



Executive
Perspectives

How AI-First Banks Are Rewriting the Rules of Retail Banking

Banking

June 2026



Introduction

Retail banks are poised to unlock significant value from GenAI and Agentic AI. They are **best positioned** with relatively higher levels of digitization, rich data (albeit legacy bound), and higher customer adoption of GenAI applications.

AI-first retail banks are driving transformation through decisive AI investments and are capturing value at scale across business functions. However, for other banks, value remains limited to select pockets.

In this edition, we explore the impact of GenAI and Agentic AI in retail banking and the approach followed to deliver impact. We address **key questions faced** by bank executives:

- Why are GenAI and Agentic AI imperatives, yet **why is at-scale value elusive**?
- Which **functions are realizing measurable impact**?
- What **approach are AI-first banks taking** across customer discovery & onboarding, service & engagement, and fulfilment in ops, credit & collections?
- How are banks building an **AI-first mindset** across the organization?

In this BCG Executive Perspective, we articulate a vision for transformative GenAI and Agentic AI deployment in retail banking

Executive summary | Retail banks are poised to derive significant value from GenAI and Agentic AI (1/2)

WHY

GenAI and Agentic AI in retail banking are imperatives but with current under-realized value

- **Structurally primed for GenAI and Agentic AI:** Retail banks have seen higher levels of digitization and a richer data repository (albeit, within legacy systems); they are **more ready to augment intelligence and create new experiences** for customers
- **Value scale-up has been limited: Six key challenges addressed by AI-first retail banks-** a) Moving beyond micro use cases, b) Upfront linkage of initiatives with value metrics, c) CEO involvement to drive accountability, d) Lead with products, build for re-use, e) Building GenAI fluency for change-the-bank roles, f) Design risk controls and processes from Day 1
- **Direct comparison with other retail experiences:** Retail consumers compare banking experiences with AI-native experiences in other sectors, such as shopping, travel, and digital media; with AI tech advancing ever more quickly, **banks need to move faster**

WHAT

key skills are being adopted by AI-first retail banks (1/2)

- **Step-change in marketing maturity with higher precision in acquisition: Over 40% sales uplift in new-to-bank products**
 - **Synthetic customer personas** created using GenAI to sharpen proposition design, targeting, and channel choices
 - Improved **visibility in generative search engine** and driving acquisition at lower CACs
 - Agentic interventions in **media allocation optimization** to reduce the number of low and non-performing channels; **high-velocity funnel experimentation** to apply best CRO practices to improve journey conversions
- **Customer service is changing from reactive support to insight-driven, personalized & proactive engagement**
 - **Human-like conversational assistance in service, reminders, and collections** by agentic bots; **70%+ of human call volumes** are enabled with voice bots at about **one-fifth the cost of traditional assistance**
 - **Always-on personal RM for every customer** that captures context from every interaction, proactively engages customers, and provides contextual solutions; **deepen customer relationships by 20%-40% through meaningful cross-sell**
 - **Step change in frontline RM productivity** across cross-sell, upsell, engagement, and customer service. Bankers augmented with customer-level agenda, personalized pitches, and knowledge assistance **have engaged 50% of clients weekly (from ~15% of clients). 5-6x conversions observed for select products**

Executive summary | Retail banks are poised to derive significant value from GenAI and Agentic AI (2/2)

WHAT

key skills are being adopted by AI-first retail banks (2/2)

- **Operating models and workflows are redefined in fulfillment channels between agentic bots and human experts:**
 - **Invisible operations or zero human-operations swim lane is increasing** by enabling the processing of non-standard files. Partial STP and non-STP cases are handled differently, with human experts engaged only where needed, while agentic execution and monitoring help ensure guaranteed SLAs; observed impact includes **70% lower turnaround time**
 - **Credit processing is becoming faster and smarter:** Bot agent-enabled underwriting engine is compressing decision cycles and follow-up communication, delivering **5x-10x faster time to quote**
 - **Smart anti-financial crime** processes reimaged with agentic bots driving customer identification and screening to onboarding and perpetual KYC monitoring; **up to 50% financial crime cost savings observed**
 - **Smart collections:** Unstructured interaction data leveraged to understand customers' reasons for bounce, to personalize outreach strategy, and enable higher resolution; **~50% reduction in collections operating costs**
 - **Harness engineering** to accelerate product and software delivery; **50% faster time to market (BRD to deploy)**
 - **Employee alter-ego** or bot agents to perform role-specific activities and other enabling tasks

HOW

banks should set themselves up for a successful AI-led transformation

- **Prioritize 3-4 chosen areas of transformation that amplify the bank's strategy in market** with a focus on business outcomes, instead of chasing many micro use cases
- For each prioritized functional reimagination, identify opportunities by focusing on nature of job instead of number of people - **structurally reallocating toil/repetitive tasks** and **reasoning tasks to AI** and augmenting **humans for expertise**
- **Build reusable products** while transforming business function, rather than adopting a GenAI platform-first approach
- **Design structured AI risk and compliance** processes from Day 1, not as an afterthought during launches
- At **early-stage maturity**, GenAI requires **centralized focus**; GenAI products can **evolve to more federated** structures in the future **once core capabilities stabilize**
- Invest in **GenAI upskilling** across the org, spanning **individual productivity improvements to building function-specific** capabilities

AI-first retail banks are addressing six key challenges to scaling GenAI initiatives



Approaching transformation through use cases

Isolated use cases with limited value; e.g., tool for branch staff to answer product FAQs, auto-generation of call summaries in contact centers, etc.



Building solutions, looking for problems

Insufficient focus on solving a real problem vs. just showcasing the art of the possible; e.g., in-call real-time nudges for human callers, video avatars for engagement, etc.



Creating clear accountability

Without clear ownership GenAI initiatives get lost in the complex coordination between business, channel, tech, AI, compliance, risk and third parties



Balancing product vs. platform objectives

Pressure to deliver immediate value leads teams to build point solutions that do not scale horizontally, e.g., RM intelligence that can not scale to D2C



Slower adoption of GenAI fluency across teams

Concentration of knowledge in specialist teams including Data Science and Engineering, with checkered understanding across other functions



Solving risk as an afterthought

Speed-to-market becomes unpredictable without institutional knowledge to navigate risk and compliance decisions

What AI-first banks are doing differently

Fewer but deeper functional reimaginations

Focus on a few priorities aligned with overall strategic agenda; e.g., reimagined frontline sales and customer service

Upfront linkage of initiatives with value metrics

Embed targets in AOPs with clear value measurement and roadmap

CEO involvement to drive joint accountability

Drive GenAI agenda from the top of the house to eliminate ambiguous accountability, with strong execution discipline

Lead with products, build for reuse

Banks that exercise patience to get foundations right, unlock exponentially faster value across initiatives through reuse

Invest in GenAI upskilling for change-the-bank roles

Invest in upskilling, so that change can be driven bottom up as well; encourage organization-wide ideation

Build in risk controls and processes

Build risk controls and new AI governance by design from day one to avoid retrofitting. Set up an accelerated approval committee to expedite decisions on architecture, security, risk, and compliance

AI-first retail banks are investing in ten key skills

Attract and Acquire

1 A. Customer Intelligence and Proposition Design: Persona agents built using real customer behavior data create a captive testing ground to validate propositions, sharpen customer targeting, and de-risk new product launches before spending even a dollar

B. Growth Marketing and Conversion:

- **AEO- and GEO-ready search:** Optimized discovery across answer and generative search engines
- **Conversational Acquisition:** Embedded chat in popular chat engine interfaces, e.g., ChatGPT and Claude
- **Agentic Media Optimization:** Autonomous real-time media buying and targeting to maximize reach and ROI
- **Agentic CRO:** Intense agent-driven customer journey funnel experimentation to boost conversions

+40%¹ Sales uplift in NTB products

Serve and Engage

2 Conversational Customer Assistance:

- Agentic bot-led **engagement, inbound service, outbound collections,** and reminders (voice and chat)
- **Persuasion by agentic bots** via customer-level contextual pitches and arguments

70%+ Human call volume handled by voice bot **1/5th** Cost of assistance with human-like efficacy

3 Always-on Personal RM for Every Customer: Always-on agent that is proactive, provides contextual engagement, and is always available

60% Increase in activation rate for NTB customers **20%-40%** Increase in cross-sell

4 Banker Productivity: Augment bankers with **contextual pitches, personalized collaterals, and process knowledge** for cross-sell, upsell, and service

Wealth illustration

50% Clients engaged weekly from 15% starting point **~20%** Increase in AUM

Fulfill and Execute

5 Invisible Operations:

- **Instant and autonomous handling** of nonstandard requests (no human touch or queue)
- **Full-context routing** to right human skill; **guaranteed SLAs**

40% Higher operations efficiency **70%** Lower turnaround time

6 Smart Credit Engine:

- **Agentic underwriting** to augment loan decision making
- **Iterative learning** from portfolio outcomes; make **policy smarter** over time

100% Review coverage for all submissions **5x-10x** Faster time to quote

7 Smart Anti-Financial Crime: AI-led CDD/EDD, AML, screening, onboarding, and monitoring with humans as needed

50% Lower financial crime costs

8 Smart Collections: Boost share of digital collections using personalization and voice agents

~50% Reduction in collections operating costs **10%-20%** Net charge-off reduction

Harness Engineering

9 Using Harness Engineering across every stage of product and software delivery to accelerate idea to impact for business

+50% Faster time to market (BRD to deploy) **3x** Engineering throughput

Employee Alter-Ego

10 Employee Alter-Ego to augment role-specific tasks, minimize toil for other enabling tasks, and provide personalized coaching

Early adoption stage, at-scale impact to be realized

1. Up to three times digital sales uplift in early-maturity functions
 Note: NTB = New To Bank; AEO/ GEO = Answer / Generative Engine Optimization; CDD = Customer Due Diligence; EDD = Enhanced Due Diligence; AML = Anti-Money Laundering
 Source: All impact numbers based on value delivered from production-grade implementations at retail banks

Attract and Acquire | Invest in strong foundations while developing GenAI capabilities to capture value

AI Beginners

AI Practitioners

AI-First

Brilliant Basics: Building the foundation for Agentic Implementation

- Segment-centric beyond P&L
- From contextual to data-driven decision-making
- Close collaboration across teams¹
- Test-and-learn mindset
- Ownership of media channels
- E2E measurement and tracking

- 1 A push toward advanced capabilities **without core foundations sets banks back by ~15% in marketing maturity**
- 2 With GenAI, the **time to build foundations is shrinking from about five years to under one year**

GenAI Initiatives: Early efforts that banks can start from Day 1

- GenAI-powered process and knowledge base
- GenAI content creation
- Customer digital twin and synthetic persona
- AEO/GEO-Ready Search

- 3 To unlock early value, **invest in quick-win GenAI initiatives in parallel** to building brilliant basics

Agentic Capabilities

- Deploy bank app in generative search engines
- Agentic media optimization
- Agentic experimentation and CRO

- 4 Layer agentic capabilities once foundations are in place to **sustain advantage as market rapidly evolves**

Foundational capabilities Customer intelligence and proposition design Growth marketing and conversion

1. Marketing, Communication, CRM, and so on
 Note: E2E = End To End; AEO/ GEO = Answer / Generative Engine Optimization; CRO = Conversion Rate Optimization
 Source: All impact numbers based on value delivered from production-grade implementations at retail banks

Attract and Acquire | Five key capabilities are providing significant sales uplift

GenAI and Agentic AI across every stage of marketing ...

Customer digital twin and synthetic persona

Learn fast without spending real resources or taking real risk

- Validate messaging with simulated **buyers before spending a single dollar on media**
- Stress-test pricing and product features against real client profiles **without real-world risk**

AEO/ GEO-Ready Search

Be found where customers are already deciding

- Generative search engine-based acquisition at lower CACs; 64% of users using AI for purchasing decisions. Companies with GEO see **~7x higher AI visibility, with users spending 30%+ more time** on site with lower bounce from sites

Agentic media optimization

Always-on Agent that makes every ad dollar work harder

- Daily agentic optimizer to enforce cross-channel best practices and eliminates non-converting media spend. Marketers-in-the-loop steer strategy while the bot agents handle execution
- Delivering up to **20% revenue uplift and 25% efficiency gain**

Agentic experimentation and CRO

Scale experimentation from a handful of tests to hundreds

- Agents rapidly generate and test variants at volume while humans lead the "big bets" that reshape core journeys; a Latin American bank using this dual **approach, scaled to 5x more experiments per month and delivered 20% sales uplift**

Deploy bank app in generative search engines

Improve funnel conversions in GEO-led acquisitions

- Conversational assistance of bank-specific information by integrating bank app/ website with generative search engines like ChatGPT. A European bank has implemented a highly integrated banking experience with personalized answers based on in-app information

...providing significant value

40%

Sales uplift in NTB products
(up to three times digital sales uplift in early-maturity functions)

30%

Sales uplift in ETB products

60%

Uplift in affluent clients

40%

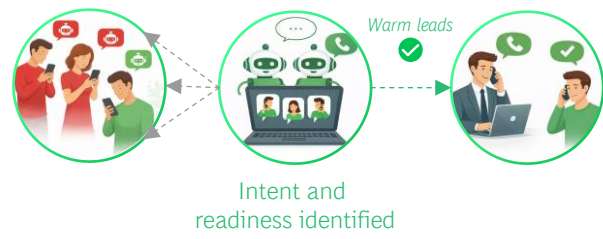
Reduced cost per acquisition

Note: AEO/ GEO = Answer / Generative Engine Optimization; CRO = Conversion Rate Optimization; NTB = New To Bank; ETB = Existing To Bank

Source: All impact numbers based on value delivered from production-grade implementations at retail banks

Conversational Customer Assistance | Significant value is realized from relationship deepening, outbound reminders, and inbound customer service

A Relationship deepening
Revenue lift: Increase in customer-to-RM ratio, at enhanced revenue per customer



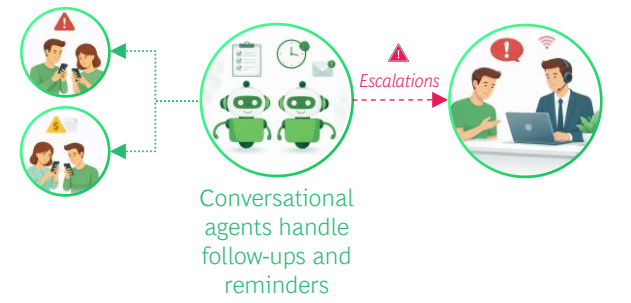
Role of AI-enabled agent

- **AI agents reach** out to customers at scale to **identify interest**
- **Personalized bot-led engagement** to validate customer intent
- **Qualify high-intent leads** and **schedule RM engagement** to deepen relationship

Impact

- 2x-3x** Customers per field RM
- 8x-10x** Customers per virtual RM

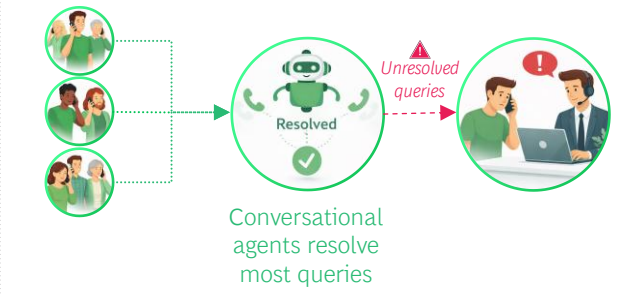
B Outbound reminders
Higher resolutions with proactive outreach and **higher NPS**



- **AI agents proactively trigger service and payment reminders** to drive timely customer action
- **Bot-led conversations convince** customers and enable query resolution
- **Human handoff** is triggered **only for true exceptions and escalations**

- 1/5th** Cost to serve at human resolution efficiency
- 70%+** Conversation volumes handled by voice bots

C Inbound customer service
Lower operating cost: Lower need for human **higher NPS**

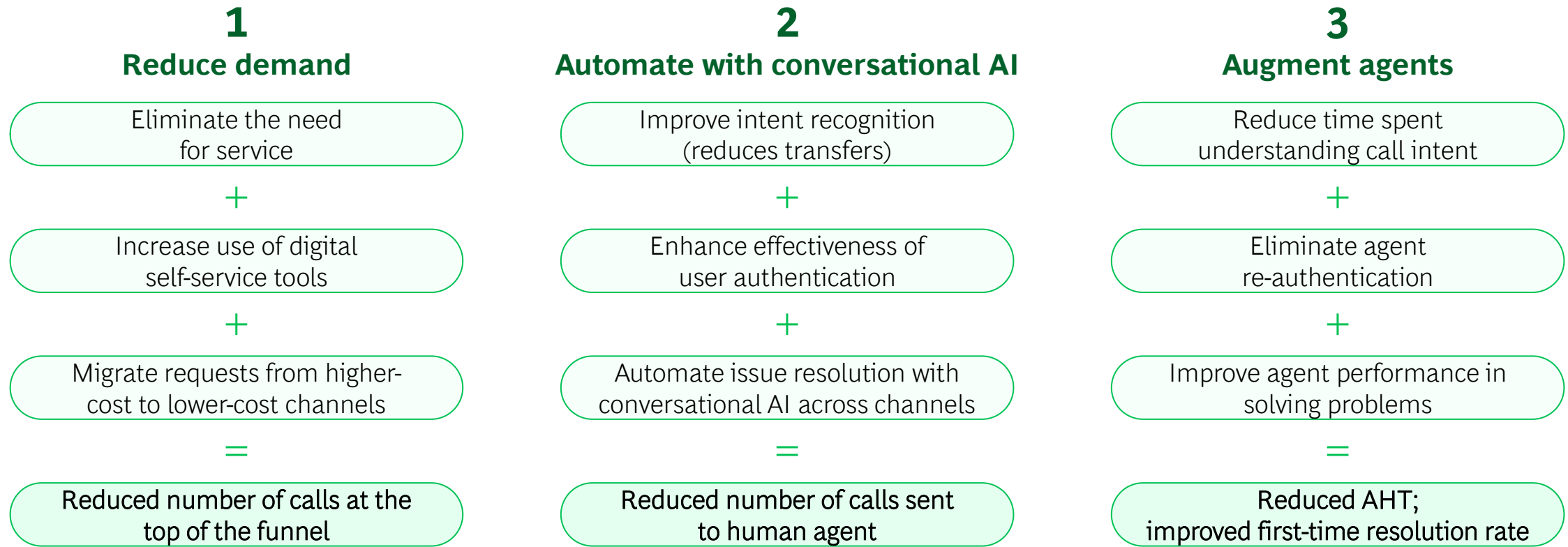


- **Chat & voice AI agents** answer **all inbound calls**
- **Bot agents resolve most simple, complex, and routine** queries
- **Escalations and exceptions** are routed to **human agents**

- 75%-90%** Lower cost to serve
- 90%-95%** Lower wait time

Note: NPS = Net Promoter Score
 Source: All impact numbers based on value delivered from production-grade implementations at retail banks

Conversational Customer Assistance | Investments in traditional call centers should be reconsidered, given high opportunity to reduce call volumes



Additional enablers to realize value include improving agent utilization, demand forecasting, and scheduling and training

Note: AHT = Average Handling Time
Source: All impact numbers based on value delivered from production-grade implementations at retail banks

Conversational Customer Assistance | Newer capabilities are required across the channel, experience, intelligence, and handling layers



Channel



Multi-channel experience



Intelligence



Handling

Traditional digital journeys

Assisted journeys	IVR	Field agent app	...
	Work apps and portals (RM, virtual RM)		
DIY journeys	App	Website	Social media ...

Based on shared data and common CRM

Traditional AI and ML models	Hyper-personalization		
	Customer persona	Next-best action	...
	Risk analytics	Propensity analytics	...

APIs for COTS systems for happy and a few unhappy swim lanes

Additional services for conversational engagement



New **machine voice, text chats, and multimodal** interfaces across all channels

Agentic orchestration to enable:



- Long-term **context**
- Curated **customer-level agenda**
- **Personalized pitches**
- **Follow-up loops** for closure



Power conversational experience with knowledge graphs:







- **Customer context**
- **Bank's process and policy context** across scenarios



- Ensure **deterministic execution of operating process flows** for secured banking (cannot be autonomous)
- Support **full spectrum of scenarios**, well beyond traditional digital journeys. Requires tooling layer of **APIs** (incl. new integrations), **data integration**, and **purpose-built tools** (e.g., calculators, schedulers)

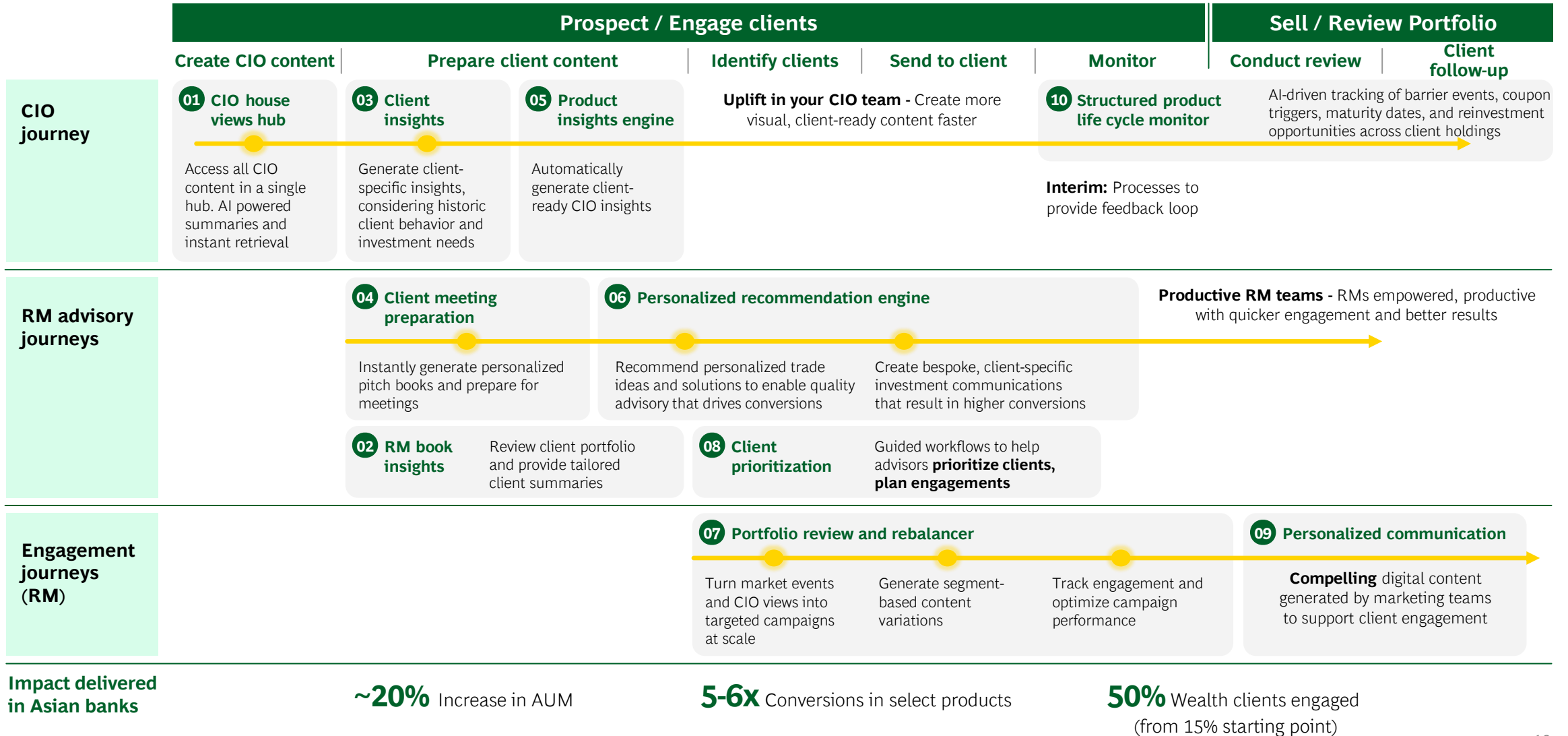
Agentic RM for Every Customer | Proactive signals and hypercontextual customer engagement are powered by six capability layers

Six new-age capability layers to change the way banks activate customers and engage with them to ensure business impact. Significant improvement seen in customer activation (**60% increase in NTB customers**) and relationship deepening, with **20%-40% improved cross-sell**

ORCHESTRATION LAYER		Orchestrates customer campaigns, engagement, and channel routing by coordinating services across the six layers			
 SENSE Never misses a moment, every event trigger captured instantly	Event monitor Captures every event trigger in real time (bank and customer initiated)	Customer behavioral signal Provides customer intent before a discussion based on previous interactions	Market context Market and competitor triggers to rescore customer intent in real time		
	 KNOW Rich customer context unified across profile data, holdings, and preferences from interactions	Customer profile 20,000+ customer DNA features; memory updated on every event	Customer context graph Unified graph linking signals, offers, responses, and outcomes	Offer relevance Reranks products vs. latest context on every update	Offer context and message Customer-specific brief per offer, channel-ready message
 DECIDE Focused customer targeting by optimizing propensity and margin		Customer-level propensity engine Campaign combination (e.g., mode, hook, and channel) scored for every event trigger	Treatment optimizer Selects winning campaign and calibrates incentive to maximize margin and propensity	Guardrail agent Complies with anti-repetition, consent, do-not-disturb, and regulatory requirements, etc	
	 ENGAGE Right pitch, hook, channel, and timing for every interaction	Emotional IQ Reads sentiment; suppresses sending and escalates when needed	Hook and timing Personalizes hook to capture customer attention in first 10 seconds; times discussion	Contextual pitch Convinces customers to take action, in a human-like manner, and handles rebuttals	
 ACT Execute actions on behalf of customers		Deterministic execution Handles and resolves customer queries as per bank's processes (deterministic flows)	Action looper Drives actions to closure by monitoring open actions and sending reminders	Handoff agent Customer handoff to relevant channel; schedules RM meeting and provides context	
	 LEARN Continuously improves every model through feedback from attribution and KPIs	Attribution agents Channel attribution for single touch vs. multi-touch; channel optimization feedback	Model update Automated retraining based on recent data; model drift detection triggers	KPI monitor Tracks conversion and channel ROI in real time; alerts on deviations	

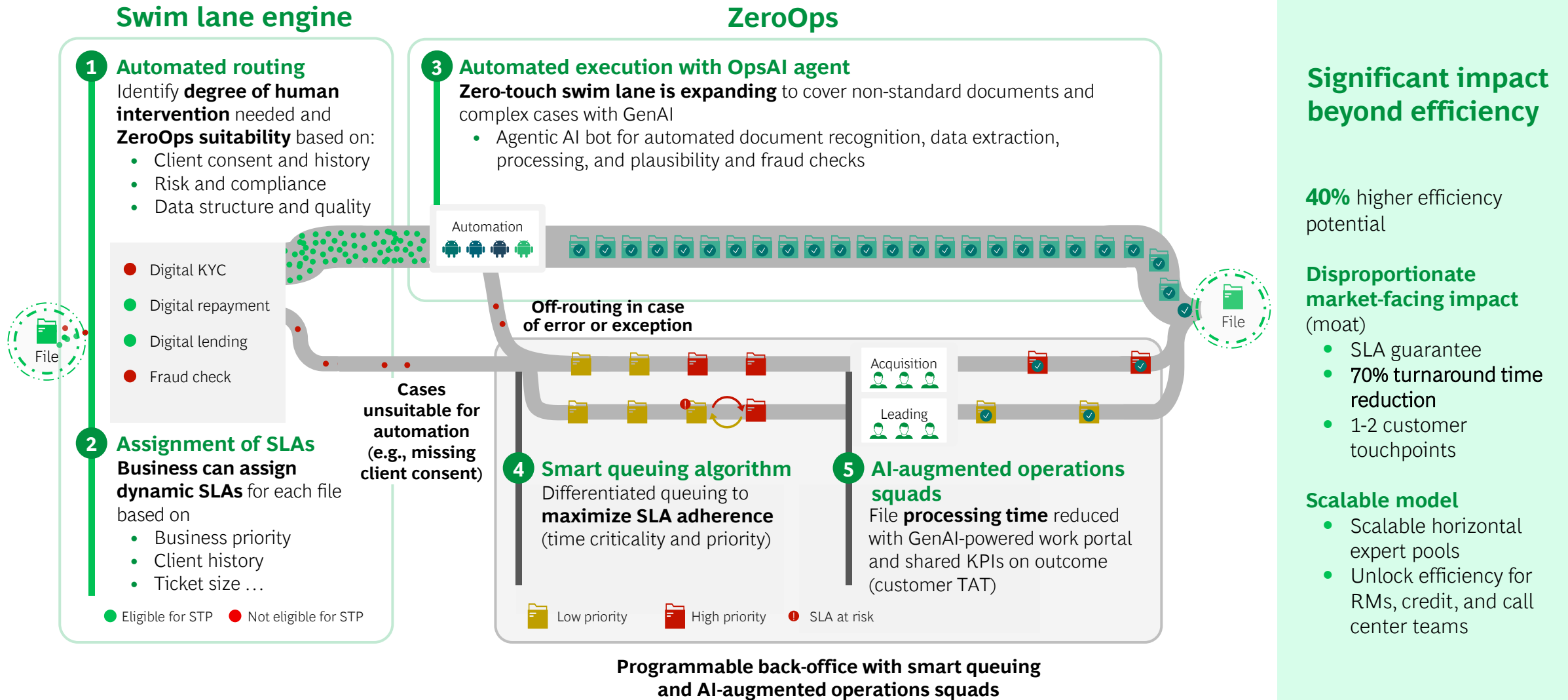
Note: NTB = New To Bank
 Source: All impact numbers based on value delivered from production-grade implementations at retail banks

Banker Productivity | Enable bankers to cross-sell, upsell, and service clients



Source: All impact numbers based on value delivered from production-grade implementations at retail banks

Invisible Operations | From cost center to competitive moat; AI-led operations deliver speed, differentiation, and customer outcomes



Significant impact beyond efficiency

40% higher efficiency potential

Disproportionate market-facing impact (moat)

- SLA guarantee
- 70% turnaround time reduction
- 1-2 customer touchpoints

Scalable model

- Scalable horizontal expert pools
- Unlock efficiency for RMs, credit, and call center teams

Note: TAT = Turnaround Time
Source: All impact numbers based on value delivered from production-grade implementations at retail banks

Smart Credit Engine | Step change in speed and scale of credit decision-making through AI agents across underwriting

Key pain points addressed

01 Manual data extraction

Time-consuming extraction from applications, financials, income proofs and third-party reports

02 Inconsistent risk assessment

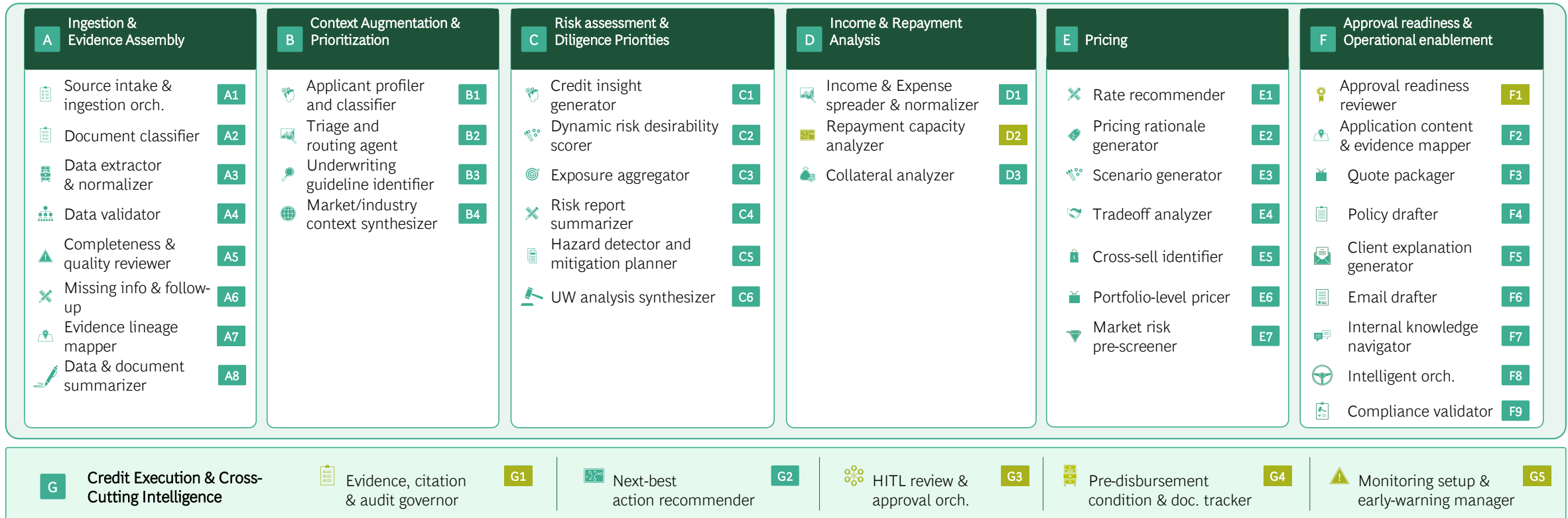
Subjective interpretation and variation in underwriter judgment leads to policy drift

03 Fragmented data sources

Hard to compile a full applicant view across bureau, income, collateral and behavioral data

04 Inefficient triage and prioritization

Suboptimal allocation of senior underwriter time across simple and complex cases



100% Submission coverage full review of all submissions received

5-10x Faster time to quote from weeks to days

100% User adoption across underwriters

Note: UW = Underwriting; HITL = Human in the loop
 Source: All impact numbers based on value delivered from production-grade implementations at retail banks

■ AI / Agentic AI capabilities ■ Deterministic calculations, workflows, controls

Smart Anti-Financial Crime | Agent bots leveraged in KYC onboarding and monitoring to realize significant cost save



Customer Identification

I. Validation Agent

II. Fraud Detection Agent

- Agentic bot validates customer identity details and flags mismatches, missing documents, or duplicates
- Biometric and data checks within seconds for AML and fraud risks
- Verifies KYC data for AML through trusted sources



Due Diligence and Financial Screening

III. Request for Information Agent

IV. Screening Agent

- Automatically screen risks (AML, PEP, and sanctions)
- Streamline customer follow-up: Define required documents, identify missing information, and handle follow-up communication
- Escalate only genuine high-risk cases to human review; risk rating created



Onboarding Decision and Control Setup

V. KYC File Assistant

VI. Admin and Pre-approval Agents

- KYC assistant analyses and synthesizes the KYC file per regulatory requirements
- Customer account created in minutes, not days, with AI quality controls
- Automated onboarding: product activated, account type & limits configured, tailored terms set for high-risk clients



Perpetual KYC Monitoring

VII. Perpetual KYC Agent

- Continuous KYC updates as life and ownership change
- Agentic bot detects anomalies, fraud, and emerging risks instantly

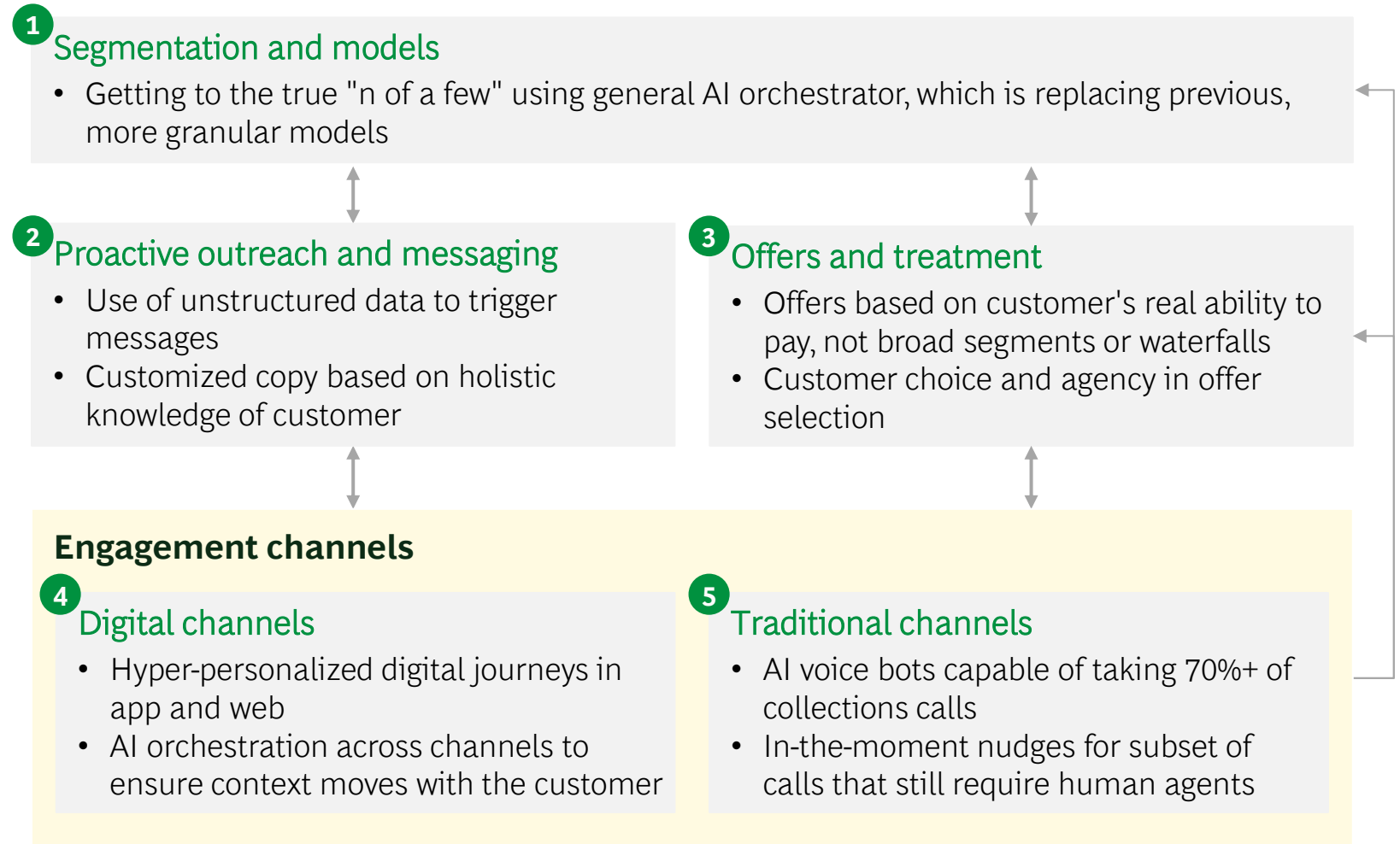
Up to **50%** cost savings realized in preventing financial crime in a European bank

Smart Collections | Key building blocks for best-in-class collections

Up to 50% Reduction in collections operating costs

10%-20% Net charge-off reduction and/or increase in amount recovered








Increase in **NPS and customer retention**



Note: All impact numbers based on value delivered from production-grade implementations at retail banks.

Harness Engineering | Transform digital builds from human-dependent manual processes to end-to-end agentic production

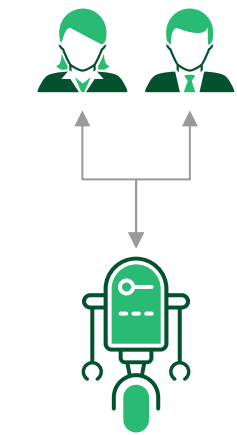
Harness engineering goes far beyond coding co-pilots into building **production-grade systems** end-to-end using AI agents. AI agents leverage **existing enterprise context** on tech architecture and systems, **validate** their own work, **recover from failures**, and **work within defined constraints**

	 Requirements	 Design	 Build	 Guardrails and traceability	 Test	 Deployment	 Operations
From	Manual BRD writing Manual requirement documentation	Manual design specifications Architects handcraft specs; no consistency, no traceability	Individual development effort Developers handcraft every feature from scratch	Individual development effort Ensure sufficient guardrails and build in observability	Individual development effort Inconsistent test case documentation	Risky manual releases High-ceremony, manual approvals, rollback fears	Reactive incident response Manual runbooks, tribal knowledge, high MTTR
To	BRD and user stories AI-generated requirements and backlog	Functional and design specs. AI-drafted functional specs. and API contracts	Agent-native build Agent-native specification-driven development	Agent-native build Agent-native specification-driven development	Test artifact generation Agentic test case generation and QA	Intelligent deployment Autonomous release orchestration	Proactive AI-Ops Anomaly detection and automated remediation

Result	50%+ Faster Time-to-market (BRD to Deploy)	90%+ Automated Test Coverage	3x Engineering Throughput	70% Reduced Toil	99%+ Deployment Reliability
---------------	--	--	-------------------------------------	----------------------------	---------------------------------------

Note: QA = Quality Assurance; MTTR = Mean Time to Recovery
Source: All impact numbers based on value delivered from production-grade implementations at retail banks

Future employees to have “Employee Alter-Ego” to perform all core and enabling tasks



Employee Alter-Ego
for each employee to perform all on-job activities

Fulfill role-specific activities

Perform other enabling tasks

	1	2	3	4	5	6
	Product and COE Teams	Credit and Ops Teams	Frontline Teams	Tech and Build Teams	Support Function Teams	Policy and Control Teams
Continuous, metric-driven innovation	Process files with accuracy and efficiency. Fast, high-quality credit decisions	Personalized pitches, quality service, and collections	Reimagined SDLC with high-quality code; builds, runs, and maintains tech stack	High efficiency and efficacy across HR, finance, marketing, and admin	Monitor and act on credit and fraud risk. Ensure regulatory and legal compliance	
<ul style="list-style-type: none"> • Design agent: BRDs, user stories, and A/B tests • Product persona agent: Iterate product solutions and propositions 	<ul style="list-style-type: none"> • Document intelligence: extract, validate, and reconcile • Credit analyst agent: auto scoring and decision-making 	<ul style="list-style-type: none"> • AI banker: personalized pitches and next-best action • Query bots: voice agents for inbound and outbound 	<ul style="list-style-type: none"> • Architecture review agents • Coding agents by skill type • QA and testing agent • Tech debt scanner 	<ul style="list-style-type: none"> • Live interview assistants • AI marketer: content generation and channel selection • Procurement agent 	<ul style="list-style-type: none"> • Regulatory agent: policy querying • Fraud agent: real-time monitoring and alerts • Compliance agent 	...
HR	Travel and Admin		Finance		IT	
<ul style="list-style-type: none"> • Career navigator: growth path mapping • Time, leave, and attendance agent • Attrition-risk agent: early disengagement signals 	<ul style="list-style-type: none"> • Travel manager: policy-compliant bookings and claims • Meeting scheduler: slots, rooms, and automated notes • Service bot: laptop and access card requests 		<ul style="list-style-type: none"> • Budget agent: spend tracking and overrun alerts • Payment agent: invoice tracking and claim processing 		<ul style="list-style-type: none"> • IT help desk: L1 query resolution and ticketing • Access provisioning: role-based setup and approvals 	

Shift from micro use cases to value-backed end-to-end function redesign



Use case centric design approach

Capability-first solutioning

Focus is on building GenAI capabilities/ applications, rather than redesigning business or customer experience (“hammer looking for a nail”)

Current tasks/ workflows automated

Legacy workflows and human tasks optimized. This restricts the opportunity to fundamentally rethink the business function

Solves multiple micro-problems

Focus is on automating multiple micro-problems. Tends to limit the impact on business function



End-to-end function reimagination

Value-based function redesign

Retail functions (e.g., service, credit, risk) are redesigned end-to-end with the focus on business outcomes

AI vs human roles and operating model

Re-define roles and operating model on what AI executes independently or with humans in the loop vs what humans execute.

Less is more

Focus is placed on fewer foundational redesigns, taken end-to-end to value, instead of many small use cases or GenAI tool implementations

AI delivers **step-change value** when it reshapes the function end-to-end

While redesigning a function, reorganize work of AI and humans based on the level of toil, reasoning, and expertise

Toil

Routine, repetitive tasks with high amounts of manual labor adding limited value

E.g., document completeness checks, system entries based on document (KYC forms, income proofs, and address proof)

70%-80%

Share of toil has the potential to be eliminated with AI

Reasoning

Tasks requiring contextual interpretation and decision-making

E.g., process-based decisions on query resolution

30%-50%

Share of reasoning has the potential to be automated

Expertise

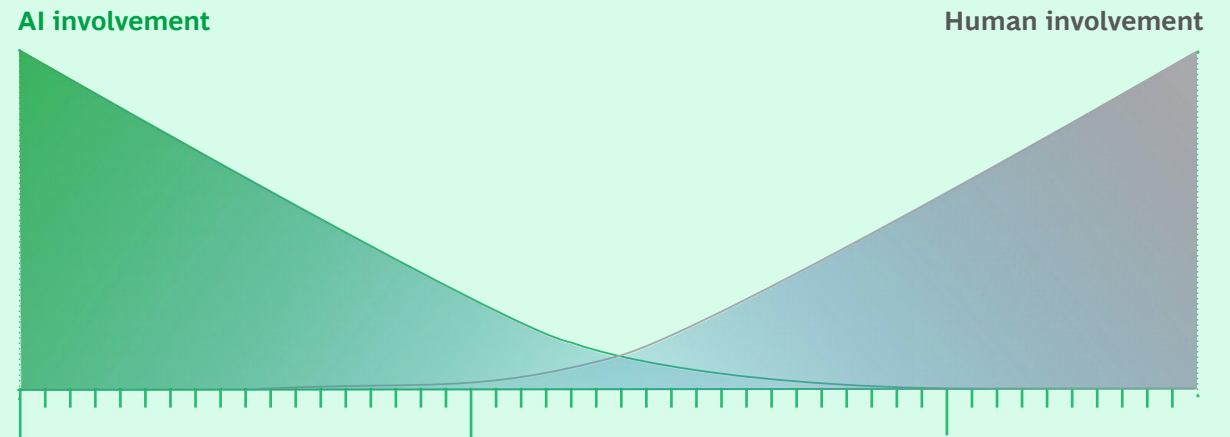
Complex tasks leveraging deep domain knowledge and specialized skills

E.g., fraud investigations and customer negotiations

AI can play the role of an **enabler** and **augment** expertise

Source: BCG case experience

Retail banks must actively decide which functions belong to AI, which require oversight, and which must remain with humans to **eliminate toil, automate reasoning, and augment expertise**



AI driven

- Service resolution for standard customer requests
- KYC data ingestion and validation

AI leads, human oversees

- Service exception handling
- RM next-best-action execution

Human leads, AI assists

- Complex credit exceptions and overrides
- Customer complaints and dispute resolution

Invest in cross-functional capabilities to power GenAI and Agentic AI

The age-old debate
is back for GenAI

Product

VS.

Platform

Start with a product approach

- Focus on a business function to build the right capabilities
- Ensure every capability is built for re-usability



Interaction insights

Enriches customer, agent, or interaction data

- Customer DNA
- Agent quality assessment
- Interaction disposition



Knowledge management system

Augments employees on bank-specific knowledge

- Product and onboarding queries by customers
- Service process queries by service agents



Document insight extraction

Enables job functions that require manual document analytics (e.g., operations)

- Document summarization
- Insight extraction from documents



Process intelligence engine

Captures fragmented human intelligence to power agentic use cases

- Process automatically captures all service scenarios (via screen recording and call analytics)



Tooling

Enables agentic action execution

- Integrations with internal systems
- Embedded functions (e.g., calculators and schedulers)



Content generation

Automates creation of content to boost productivity and efficiency

- Marketing collaterals creation
- Report generation (e.g., KYC and CDD)

Note: KYC = Know Your Customer; CDD = Customer Due Diligence
Source: BCG case experience

Manage AI risk and compliance from Day 1 by design, not as an afterthought

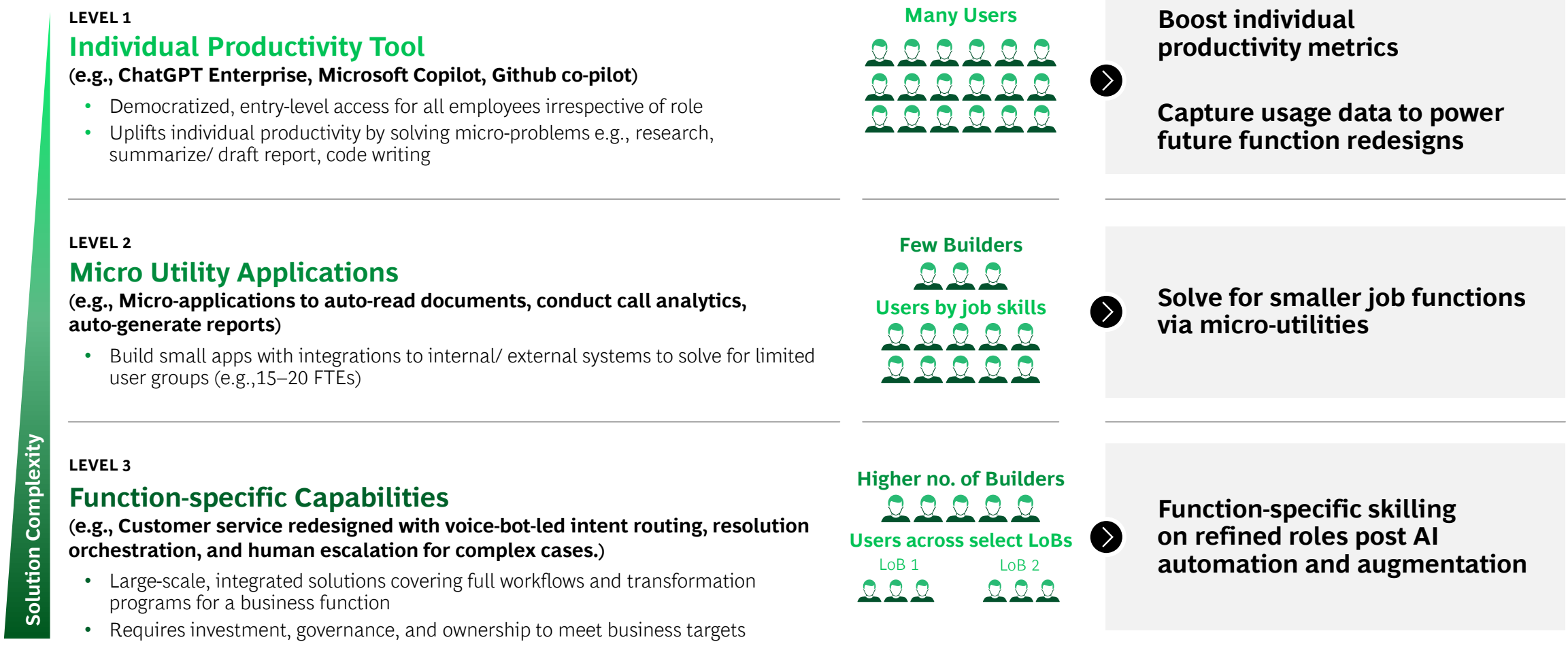
Dimensions

Components

<p>AI risk strategy</p>	<p>AI risk taxonomy AI-specific taxonomy, covering Agentic AI and agent-to-agent systemic risks. Built for enterprise-wide application, with scope to evolve with tech</p>	<p>AI risk strategy and appetite Board-approved thresholds tied to organization's AI ambition, not informally understood across teams</p>	<p>AI programs and policies (incl. Responsible AI program) Code of conduct, acceptable use, data handling, ethics, etc., actively monitored for adoption and refreshed as the landscape shifts</p>		
<p>AI risk governance and operating model</p>	<p>Risk operating model 1st/2nd/3rd line accountability clearly split across AI risk, security, privacy, and model risk; no overlaps, no gaps</p>	<p>Risk committees Dedicated AI risk committee laddering to ERM, with oversight anchored to AI initiatives not just models</p>		<p>Roles, responsibilities and decision rights Every function knows who approves what, at which risk threshold, and who owns residual risk if something goes wrong</p>	
<p>Risk management process (prevent and control)</p>	<p>AI intake, assessment and tiering Structured intake and risk-mapping routes every initiative to a governance tier (based on customer impact, data sensitivity, reversibility). Tracked SLAs and a live inventory ensure nothing slips through</p>	<p>PDLC integration and controls Risk-appropriate gates and controls from design to deployment, not only at launch</p>	<p>Controls design and automation Rules by agent type with action-space limits, multi-agent interaction controls, data passing restrictions. Level of automation defined with human oversight to evolve system</p>	<p>Resilience planning Rehearsed degraded-state procedures, incident playbooks and on-call rotations ensure rapid response and continuity</p>	<p>Third party risk management Vendors held to the same standards as internal models. Contracts enforce observability, auditability, and quality accountability</p>
<p>Risk management processes (detect and respond)</p>	<p>Testing and assurance Adversarial testing, second and third-order business impact assessment, and behavioral validation to confirm AI systems perform as intended across all conditions</p>	<p>Continuous monitoring Real-time tracking of accuracy, fairness, drift, hallucinations. Audit trail and active detection of cross-agent error propagation</p>	<p>Risk reporting and communication Portfolio-level AI risk visibility to board and regulators. Proactive failure-mode awareness before incidents</p>	<p>Issues management and remediation AI playbooks by failure type. Features to ensure fully traceable investigation. Response speed tracked as a risk metric</p>	

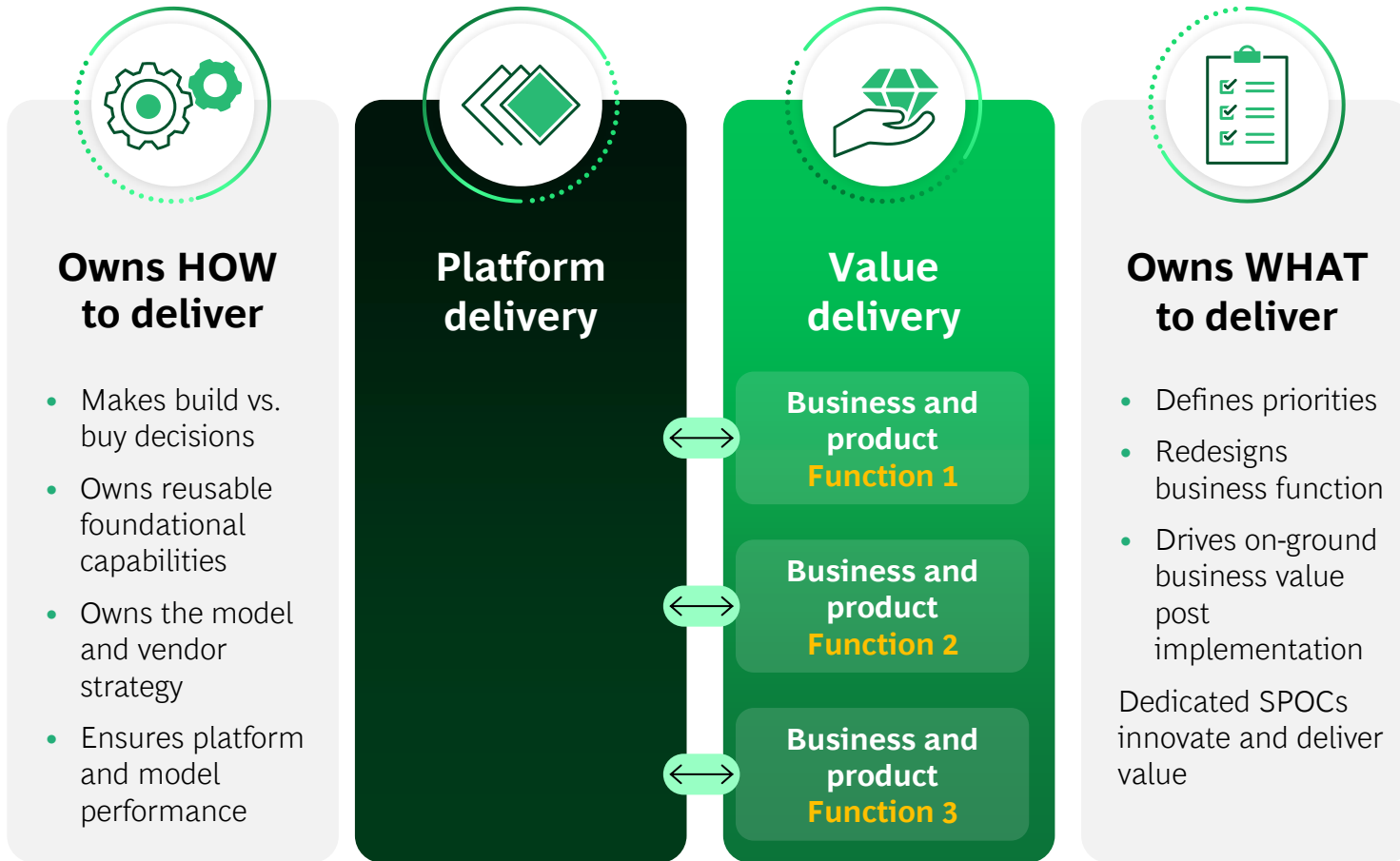
Note: ERM = Enterprise Risk Management Committee; PDLC = Product Development Life Cycle
Source: BCG case experience

Organization-wide skilling and adoption to be at 3 levels, from individual productivity to function-specific capabilities



Note: FTE = Full-time Equivalent; LoB = Line of Business
 Source: BCG case experience

Start with a centralized tech COE before moving to a more federated structure for function-specific transformations



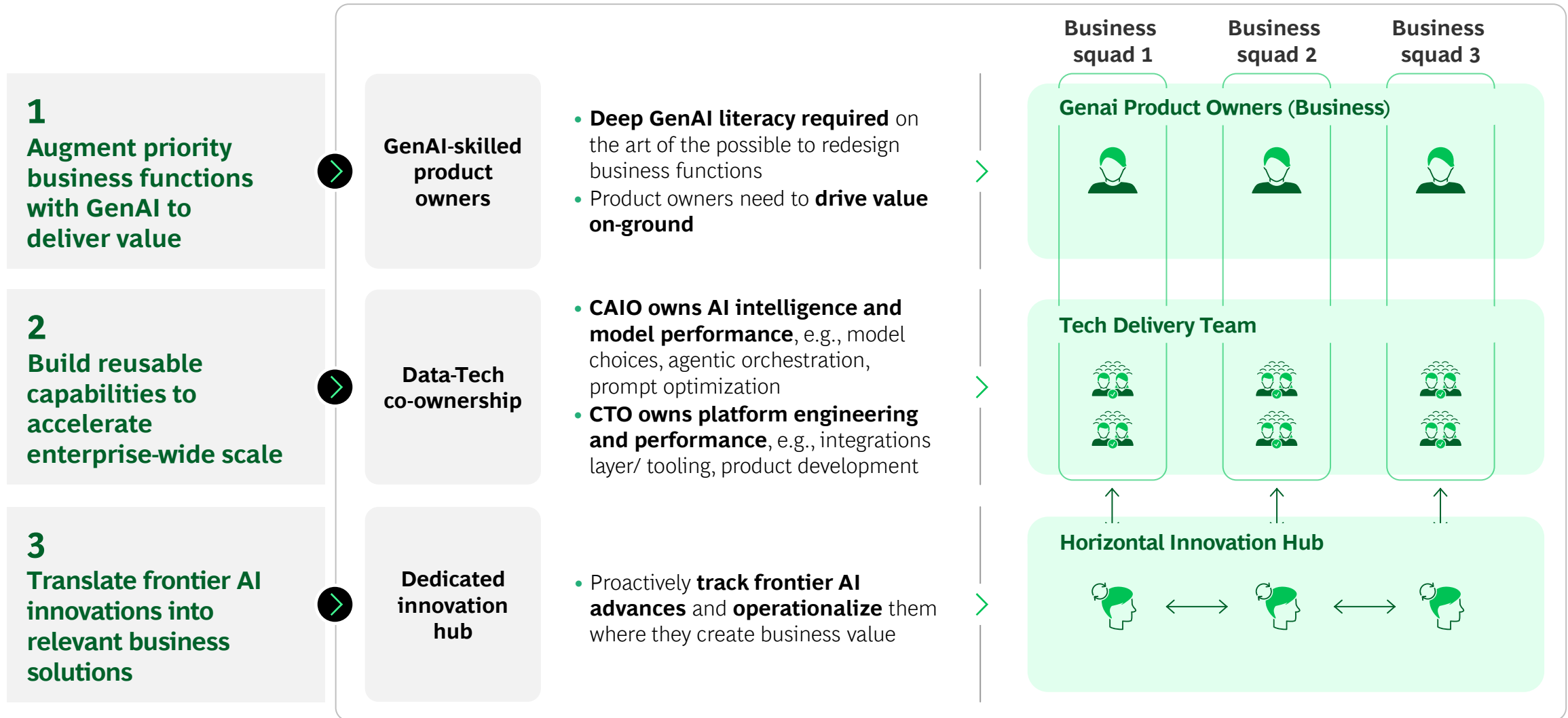
At all points in time, both the **GenAI platform and business function need to co-own** the business impact

At **early GenAI-maturity levels**, centralizing GenAI tech is a good starting point:

- Builds **reusable GenAI capabilities** in parallel, accelerating functional scale and overall maturity
- Establishes **enterprise-grade standards** across models, security, and risk to enable safe scaling
- **Concentrates scarce GenAI talent** to drive depth, quality, and repeatable excellence

As **GenAI maturity increases**, the **product tech must progressively evolve** toward a more **federated structure**, with stronger functional domain autonomy

Set up a GenAI COE for three key purposes



Note: CAIO = Chief AI Officer; CTO = Chief Technology Officer
 Source: BCG case experience

BCG experts | Key contacts for AI-First Retail Banks

Key BCG experts for AI in Financial Institutions



Bharat Poddar
MD & Sr. Partner
Retail Banking
Global Lead
New York



Nipun Kalra
MD & Sr. Partner,
AI @ FI Global Lead-
Retail Banking
Mumbai



Stiene Riemer
MD & Partner
AI @ FI Global Lead-
Wholesale Banking
Munich



Aparna Kapoor
MD & Partner
AI in Wealth
Singapore



Anne Kleppe
MD & Partner
AI in Compliance
Berlin



Javier Perez Moino
MD & Partner
AI in Customer Acquisition
Madrid



Lorenzo Fantini
MD & Sr. Partner
AI in Fin Crime
Milan



Lukas Haider
MD & Sr. Partner
AI in Operations
Vienna



Matthew Barton
MD & Partner
AI in Collections
Philadelphia



Peshotan Kapadia
MD & Partner
Conversational Banking
Mumbai



Semih Durmus
MD & Partner
AI in Credit
Minneapolis



Yogesh Mishra
MD & Sr. Partner
AI in Tech
Dallas

Author Team



Nipun Kalra
Kalra.Nipun@bcg.com



Sweta Jayasankar
Jayasankar.Sweta@bcg.com



BCG