Simen Owen Academic Proceedings Series

4th International Conference on Media, Economy, Communication and Intelligence Management (MECI 2026)

Review

Artificial Intelligence and Commercial Bank Operational Efficiency in the Fintech Era

Qun Zhang[1,*]

[1] BIC Division, Beijing No. 8 High School Yihai Campus, Beijing, China
* Correspondence: Qun Zhang, BIC Division, Beijing No. 8 High School Yihai Campus, Beijing, China

Abstract: Artificial intelligence can accelerate information processing in commercial banks, but faster individual tasks do not necessarily produce lower operating costs or more efficient financial intermediation. This article develops a critical, integrative review of research on banking efficiency, financial technology, machine learning and generative artificial intelligence. It distinguishes evidence about prediction and workflow performance from estimates of bank-level efficiency and profitability. The synthesis identifies three principal channels: automating document-intensive operations, improving the allocation of screening and monitoring effort, and supporting employees in customer service and knowledge work. Evidence from digital lending supports processing gains, while bank-level studies suggest performance associations that remain sensitive to technology measurement and identification assumptions. Generative artificial intelligence offers additional opportunities, although influential productivity estimates come from settings outside banking. The article proposes an evaluation framework linking application-specific outcomes to full lifecycle costs, service quality and risk-adjusted bank performance. Data governance, complementary organisational investment and appropriate human oversight determine whether local productivity gains persist after implementation, validation and remediation costs. The central implication is that commercial banks should evaluate verified improvements in completed workflows rather than adoption counts or model accuracy alone. Further research requires deployment-level data, credible comparison groups and longer observation of credit and operational risks.

Keywords: artificial intelligence; commercial banks; operational efficiency; financial technology; human augmentation

Received: 09 September 2026
Revised: 18 October 2026
Accepted: 31 October 2026
Published: 05 November 2026

Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Commercial banks transform deposits and other funding into loans, payments and financial services through repeated information-intensive decisions. Verifying an applicant's identity, evaluating repayment capacity, investigating suspicious transactions and answering a customer's question all consume staff time and organisational resources. Financial technology changes how these tasks are performed, and artificial intelligence adds methods for extracting patterns from data and generating usable content. The resulting opportunities concern both the cost of producing banking services and the quality of the decisions underlying them [1,2].

Adoption is already substantial, although its interpretation requires care. The Bank of England and Financial Conduct Authority's 2024 survey received responses from 118 financial firms across several sectors; 75% of respondents reported using artificial intelligence. This figure describes the responding financial firms, rather than commercial banks alone, and does not measure realised cost savings [3]. Banks can deploy many models while still carrying duplicate manual processes, fragmented databases and

substantial monitoring costs. Conversely, a limited application embedded in a well-designed workflow may generate a measurable operational improvement.

The academic evidence is dispersed across literatures that use different units of analysis. Banking studies estimate distance from an efficiency frontier or examine accounting performance. Lending studies investigate processing time, default prediction and access to credit. Organisational research measures worker productivity, while supervisory assessments examine governance and financial stability. Combining these findings without distinguishing their outcomes creates an inference problem: better prediction can be presented as lower bank costs, and a nonbank productivity estimate can be presented as a banking effect. Neither interpretation follows automatically from the original evidence.

This review asks how artificial intelligence can affect commercial bank operational efficiency, what the available evidence establishes, and which organisational conditions determine the net result. Its contribution is an outcome hierarchy connecting model performance, completed workflows and bank-level efficiency. The analysis separates predictive applications from generative applications and incorporates the costs of control and correction into efficiency assessment. It thereby provides a basis for evaluating banking applications without assuming that every technological improvement produces a corresponding financial return.

2. Literature Review and Conceptual Foundations

2.1. Artificial Intelligence within Financial Technology

Financial technology includes digital distribution, payment infrastructure, cloud services and other innovations that need not involve artificial intelligence. Digital disruption also changes competition between incumbent banks, specialised lenders and large technology platforms [4]. An online application portal, an application programming interface and an automated rules engine should therefore not be treated as equivalent evidence of machine learning adoption. Broader technology studies remain relevant to the operating environment, but their findings cannot isolate artificial intelligence's incremental contribution.

Institutional differences also affect comparisons between banks and fintech lenders. Buchak and colleagues attribute the expansion of US shadow mortgage lending to both regulatory and technological factors [5]. A lender's market-share growth or convenience advantage can therefore reflect its business model and regulatory position as well as its production technology. These factors should be considered before transferring a fintech result to a deposit-taking commercial bank.

Predictive artificial intelligence estimates an outcome or classifies an observation. Credit scoring, fraud detection and transaction monitoring are representative applications. Generative artificial intelligence produces text, code or other content from a prompt and contextual information. It can support document summarisation, internal knowledge retrieval and draft responses. These categories describe principal uses rather than mutually exclusive algorithmic families; this review focuses on contemporary machine-learning and generative applications. The distinction matters because prediction quality is commonly evaluated against labelled outcomes, whereas generated content also requires assessment of factual accuracy, completeness and adherence to authorised instructions. Information-processing improvements can support financial intermediation, but the two application types introduce different operating and governance requirements [6].

Banking information is also heterogeneous. Hard information can be stored and transmitted with relatively stable meaning, while soft information depends more strongly on the person and circumstances through which it is acquired [7]. Machine learning can expand the information available for screening, yet historical data may omit an entrepreneur's changing circumstances or the context behind a distressed customer's

request. The operational question is consequently which decisions can be standardised, which require additional evidence, and how exceptions move through the bank.

2.2. Operational Efficiency and Related Outcomes

Operational efficiency concerns the resources required to deliver a defined set of services at an acceptable quality and risk level. Technical efficiency concerns the relationship between inputs and outputs; cost efficiency additionally incorporates input prices. Profit efficiency allows for revenue opportunities and the choice of outputs. These concepts differ from profitability, which also reflects market power, funding conditions and risk-taking. Reviews of financial institution efficiency show that results depend on the selected frontier method and the treatment of banking outputs [8]. Bank-level comparisons must additionally account for business mix, financial capital and asset quality [9].

Data envelopment analysis constructs a frontier from observed input-output combinations, while stochastic frontier analysis separates an assumed inefficiency component from statistical noise [10,11]. Each approach has limitations. A data envelopment analysis score is relative to the comparison sample and sensitive to measurement error. A stochastic frontier result depends on functional and distributional assumptions. Neither technique, by itself, establishes that artificial intelligence caused an efficiency change. An efficient retail lender and an efficient corporate bank may also occupy different production environments.

Accounting ratios remain useful operational indicators when their limitations are explicit. A lower cost-to-income ratio may reflect declining operating expenses, rising interest margins or a change in business composition. Higher return on assets may result from lower funding costs or greater risk exposure. Staff reductions can lower recorded expenditure while increasing waiting times, outsourced costs or unresolved complaints. Efficiency assessment must therefore link financial ratios to comparable service outputs and quality constraints rather than select a single favourable indicator.

Credit losses should be reported alongside operating expenses. Stronger underwriting may improve risk-adjusted performance without lowering production costs, while faster processing can conceal deterioration in loan quality. The framework developed here consequently evaluates operating efficiency subject to risk and service-quality constraints, keeping production costs and broader economic consequences distinguishable.

2.3. Organisational Complements and Delayed Returns

Technology investment produces value through its interaction with business processes and employee capabilities. Research on information technology and organisational transformation identifies complementary investment as a central explanation for differences in realised productivity [12]. Banking technology research likewise connects technological progress with changes in production, service delivery and industry structure [13]. For artificial intelligence, these complements include accessible data, integration with core systems, revised decision rights and staff who can recognise inappropriate outputs.

The productivity J-curve framework explains why measured returns can initially lag investment when organisations build complementary intangible assets [14]. Its application to banking is a theoretical expectation rather than proof that every bank follows the same trajectory. Integration expenditure and parallel processing can increase costs before benefits appear. Later improvements may reflect accumulated organisational changes as well as the model itself. Automation and augmentation also interact: delegating routine tasks changes the expertise employees need to handle the remaining cases [15]. This suggests examining the whole workflow and the timing of complementary investment.

3. Methodology

This article uses a critical integrative review rather than an original econometric study. It combines peer-reviewed research with a limited set of primary supervisory publications available by 12 October 2026. Sources were identified through targeted scholarly searches, publisher records and citation links. Search concepts combined artificial intelligence, machine learning, generative artificial intelligence, financial technology, commercial banks, lending, productivity and efficiency. Foundational efficiency and information-economics studies were retained to define the constructs against which newer applications are assessed. Titles, publication details and digital object identifiers were checked against primary publication records where available.

Selection prioritised relevance to a specified banking task or an explicit efficiency mechanism. Empirical lending papers were assessed for their technology definition, institutional setting, outcome and comparison strategy. Bank-level studies were examined for the relationship between their artificial intelligence proxy and actual deployment. Cross-sector productivity studies were included only as evidence about transferable mechanisms, with their setting identified. Supervisory publications informed adoption context and risk analysis, rather than estimates of causal productivity effects. Vendor forecasts and unverified cost-saving claims were excluded from the evidentiary basis.

The synthesis distinguishes observed associations, estimates based on experimental or quasi-experimental variation, and conceptual implications. It does not pool effect sizes because the studies concern different technologies, populations and outcomes. The search was targeted and non-exhaustive; no systematic screening protocol or comprehensive coverage claim is made. Publicly accessible abstracts and metadata also provide less detail than full research records for some studies. These restrictions motivate cautious interpretation of bank-level causal effects and explicit identification of the evidence supporting each mechanism.

4. Analytical Findings

4.1. Document Processing and Workflow Automation

Commercial banking operations frequently involve moving information between forms, statements and internal systems. Optical character recognition combined with classification or language models can extract relevant fields and route cases for review. Generative systems can summarise a file or prepare a draft explanation. The potential efficiency gain arises when these outputs reduce the time required to complete a verified case. Generating a document faster is insufficient if employees subsequently spend the saved time locating omissions, reconciling inconsistencies or reproducing work in another system.

Evidence from mortgage lending illustrates the value of workflow measurement. Fuster and colleagues find that technology-based mortgage lenders processed applications approximately 20% faster than other lenders without evidence of higher default rates [16]. Their analysis concerns a broader digital lending model, including many nonbank lenders, rather than an isolated artificial intelligence intervention. It supports a processing-efficiency mechanism but cannot determine how much of the improvement came from machine learning, online distribution, staffing or standardisation. For commercial banks, the relevant adaptation is to measure elapsed completion time and staff effort across the entire application process.

Automation also changes the location of bottlenecks. Faster document intake may increase the queue awaiting legal review or exception handling. A bank should therefore distinguish throughput from the duration of individual automated steps. Duplicate checks can remain necessary during validation, but they should be included in operating costs and reviewed when evidence supports a change. The managerial implication is to redesign the sequence of work around a defined service outcome, rather than attach an artificial intelligence component to each existing stage and assume that their local gains will accumulate.

4.2. Screening and the Allocation of Monitoring Effort

Machine learning can improve the use of information in credit assessment. Using deferred-payment transactions at a German e-commerce firm, Berg and colleagues show that digital footprints contain information useful for predicting consumer default, including information complementary to conventional credit bureau scores [17]. Jagtiani and Lemieux similarly examine alternative data and machine learning in fintech credit scoring [18]. These findings support the possibility of screening borrowers with limited conventional histories. They concern particular lending settings, however, and do not establish a universal reduction in commercial bank operating expenses.

Prediction becomes an efficiency mechanism when it changes resource allocation. A more informative score may allow staff to concentrate investigation on cases for which additional evidence has value. It may also support more consistent monitoring after origination. The benefit depends on the costs of acquiring the data, maintaining the model and acting on its predictions. A small increase in predictive accuracy can have little practical value if the lending decision remains unchanged. Conversely, an unchanged average accuracy can conceal improvements in a costly segment of the portfolio.

Evidence on credit scoring also highlights the importance of benchmarks and economic conditions. Using data from a Chinese fintech firm, Gambacorta and colleagues find predictive benefits from machine learning and nontraditional data during a negative credit-supply shock [19]. A bank must accordingly test calibration and decision quality outside its model-development sample, including periods with changing employment, interest rates or borrower behaviour. Fraud and anti-money-laundering applications require a related distinction between detecting suspicious activity and creating an excessive investigation workload. Better ranking of alerts could reduce avoidable reviews, but missed cases and the costs of investigation must enter the same assessment. Supervisory publications identify these applications as important use cases, rather than providing a general causal estimate of savings [3,20].

Longer-run borrower outcomes can qualify an initially favourable interpretation. Di Maggio and Yao find higher subsequent default and indebtedness among fintech personal-credit borrowers in their comparisons [21]. Borrower selection and lending incentives prevent attributing these differences to artificial intelligence itself. The study nevertheless shows why rapid origination and easier access should be evaluated alongside subsequent repayment and debt outcomes, with product-specific evidence rather than a single narrative about digital lending.

4.3. Employee Augmentation and Customer Service

Generative artificial intelligence can help employees retrieve procedures, summarise interactions and prepare responses. The strongest productivity estimates often concern applications outside commercial banking. Brynjolfsson, Li and Raymond study staggered access to an assistant among 5,172 customer-support agents and report an average 15% increase in issues resolved per hour in the published article [22]. The setting is a nonbank enterprise-software support operation. The result supports a knowledge-assistance mechanism, particularly for less experienced workers, but cannot be translated into a 15% reduction in bank costs.

Noy and Zhang's experiment with professional writing tasks provides additional evidence that generative assistance can change completion time and evaluated output quality [23]. Banking applications impose further requirements concerning sensitive data, current product terms and the consequences of incorrect advice. A useful internal assistant should make its supporting information inspectable and direct uncertain or unauthorised requests to an accountable employee. Its evaluation should include the accuracy of completed responses, repeat contacts and correction time rather than the amount of text produced.

Customer behaviour can also affect the operational result. Luo and colleagues' field experiment shows that chatbot disclosure changes purchasing behaviour in a structured outbound-sales setting [24]. This is evidence about customer response in that setting, rather than justification for concealing artificial intelligence use in banking. Banks need

transparent interaction design and accessible human support when requests involve financial distress, disputes or unusual circumstances. A chatbot that shortens initial contacts while increasing repeat enquiries can raise the total cost of resolving a customer's problem.

Human involvement must itself be designed and tested. Vaccaro and colleagues' cross-domain meta-analysis finds that combining humans and artificial intelligence does not automatically outperform the better of the two alone [25]. The finding does not remove the need for accountability in financial services. It shows why a generic human review step cannot be assumed to create superior decisions. Reviewers need appropriate information, time and authority, and escalation rules should identify the cases in which their contribution is most valuable.

4.4. Evidence at Bank and Industry Levels

Bank-level evidence is closer to the article's central outcome, but often relies on indirect measures of technology. Lee and colleagues find a positive association between fintech development and bank cost efficiency in China using frontier methods and panel analysis [26]. The study concerns the fintech environment rather than artificial intelligence deployment alone. It is useful for understanding how technological change can interact with banking efficiency, while leaving the incremental contribution of a particular artificial intelligence application unresolved.

More direct studies also require careful interpretation. Li examines artificial intelligence adoption proxies, profitability and operating-cost channels using a panel of 430 Chinese banks over 2007–2022 [27]. Baffour Gyau and colleagues analyse artificial intelligence technology innovation and banking financial performance in a 20-country panel [28]. These studies provide evidence about performance associations under their respective measures and identification assumptions. Annual-report language, patents and wider innovation indicators may capture strategic orientation or the surrounding technology environment as well as deployed systems. Profitability and estimated cost mechanisms should not be relabelled as independently observed workflow efficiency.

Profitable banks may have greater capacity to invest, disclose more technological activity or attract specialised staff. Management quality, market structure and public digital infrastructure can also influence both adoption and performance. Fixed effects and dynamic panel estimators address some sources of confounding, but their validity depends on assumptions about remaining variation and instruments. The evidence therefore supports examining artificial intelligence as a potential contributor to efficiency, with heterogeneous effects, rather than asserting a settled average causal effect across commercial banks.

Market structure provides an additional boundary condition. Balyuk and colleagues find that fintech small-business lending substitutes more strongly for lending by large or out-of-market banks than for lending by small or in-market banks [29]. This is consistent with advantages in processing hard information rather than a general replacement of relationship-based assessment. For incumbent banks, it suggests that the relevance of automated screening depends partly on the customers and information that already support their lending model.

Table 1 compares selected empirical results and identifies the limits on transferring them to commercial bank efficiency. Its purpose is to preserve the connection between each reported result and its original unit of analysis.

Table 1. Representative Empirical Evidence and Its Relevance to Banking Efficiency

Study and setting	Observed outcome or evidence	Interpretation for commercial banks
Fuster et al. [16]; US mortgage lenders	Technology-based lenders processed applications approximately 20% faster.	Supports workflow gains; digital lending and nonbank participation prevent attribution to AI alone.

Berg et al. [17]; German e-commerce credit	Digital footprints added default-prediction information.	Supports screening; prediction gains do not measure bank operating costs.
Lee et al. [26]; Chinese banks	Fintech development analysed against frontier efficiency.	Direct efficiency outcome; broader fintech exposure is not a bank AI deployment measure.
Li [27]; 430 Chinese banks	AI proxies positively associated with profitability over 2007–2022.	Bank-level association; proxy validity and adoption endogeneity remain important.
Brynjolfsson et al. [22]; nonbank support	Issues resolved per hour increased by 15% on average.	Supports augmentation; bank-specific costs and quality require separate validation.
Vaccaro et al. [25]; cross-domain experiments	Human–AI combinations performed below the better standalone performer on average.	Oversight and task allocation need testing; human involvement is not automatic synergy.

4.5. Distributional Effects and Service Quality

An efficiency improvement for a bank can have uneven effects on customers. Fuster and colleagues show that machine learning in mortgage screening can change the distribution of benefits across borrower groups [30]. Bartlett and colleagues document lending discrimination in the fintech era, with results that vary across mortgage market segments [31]. These studies do not support assuming that automated decisions eliminate unequal treatment. Differences in historical representation, data availability and pricing objectives can persist within apparently accurate systems.

Distributional analysis belongs within operational evaluation because exclusion, complaints and remediation affect service outcomes and operating costs. Relevant comparisons include approval and pricing outcomes among appropriately comparable applicants, access for customers with limited digital records, and the availability of review for contested decisions. The appropriate constraints depend on the jurisdiction and product. A bank should establish them before optimisation, rather than infer acceptable treatment from a model's aggregate accuracy after deployment.

5. An Integrated Framework for Evaluating Net Efficiency

5.1. Linking Models to Completed Banking Services

Figure 1 summarises the proposed framework. Artificial intelligence capabilities act through information processing and changes in the allocation of work. Observable workflow outcomes then contribute to bank-level efficiency, subject to operating context and quality constraints. The arrows describe analytical mechanisms for evaluation rather than estimated causal relationships. The framework deliberately retains implementation and governance costs between a successful model test and a claim about banking performance.

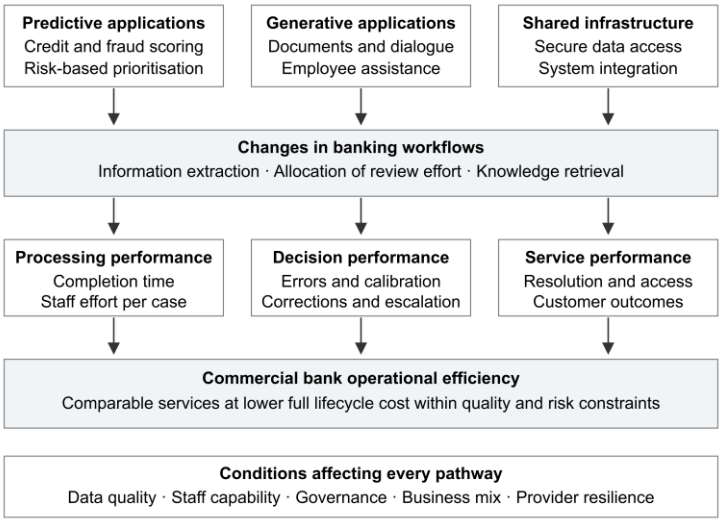


Figure 1. Conceptual Pathways Linking Artificial Intelligence Applications to Commercial Bank Operational Efficiency

Note. Author's synthesis of the reviewed literature. The pathways are conditional propositions, not an estimated structural model.

Evaluation should begin with a precisely specified service. For a loan application, the relevant endpoint may be a documented decision following all required checks. For customer service, it may be a resolved request without an avoidable repeat contact. For transaction monitoring, it may be an appropriately investigated alert. Defining this endpoint makes it possible to compare resource use while preserving the quality of the service. It also discourages reporting a convenient intermediate indicator as if it were the final result.

Table 2 identifies measures at the application, workflow and bank levels. No single measure is sufficient. The same application can improve completion time while increasing correction costs or disadvantaging particular customers. Measurement should connect those outcomes to a common operational objective and include a comparison with the process that would otherwise have been used.

Table 2. Measures for Evaluating Artificial Intelligence and Bank Operational Efficiency

Evaluation level	Primary measure	Necessary companion measures
Credit prediction	Calibration and decision value	Realised losses, approval outcomes and performance under stress.
Document processing	Cost per verified completed case	Omission rate, correction time and escalation share.
Customer service	Cost per resolved request	Repeat contacts, factual accuracy, complaints and accessibility.
Fraud and transaction alerts	Investigation workload at a defined detection standard	Missed cases, false alerts, customer disruption and investigation quality.
Bank operations	Comparable unit costs and frontier efficiency	Business mix, input prices, asset quality and service quality.
Technology lifecycle	Net operating benefit over an explicit horizon	Integration, inference, monitoring, human review and exit costs.

5.2. Full Lifecycle Costs and Organisational Conditions

Net efficiency depends on costs incurred beyond the production of model outputs. Data preparation, secure access, system integration and employee training precede use. Inference charges, maintenance, validation and human review continue after deployment. Errors introduce correction costs and potential customer harm. Some costs are fixed and

shared across applications, whereas others increase with usage. Their allocation should be explicit so that a successful pilot is not evaluated using a cost basis unavailable at scale.

Data quality influences both predictive and generative applications. Inconsistent customer identifiers, outdated documents and unclear access rights can undermine an otherwise capable model. Improved data infrastructure can benefit multiple processes, which makes its contribution difficult to isolate but important to record. Staff also need incentives to report errors and authority to challenge outputs. If performance targets reward rapid closure alone, employees may accept unverified suggestions and transfer the resulting work to complaints or compliance teams.

The chosen objective also influences decision quality. Dobbie and colleagues show that lending bias can depend on whether loan examiners are assessed against short-run default outcomes or longer-run profitability [32]. Their study examines human lending decisions, but motivates an analogous question for artificial intelligence: does the optimisation target represent the banking outcome that managers actually seek? A technically effective model can reproduce an unsuitable objective more consistently without improving the underlying service.

Model choice should follow the decision requirement. Rudin argues for interpretable models in high-stakes decisions when suitable interpretable alternatives are available [33]. This is a relevant design principle for banking, rather than proof that a single model class is always best. Banks should compare simpler benchmarks with more complex systems on economically relevant outcomes. A modest predictive improvement may not justify added validation difficulty, particularly where decision explanations and reliable challenge procedures are essential.

Scale can moderate the result. Large banks may spread fixed investments over more transactions, but complex legacy systems can increase integration costs. Smaller banks may obtain capabilities through shared providers, while becoming more dependent on external pricing and service continuity. Lending and payments also differ in how rapidly outcomes become observable. A payment-processing error may appear immediately; a credit decision's consequences can emerge much later. These differences make application-specific evaluation more credible than a uniform target for artificial intelligence expenditure.

5.3. Governance and Operational Resilience

Governance is part of the operating model because it determines how errors are detected, corrected and prevented from recurring. Generative systems require controls against unsupported content, inappropriate data disclosure and unauthorised actions. Retrieval from approved documents may improve the evidentiary basis of an answer, but does not guarantee correctness. The bank still needs version control, access restrictions, testing of difficult cases and a procedure for withdrawing a system when its performance becomes unacceptable [20].

The Financial Stability Board identifies vulnerabilities associated with common providers, correlated behaviour and model or data weaknesses [34]. Its later monitoring framework highlights gaps in adoption data and the importance of dependencies across the artificial intelligence supply chain [35]. These are system-level considerations, not estimates that every bank incurs the same losses. They nevertheless affect bank-level efficiency when service interruptions, supplier changes or unavailable substitutes create costs that routine performance measures omit.

Provider assessment should therefore consider how long a service can remain unavailable, which operations depend on it and how difficult replacement would be. Maintaining an alternative process has a cost, but may preserve essential banking services during an outage. The decision is not to maximise automation without constraints. It is to choose a process that delivers the required service reliably at an acceptable total cost. Independent validation and internal audit should assess whether reported benefits persist after these requirements are included.

6. Discussion and Future Research

6.1. Implications for Commercial Bank Management

The synthesis supports beginning with operational problems whose completion criteria and quality standards are observable. Document intake and internal knowledge assistance may be suitable candidates when data permissions and escalation rules are clear. Credit approval and customer advice involve different consequences and require evaluation appropriate to the product. Application selection should follow the expected value of improved decisions or reduced effort, together with the capacity to detect and correct failures.

Banks should record both assisted and unassisted performance under comparable conditions. Where feasible, phased or randomised rollouts can reveal whether apparent gains reflect the tool or differences in staff and customers. Learning periods should be observed separately from later operation, while retaining integration costs in the investment assessment. Managers also need to examine where saved time goes. It may support more customer contact or reduce a backlog before any staff expenditure changes. These outcomes can be valuable even when the short-run cost-to-income ratio remains unchanged.

Governance should be proportionate to the use case, with clear ownership of the model, workflow and customer outcome. A low-risk drafting aid and an automated lending decision need different controls. Shared inventories can document the technology, data sources, affected process, deployment date and provider. Such records make it easier to compare applications, identify dependencies and retire systems whose maintenance costs exceed their realised contribution.

6.2. Priorities for Empirical Research

Future studies need deployment measures that distinguish installed capability from actual use. Annual-report keywords and patents can indicate strategic attention, but should be validated against model inventories, transaction coverage or staff usage. Researchers should identify whether the application concerns prediction, generation or broader digital infrastructure. Without that distinction, the estimated treatment may combine several organisational changes and obscure the mechanism under investigation.

Stronger identification requires comparison groups and timing that address selective adoption. Profitable banks may invest earlier, while struggling banks may adopt tools during restructuring. Bank fixed effects alone do not remove changes in management, product strategy or borrower composition that occur alongside deployment. Event studies and differences-in-differences designs should document pre-adoption trends and treatment timing, using methods appropriate to staggered adoption and heterogeneous effects [36]. Workflow experiments can offer stronger local identification, although their external validity across banks remains an empirical question.

Efficiency estimation also needs disciplined inference. Frontier studies should assess sensitivity to the definition of outputs, input prices and environmental differences. When estimated data envelopment analysis scores enter a second-stage regression, conventional inference can be misleading because the scores are generated and statistically dependent [37]. Appropriate estimation methods improve inference, but cannot compensate for an artificial intelligence measure that fails to represent deployment. Combining workflow logs with accounting and frontier measures would make the outcome hierarchy testable.

Longer observation is especially important in lending. Early processing gains may coexist with risks that become visible only when economic conditions change. Studies should follow loan cohorts, monitor revisions to decision rules and report service outcomes for different customer groups. Differences by bank size, ownership, legacy systems and staff capability can then be examined as pre-specified sources of heterogeneity. Cross-country comparisons must additionally account for regulatory and institutional differences that influence both adoption and measured performance.

6.3. Limitations of the Review

This review organises heterogeneous evidence and develops an evaluative framework; it does not estimate an average effect or rank commercial banks. Its targeted search may omit relevant studies, and public disclosure provides an incomplete view of deployed systems and implementation costs. Positive results may also be more visible than unsuccessful applications. Foundational fintech studies concern technologies and market conditions that differ from current generative systems. Cross-sector experiments establish mechanisms within their own settings, while their transfer to regulated banking remains conditional. These limitations strengthen the case for transparent deployment records and bank-specific outcome measurement.

7. Conclusion

Artificial intelligence can improve commercial bank operations by reducing repetitive processing, supporting credit and monitoring decisions, and assisting employees with information-intensive work. The available evidence is strongest for particular tasks and lending processes. Broader claims about bank-level efficiency remain conditional on how technology is measured, whether adoption is credibly identified and which costs and service outcomes are included.

A defensible efficiency assessment connects model performance to completed banking services and then to comparable bank-level outcomes. It includes implementation, validation, human review and remediation costs, together with service quality and risk. Data infrastructure, organisational redesign and employee capability determine whether local productivity gains become durable operating improvements. Commercial banks should therefore evaluate artificial intelligence through verified workflow outcomes and full lifecycle costs, while future research tests these mechanisms with deployment-level data and credible counterfactuals.

References

1. A. V. Thakor, "Fintech and banking: What do we know?," *J. Financial Intermediation*, vol. 41, Art. no. 100833, 2020, doi: 10.1016/j.jfi.2019.100833.
2. I. Goldstein, W. Jiang, and G. A. Karolyi, "To FinTech and beyond," *Rev. Financial Stud.*, vol. 32, no. 5, pp. 1647–1661, 2019, doi: 10.1093/rfs/hhz025.
3. Bank of England and Financial Conduct Authority, *Artificial Intelligence in UK Financial Services—2024*, 2024. [Online]. Available: https://www.bankofengland.co.uk/report/2024/artificial-intelligence-in-uk-financial-services-2024
4. X. Vives, "Digital disruption in banking," *Annu. Rev. Financial Econ.*, vol. 11, no. 1, pp. 243–272, 2019, doi: 10.1146/annurev-financial-100719-120854.
5. G. Buchak, G. Matvos, T. Piskorski, and A. Seru, "Fintech, regulatory arbitrage, and the rise of shadow banks," *J. Financial Econ.*, vol. 130, no. 3, pp. 453–483, 2018, doi: 10.1016/j.jfineco.2018.03.011.
6. I. Aldasoro, L. Gambacorta, A. Korinek, V. Shreeti, and M. Stein, *Intelligent Financial System: How AI Is Transforming Finance*, BIS Working Papers No. 1194. Basel: Bank for International Settlements, 2024. [Online]. Available: https://www.bis.org/publ/work1194.htm
7. J. M. Liberti and M. A. Petersen, "Information: Hard and soft," *Rev. Corporate Finance Stud.*, vol. 8, no. 1, pp. 1–41, 2019, doi: 10.1093/rcfs/cfy009.
8. A. N. Berger and D. B. Humphrey, "Efficiency of financial institutions: International survey and directions for future research," *Eur. J. Oper. Res.*, vol. 98, no. 2, pp. 175–212, 1997, doi: 10.1016/S0377-2217(96)00342-6.
9. A. N. Berger and L. J. Mester, "Inside the black box: What explains differences in the efficiencies of financial institutions?," *J. Banking Finance*, vol. 21, no. 7, pp. 895–947, 1997, doi: 10.1016/S0378-4266(97)00010-1.
10. A. Charnes, W. W. Cooper, and E. Rhodes, "Measuring the efficiency of decision making units," *Eur. J. Oper. Res.*, vol. 2, no. 6, pp. 429–444, 1978, doi: 10.1016/0377-2217(78)90138-8.
11. D. Aigner, C. A. K. Lovell, and P. Schmidt, "Formulation and estimation of stochastic frontier production function models," *J. Econometrics*, vol. 6, no. 1, pp. 21–37, 1977, doi: 10.1016/0304-4076(77)90052-5.
12. E. Brynjolfsson and L. M. Hitt, "Beyond computation: Information technology, organizational transformation and business performance," *J. Econ. Perspect.*, vol. 14, no. 4, pp. 23–48, 2000, doi: 10.1257/jep.14.4.23.
13. A. N. Berger, "The economic effects of technological progress: Evidence from the banking industry," *J. Money Credit Bank.*, vol. 35, no. 2, pp. 141–176, 2003, doi: 10.1353/mcb.2003.0009.
14. E. Brynjolfsson, D. Rock, and C. Syverson, "The productivity J-curve: How intangibles complement general purpose technologies," *Am. Econ. J.: Macroecon.*, vol. 13, no. 1, pp. 333–372, 2021, doi: 10.1257/mac.20180386.

15. S. Raisch and S. Krakowski, "Artificial intelligence and management: The automation–augmentation paradox," *Acad. Manage. Rev.*, vol. 46, no. 1, pp. 192–210, 2021, doi: 10.5465/amr.2018.0072.
16. A. Fuster, M. Plosser, P. Schnabl, and J. Vickery, "The role of technology in mortgage lending," *Rev. Financial Stud.*, vol. 32, no. 5, pp. 1854–1899, 2019, doi: 10.1093/rfs/hhz018.
17. T. Berg, V. Burg, A. Gombović, and M. Puri, "On the rise of FinTechs: Credit scoring using digital footprints," *Rev. Financial Stud.*, vol. 33, no. 7, pp. 2845–2897, 2020, doi: 10.1093/rfs/hhz099.
18. J. Jagtiani and C. Lemieux, "The roles of alternative data and machine learning in fintech lending: Evidence from the LendingClub consumer platform," *Financial Manage.*, vol. 48, no. 4, pp. 1009–1029, 2019, doi: 10.1111/fima.12295.
19. L. Gambacorta, Y. Huang, H. Qiu, and J. Wang, "How do machine learning and non-traditional data affect credit scoring? New evidence from a Chinese fintech firm," *J. Financial Stability*, vol. 73, Art. no. 101284, 2024, doi: 10.1016/j.jfs.2024.101284.
20. J. C. Crisanto, C. B. Leuterio, J. Prenio, and J. Yong, *Regulating AI in the Financial Sector: Recent Developments and Main Challenges*, FSI Insights on Policy Implementation No. 63. Basel: Bank for International Settlements, 2024. [Online]. Available: https://www.bis.org/fsi/publ/insights63.htm
21. M. Di Maggio and V. Yao, "Fintech borrowers: Lax screening or cream-skimming?," *Rev. Financial Stud.*, vol. 34, no. 10, pp. 4565–4618, 2021, doi: 10.1093/rfs/hhaa142.
22. E. Brynjolfsson, D. Li, and L. Raymond, "Generative AI at work," *Quart. J. Econ.*, vol. 140, no. 2, pp. 889–942, 2025, doi: 10.1093/qje/qjae044.
23. S. Noy and W. Zhang, "Experimental evidence on the productivity effects of generative artificial intelligence," *Science*, vol. 381, no. 6654, pp. 187–192, 2023, doi: 10.1126/science.adh2586.
24. X. Luo, S. Tong, Z. Fang, and Z. Qu, "Frontiers: Machines vs. humans: The impact of artificial intelligence chatbot disclosure on customer purchases," *Marketing Sci.*, vol. 38, no. 6, pp. 937–947, 2019, doi: 10.1287/mksc.2019.1192.
25. M. Vaccaro, A. Almaatouq, and T. Malone, "When combinations of humans and AI are useful: A systematic review and meta-analysis," *Nat. Hum. Behav.*, vol. 8, no. 12, pp. 2293–2303, 2024, doi: 10.1038/s41562-024-02024-1.
26. C.-C. Lee, X. Li, C.-H. Yu, and J. Zhao, "Does fintech innovation improve bank efficiency? Evidence from China's banking industry," *Int. Rev. Econ. Finance*, vol. 74, pp. 468–483, 2021, doi: 10.1016/j.iref.2021.03.009.
27. Z. Li, "Does artificial intelligence enhance bank profitability? Evidence from China," *Finance Res. Lett.*, vol. 90, Art. no. 109366, 2026, doi: 10.1016/j.frl.2025.109366.
28. E. Baffour Gyau, M. Appiah, B. A. Gyamfi, T. Achie, and M. A. Naeem, "Transforming banking: Examining the role of AI technology innovation in boosting banks financial performance," *Int. Rev. Financial Anal.*, vol. 96, Art. no. 103700, 2024, doi: 10.1016/j.irfa.2024.103700.
29. T. Balyuk, A. N. Berger, and J. Hackney, "What is fueling FinTech lending? The role of banking market structure," *Rev. Corporate Finance Stud.*, vol. 15, no. 2, pp. 305–351, 2026, doi: 10.1093/rcfs/cfae028.
30. A. Fuster, P. Goldsmith-Pinkham, T. Ramadorai, and A. Walther, "Predictably unequal? The effects of machine learning on credit markets," *J. Finance*, vol. 77, no. 1, pp. 5–47, 2022, doi: 10.1111/jofi.13090.
31. R. Bartlett, A. Morse, R. Stanton, and N. Wallace, "Consumer-lending discrimination in the FinTech era," *J. Financial Econ.*, vol. 143, no. 1, pp. 30–56, 2022, doi: 10.1016/j.jfineco.2021.05.047.
32. W. Dobbie, A. Liberman, D. Paravisini, and V. Pathania, "Measuring bias in consumer lending," *Rev. Econ. Stud.*, vol. 88, no. 6, pp. 2799–2832, 2021, doi: 10.1093/restud/rdaa078.
33. C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nat. Mach. Intell.*, vol. 1, no. 5, pp. 206–215, 2019, doi: 10.1038/s42256-019-0048-x.
34. Financial Stability Board, *The Financial Stability Implications of Artificial Intelligence*, 2024. [Online]. Available: https://www.fsb.org/2024/11/the-financial-stability-implications-of-artificial-intelligence/
35. Financial Stability Board, *Monitoring Adoption of Artificial Intelligence and Related Vulnerabilities in the Financial Sector*, 2025. [Online]. Available: https://www.fsb.org/2025/10/monitoring-adoption-of-artificial-intelligence-and-related-vulnerabilities-in-the-financial-sector/
36. B. Callaway and P. H. C. Sant'Anna, "Difference-in-differences with multiple time periods," *J. Econometrics*, vol. 225, no. 2, pp. 200–230, 2021, doi: 10.1016/j.jeconom.2020.12.001.
37. L. Simar and P. W. Wilson, "Estimation and inference in two-stage, semi-parametric models of production processes," *J. Econometrics*, vol. 136, no. 1, pp. 31–64, 2007, doi: 10.1016/j.jeconom.2005.07.009.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of Publisher and/or the editor(s). Publisher and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.