has been utilizing traditional machine learning models (conventional AI technologies) for nearly a decade. HSBC has around 1,000 applications using some form of AI technology.

HSBC holds a “cautious optimism” towards generative AI.<sup>\*23</sup> As a regulated industry, the bank requires a certain level of accuracy and certainty in technology usage. HSBC has hundreds of use cases undergoing proof of concept (PoC) within the bank, with some use cases, such as coding assistants and chatbots, entering the pilot stage.<sup>\*24</sup> However, there have been no official announcements confirming whether these pilot use cases have reached full-scale operation.

In June 2024, HSBC joined Quantexa’s early adopter Lighthouse program, expecting significant productivity improvements within a year of implementation through streamlined analysis and accelerated processes.<sup>\*25</sup> Quantexa’s customers can operationalize generative AI and achieve transformation without substantial investments in infrastructure, tools, or additional skilled resources, thanks to the new Q Assist technology suite. Thus, HSBC is forming an ecosystem with prominent third-party vendors, including fintech companies, rather than solely focusing on internal generative AI initiatives.

## **(3) Bank of America (BofA): Adopting a “Controlled AI” Approach**

Bank of America (BofA) is a global banking group with \$171.9 billion in revenue, around 69 million customer accounts (including about 57 million digital users), and 213,000 employees. The bank has been investing in digital technologies for over a decade, with AI investments being a key focus of its management strategy.<sup>\*26</sup> Currently, over 90% of BofA’s interactions with customers occur through digital channels, and customers are increasingly accustomed to interacting through these channels.

The most significant factor in operational efficiency and customer retention is the AI virtual assistant “Erica,” released in June 2018. As of the end of 2023, Erica has reached 42 million users and surpassed 2 billion interactions since its release.<sup>\*27</sup> Erica is an application leveraging natural language processing (NLP) and predictive analytics, without incorporating LLM-based generative AI.<sup>\*28</sup> Since its release, Erica has undergone over 50,000 updates (including feature adjustments, expansions, and fine-tuning) to improve performance in responding to customer inquiries and providing insights, maintaining timeliness, and achieving high customer experience.

BofA’s approach is “controlled AI,” carefully considering where to apply generative AI, given the risks of hallucinations and errors. While pilots are being conducted in areas such as customer service, employee efficiency, and coding, there have been no announcements regarding full-scale implementation.<sup>\*29</sup>

\*23 The Banker (January 2024) [“The story of generative AI is ‘cautiously optimistic’”](#)

\*24 HSBC (2024) [“There’s a risk in missing the AI opportunity”](#)

\*25 Quantexa (Jun 10th, 2024) [“Quantexa Debuts Q Assist, New Context Aware Generative AI Technology Suite”](#)

\*26 Bank of America (October 17, 2023) [“Third Quarter 2023 Earnings Announcement”](#)

\*27 Bank of America (April 8, 2024) [“BofA’s Erica Surpasses 2 Billion Interactions, Helping 42 Million Clients Since Launch”](#)

\*28 In other words, Erica is based on a natural language model.

Banking Dive (April 15, 2024) [“BofA evolves AI-powered assistant toward deeper digital integration”](#)

\*29 Aim Research (July 30, 2024) [“How Bank Of America’s Erica Boosted Earnings by 19% and What’s Coming Next”](#)

# Trends from the Case Studies of Three Major Banks

## 1) Cautious Progress with Generative AI

All three banks are cautiously progressing with generative AI, balancing its potential and risk evaluation, and are in the early stages of adoption. However, the pace of practical implementation varies. Wells Fargo is a step ahead, implementing and operating some front-end use cases.

## 2) Collaboration with Leading Vendors

Wells Fargo and HSBC are collaborating with leading model development vendors, platform providers aggregating models and knowledge bases, fintech companies, and major technology firms. BofA, on the other hand, holds numerous AI/machine learning patents related to the financial industry and has development talent, indicating a strategy of self-development and utilization.

## 3) Extensive AI Application

All three banks have implemented and operationalized numerous applications using traditional AI focused on specific tasks. The use of AI technology in banking is proactive, with a wealth of accumulated knowledge. From the success of the virtual assistant “Erica,” BofA is focusing on leveraging NLP models rather than LLM-based generative AI in the short term, aiming to capitalize on its technological resources in NLP.

## 4) Four Methods of Generative AI Utilization

Based on the practices of the three banks, the following four methods of generative AI utilization can be considered.

1. Customer/Employee Engagement (Virtual Assistants): 24/7 chatbots, etc.
2. Content Integration (Virtual Experts): Summarizing and generating reports, utilizing knowledge bases, etc.
3. Software Transformation: Automation in Software Engineering.
4. Use Cases Influencing Customer Decision-Making: Implementation is limited, still in the learning and proof of concept stage.

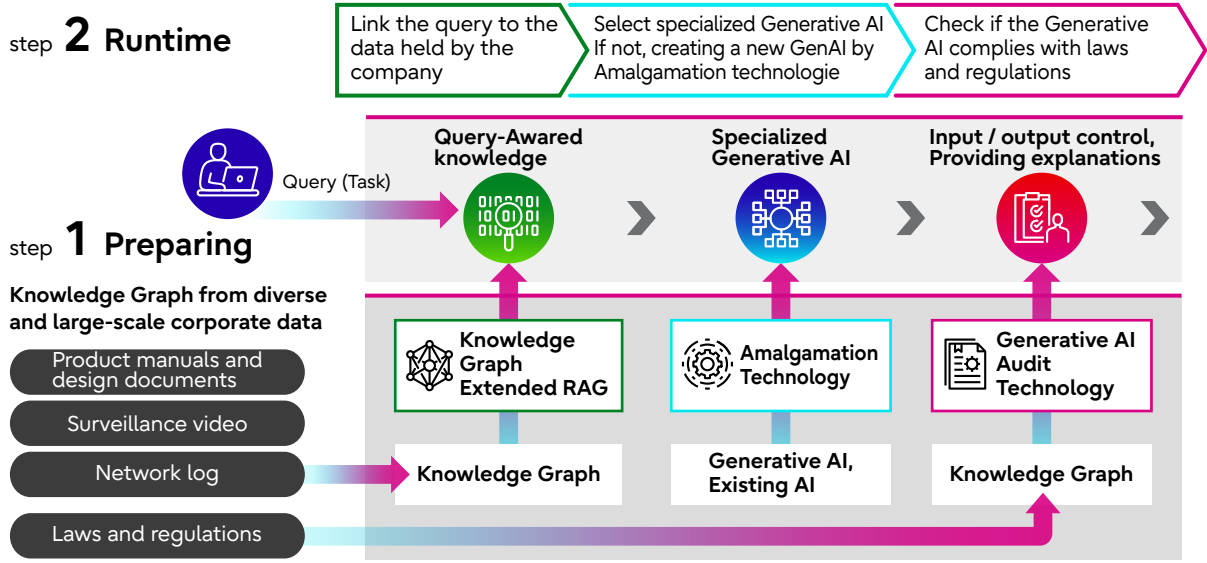
# Vendor Example: Fujitsu's Enterprise Generative AI Framework

In recent years, challenges such as handling large-scale data, meeting diverse requirements like cost and response speed, and complying with corporate regulations and laws have become barriers to the enterprise adoption of generative AI. Additionally, the analytical methods of Gartner and McKinsey and the case studies of major banks have focused on the use case level. To transform the potential of generative AI into value, it is necessary to set up generative AI architectures and develop components across the entire enterprise, rather than focusing solely on individual use cases.

Aiming to become a global top player in driving enterprise generative AI utilization, Fujitsu has begun providing an “Enterprise Generative AI Framework” globally. This framework addresses issues in generative AI utilization within enterprises (banks), alleviates security concerns, and ensures high-reliability outputs.<sup>\*30</sup> As shown in Figure 5, it consists of the following three technologies:

<sup>\*30</sup> Fujitsu Press Release (June 4, 2024) [“Fujitsu to provide the world's first enterprise-wide generative AI framework technology to meet changing needs of companies”](#)

Figure 5 Overview of a generative AI framework for enterprises



Source: [Fujitsu Press Release](#) (June 4, 2024)

- (1) Knowledge Graph-Enhanced RAG: Handling diverse and large-scale enterprise data.
- (2) Generative AI Amalgamation Technology: Flexibly responding to changing enterprise needs, utilizing appropriate AI, including existing AI.
- (3) Generative AI Audit Technology: Alleviating concerns about AI utilization through behavior control and ensuring compliance with rules.

Fujitsu's technology allows user companies to not only use various general-purpose LLMs like ChatGPT, Claude, and Gemini as the main models but also select or combine models that deliver the highest performance based on task content. Additionally, it enables systematic expansion of LLMs with knowledge graphs<sup>\*31</sup> that organize regulatory insights within the organization.

For banks, implementing Fujitsu's Enterprise Generative AI Framework is expected to enhance ROI by maximizing asset efficiency through the combination (synthesis) of already established AI models, strengthen competitiveness through differentiation by fully utilizing proprietary data, and improve output accuracy while reducing or eliminating risks such as leakage of confidential information.

\*31 Knowledge graph: A network of knowledge that systematically represents different types of knowledge in a graph structure



## 6. Insights and Recommendations for Bank Executives

As revealed by the research findings, the expansion of generative AI adoption in the banking industry is hindered by several key issues: concerns about high accuracy and reliability related to hallucinations, data leakage risks and compliance issues, and questions about return on investment. Here, we summarize insights on overcoming these challenges and transforming the potential of generative AI into business value.

---

### Insight and Recommendations 1

#### Ensuring High Accuracy through Hybrid Models

To ensure high accuracy and reliability in generative AI, there are two main approaches: optimizing the model itself and optimizing through external context (e.g., using RAG). By hybridizing the model, balancing the general and specialized capabilities of generative AI, and retraining it with proprietary data, banks can provide highly accurate results and customized solutions. However, excessive retraining may limit the model's "zero-shot learning" ability, generality, and emergent properties. To minimize these drawbacks, strategies such as hybrid approaches (multi-model with base and specialized models), modular approaches (modularizing the model to handle specific tasks with dedicated models), and continuous learning can be effective. This allows banks to maximize the strengths of generative AI while addressing specific business needs.

The hybrid model approach can achieve both 'adaptability' and 'context understanding' for a wide range of tasks and high accuracy and reliability within specialized domains by integrating generative AI with traditional NLP models. As mentioned earlier, Wells Fargo's generative AI virtual assistant "Fargo," leveraging LLMs, and BofA's AI virtual assistant "Erica," utilizing NLP and predictive analytics, have both been successful. Combining generative AI models with NLP models can significantly enhance reliability and user experience.

Fujitsu's "Enterprise Generative AI Framework," which includes generative AI hybrid technology, can automatically select the necessary AI models and generate required models if none are available. This automatically extends the model lake for hybridization, enhancing customer reliability and user experience to new levels.

---

## Insight and Recommendations 2

### **Ensuring Reliability and Accuracy with RAG, and Securing Privacy and Compliance**

RAG (Retrieval-Augmented Generation) ensures that generated responses are accurate and up-to-date by incorporating data from diverse sources (public databases, industry-specific databases, proprietary databases, news and media sources, and real-time information) without altering the model itself. By integrating retrieval into the generative AI process, RAG ensures high reliability and accuracy. RAG models have been reported to reduce hallucination rates by up to 30% compared to traditional generative models. Such improvements are crucial for banking applications, where error tolerance is very low, and factual accuracy is paramount.

Another advantage of RAG is that it allows the use of private data sources without exposing them to external LLMs, ensuring data privacy and compliance. Thus, RAG addresses both the accuracy and data-related risk concerns that banks have regarding generative AI.

In practice, Fujitsu's knowledge graph-enhanced RAG has expanded its ability to analyze over 10 million characters of documents with high precision. The accuracy of complex question answering has reached world-class levels.

---

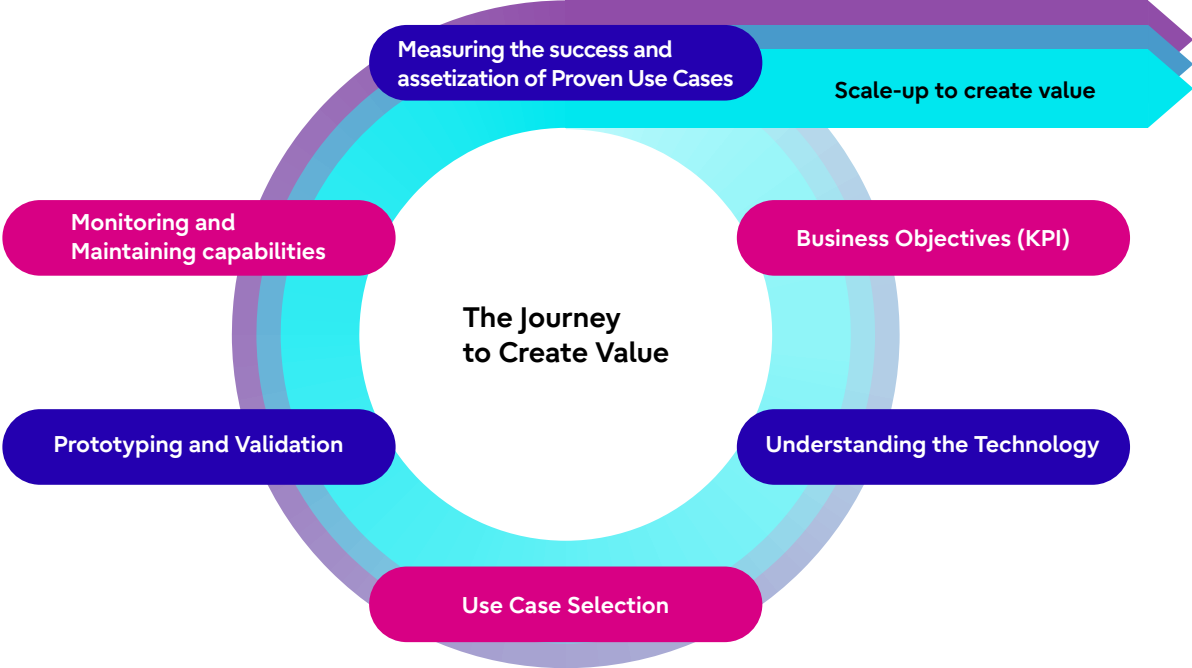
## Insight and Recommendations 3

### **Reducing Costs and Improving ROI through Modularization and Assetization of Common Functions**

As previously mentioned, companies are increasingly concerned about the high costs associated with generative AI. These costs mainly include the development and implementation costs of models and applications and the operational costs, including change management. Operational costs are said to be significantly higher than development and implementation costs. When hundreds or thousands of models are deployed across the enterprise, costs can skyrocket. Therefore, from a management perspective, the generative AI utilization strategy should not be limited to use cases or siloed productivity improvements but should aim for enterprise-level processes and comprehensive outcomes.

By automatically selecting and implementing the optimal generative AI model for individual tasks through multi-modeling, sharing cost-reduction tools like RAG, and improving utilization rates, costs can be reduced. Additionally, in an insight paper released in February 2024, the author suggested that "technologies, know-how, and use cases (such as architectures) proven through repeated practice should be assetized for rapid scaling. At the scaling stage, it is also important to skip PoCs and directly proceed to implementation" (see Figure 6). To achieve this, it is necessary to consolidate and reuse common function modules and assetized modules on a unified enterprise platform.

Figure 6 Unleashing the power of generative AI: Effective steps and scaling



Source: Author Creation

As summarized in these insights, the issues of high accuracy and reliability required in banking, data leakage risks, and compliance issues are being significantly addressed through technological advancements and AI architecture innovations. Additionally, business model initiatives are reducing the construction and operational costs of generative AI, enhancing ROI. As shown in Figure 2, the utilization of generative AI in banking is shifting from human-led processes and decision-making to AI-enhanced processes, where AI, empowered by LLMs, takes the lead, and humans act as overseers and final approvers. This AI transformation in banking is progressing and holds great promise.

# Key References

1. Accenture (January 2024) "[The age of AI: Banking's new reality](#)"
2. Aim Research (July 30, 2024) "[How Bank Of America's Erica Boosted Earnings by 19% and What's Coming Next](#)"
3. Bank of America (October 17, 2023) "[Third Quarter 2023 Earnings Announcement](#)"
4. Bank of America (April 8, 2024) "[BofA's Erica Surpasses 2 Billion Interactions, Helping 42 Million Clients Since Launch](#)"
5. Banking Dive (April 15, 2024) "[BofA evolves AI-powered assistant toward deeper digital integration](#)"
6. Forbes (April 22, 2024) "[Wells Fargo Senior EVP Ather Williams On Digital And AI In Banking](#)"
7. Gartner (2024) "[Building a Value-Driving AI Strategy for Your Business](#)"
8. Gartner (June 2024) "[How Your CEO Is Thinking About AI](#)"
9. Gartner (August 2023) "[Take This View to Assess ROI for Generative AI](#)"
10. Goldman Sachs (June 25, 2024) "[Gen AI: too much spend, too little benefit?](#)"
11. HSBC (2024) "[There's a risk in missing the AI opportunity](#)"
12. KPMG (March 2024) "[Frontiers in Finance](#)"
13. KPMG in China (June 2024) "[A Preliminary Study on the Application Value of Large Models and Generative AI in the Banking](#)"
14. McKinsey (December 2023) "[Capturing the full value of generative AI in banking](#)"
15. McKinsey (May 2024) "[Moving past gen AI's honeymoon phase: Seven hard truths for CIOs to get from pilot to scale](#)"
16. McKinsey (July 2023) "[Technology's generational moment with generative AI: A CIO and CTO guide](#)"
17. The Banker (January 2024) "[The story of generative AI is 'cautiously optimistic'](#)"
18. The Stack (July 15, 2024) "[We'll move slowly": Wells Fargo CEO reveals bank's cautious approach to Generative AI](#)"
19. Vinoth Nageshwaran (August 16, 2024) "[RAG: The Future of Reliable and Accurate Generative AI](#)"
20. [Wells Fargo \(WFC\) Q2 2024 Earnings Call Transcript](#) (Jul 12, 2024)
21. Fujitsu Press Release (June 4, 2024) "[Fujitsu to provide the world's first enterprise-wide generative AI framework technology to meet changing needs of companies](#)"
22. Jianmin Jin (February 2024) "[Leveraging the LLM: Strategy from Model Selection to Optimization –Insight for top management-](#)"

## About the author



### Dr. Jianmin Jin

2020- Fujitsu Ltd., Chief Digital Economist

1998-2020 Fujitsu Research Institute, Senior Fellow

Dr. Jin's research mainly focuses on global economic, digital innovation/ digital transformation, and Dr. Jin has published books such as "Free Trade and Environmental Protection", etc.

Recent writings: the following Fujitsu Insight Paper, etc.

- [Leveraging the LLM: Strategy from Model Selection to Optimization –Insight for top management-](#) (2024)
- [Generative AI: Use Cases as the Pathway to Value Creation](#) (2024)
- [Transformative Quantum Computing: Striving for Greater Heights in Pursuit of Steady Progress](#) (2023)
- [Transforming Supply Chains to Be More Productive, Resilient, and Sustainable](#) (2023)
- [Transformative Enterprise 5G: To Become an Attractive Enabler for DX](#) (2023)

The author would like to thank Keisuke Yano, Shinichi Komeda, Hiroshi Nishikawa, Nick Cowell, Nicolas Sautier, and Takashi Shinden for their review and invaluable advice during the development of this paper. The author would also like to thank Michiyo Hano, Yukiko Sato, and Mitsuo Tsukihara for their unwavering daily support.



© Fujitsu 2024. All rights reserved. Fujitsu and Fujitsu logo are trademarks of Fujitsu Limited registered in many jurisdictions worldwide. Other product, service and company names mentioned herein may be trademarks of Fujitsu or other companies. This document is current as of the initial date of publication and subject to be changed by Fujitsu without notice. This material is provided for information purposes only and Fujitsu assumes no liability related to its use.

October, 2024 v1.0