Exceleron AI Layer

AI INTEGRATED, ON-PREM OR CLOUD

AI inside your operations, on your terms

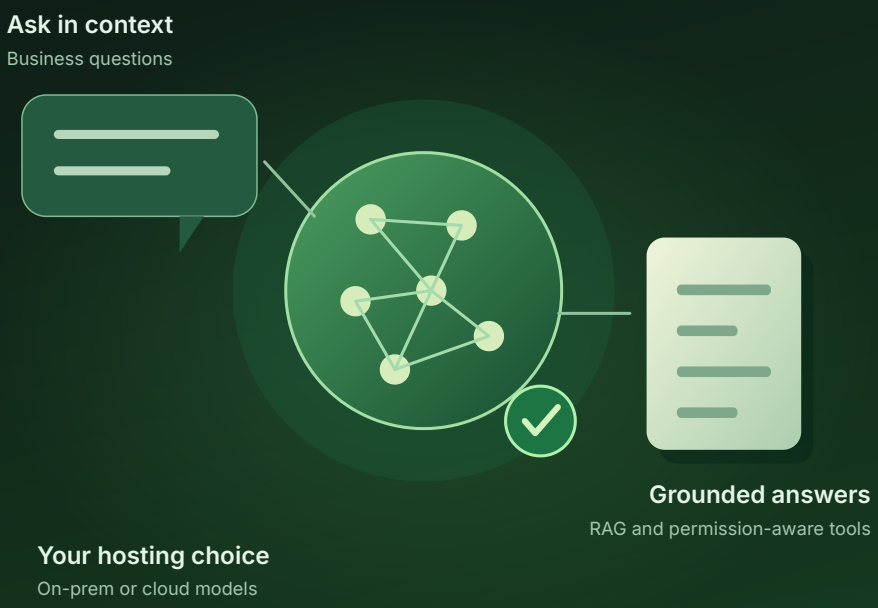
Every Exceleron product integrates with AI, deployed on your own servers or in the cloud. RAG grounds answers in your knowledge, MCP connects assistants to your tools, and people stay in control.

DEPLOYMENT OPTIONS
2

GROUNDED ANSWERS
RAG

TOOL CONNECTION
MCP

IN THE LOOP
Human



On-prem models
Self-hosted models for regulated teams.

Cloud models
Connect to a cloud provider where allowed.

RAG
Answers grounded in your own content.

OPERATIONAL CONTEXT

A clearer way to work

Every Exceleron product integrates with AI, deployed on your own servers or in the cloud. RAG grounds answers in your knowledge, MCP connects assistants to your tools, and people stay in control.

01

Data control

Sensitive data cannot leave freely.

02

Untrusted answers

Generic AI does not know your business.

03

Disconnected tools

Assistants cannot reach operational systems.

THE CHANGE IN DAILY OPERATIONS

Before	With AI Layer
– Public AI risks	✓ Your hosting choice
– Generic answers	✓ Grounded answers
– Copy and paste	✓ In-product AI
– Disconnected tools	✓ Connected tools
– Unclear access	✓ Permission-aware

AN OPERATIONAL SCENARIO

Answers connected to business context

A user asks a question that depends on internal documents or operational data. Grounded retrieval and permission-aware tools connect the request to relevant information.

Your
Choice of hosting

Grounded
Answers

Governed
Access

All
Products

CAPABILITIES 1

The capabilities that do the work

Explore the core functions of Exceleron AI Layer and the operational tasks they support.

01

On-prem models

Self-hosted models for regulated teams.

02

Cloud models

Connect to a cloud provider where allowed.

03

RAG

Answers grounded in your own content.

04

MCP

Standard connection to tools and data.

05

Data assistant

Read-only questions on live data.

06

Governed access

Answers respect user permissions.

DESIGNED AROUND THE USERS

Banks On-prem control.

Operators Ask network data.

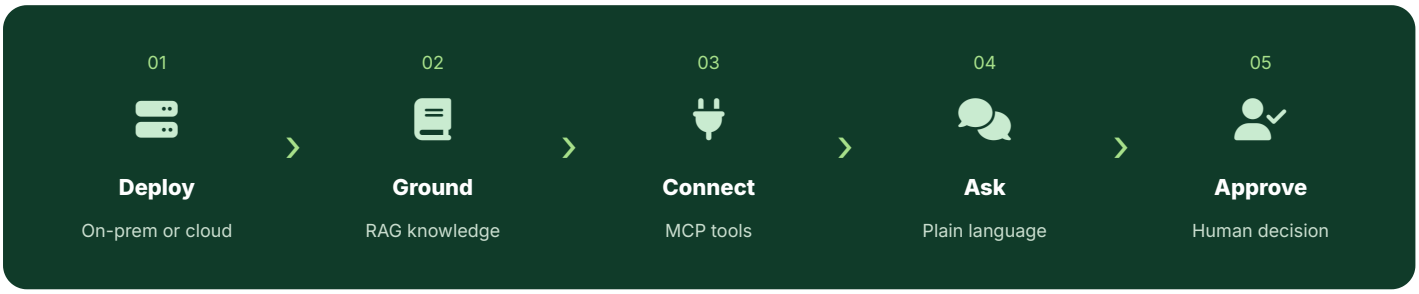
Security Governed access.

IT A choice of models.

WORKFLOW

From the first step to the outcome

A connected sequence makes the handoffs and responsibilities easier to follow.



01 Choose
Pick on-prem, cloud or both.

02 Connect
Link content and tools through RAG and MCP.

03 Scope
Apply each user's permissions.

04 Assist
Ask, summarise, draft and chart.

05 Decide
People review and approve.

A PRACTICAL WALKTHROUGH

Answers connected to business context

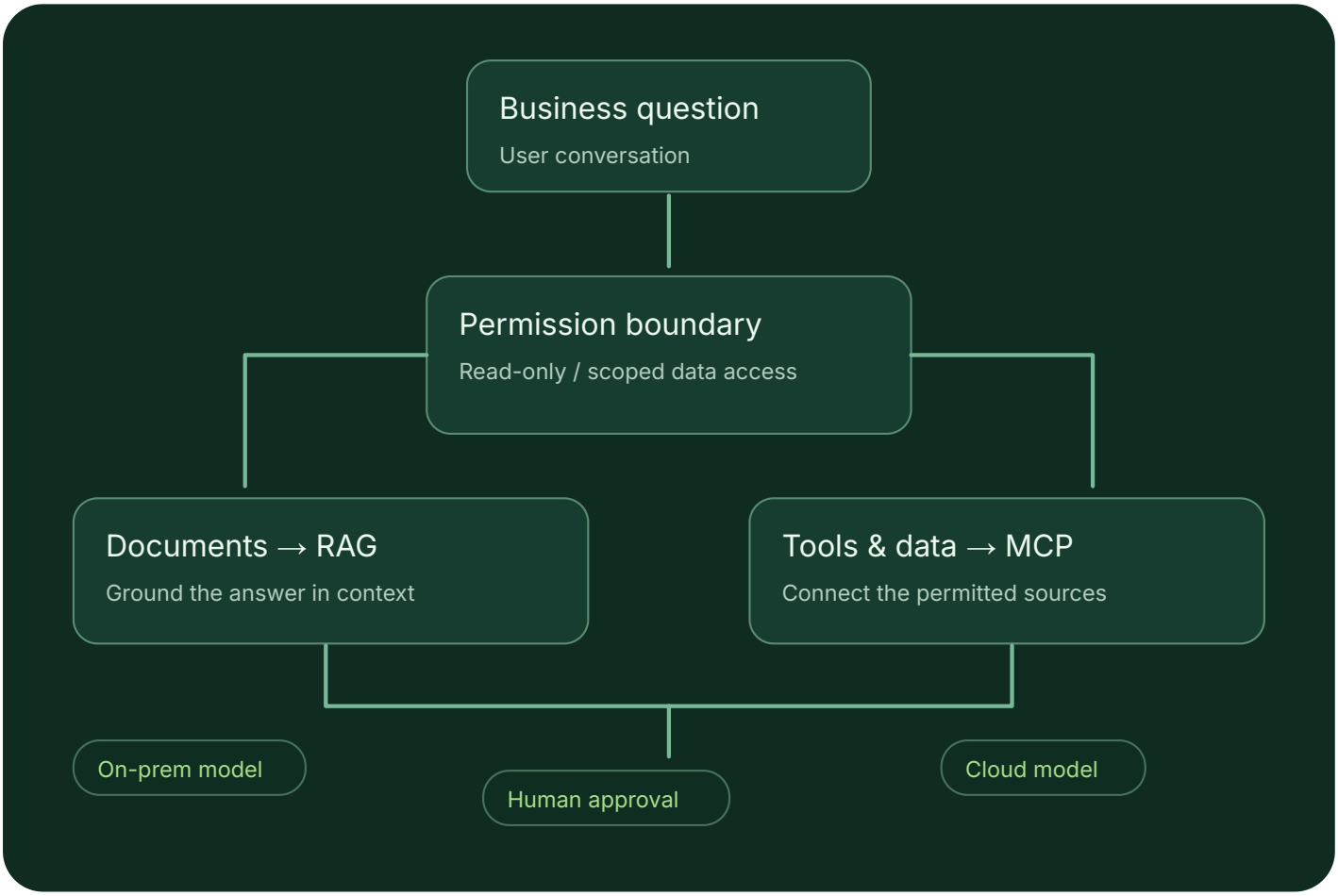
A user asks a question that depends on internal documents or operational data. Grounded retrieval and permission-aware tools connect the request to relevant information.

Review the retrieved context, permitted data scope and approval boundary before enabling actions. Agree whether model hosting should remain on-prem or use a cloud provider.

ARCHITECTURE & HANDOFFS

Connect answers to permitted context

Retrieval and tools bring relevant business context to the model, while the permission and approval boundaries remain explicit.



Conceptual product flow · Exceleron AI Layer

01

RAG connects the conversation to relevant documents.

02

MCP connects tools and data within the configured permissions.

03

Select on-prem or cloud hosting and agree when human approval is required.

DISCUSS YOUR ENVIRONMENT

Confirm the source systems, responsibilities and integration boundaries during the walkthrough. The diagram explains the product concept; the deployment scope is agreed for your environment.

TECHNOLOGY 1

Understand the connected pieces

Formats, protocols and tools in the product scope, with plain-language explanations for technical and business readers.

On-prem

Discuss On-prem in the context of your environment during the technical walkthrough.

Cloud

Discuss Cloud in the context of your environment during the technical walkthrough.

RAG

Discuss RAG in the context of your environment during the technical walkthrough.

MCP

Discuss MCP in the context of your environment during the technical walkthrough.

OpenAI compatible

Discuss OpenAI compatible in the context of your environment during the technical walkthrough.

Ollama

Discuss Ollama in the context of your environment during the technical walkthrough.

llama.cpp

Discuss llama.cpp in the context of your environment during the technical walkthrough.

Charts

Discuss Charts in the context of your environment during the technical walkthrough.

THE AI LAYER OPERATING FLOW



INTEGRATION PLANNING

Agree the source systems, supported versions, data ownership and access requirements during discovery. Validate the relevant interfaces against your environment before production rollout.

ASSURANCE & NEXT STEPS

Prepare for a successful rollout

Match the product scope to your people, systems and operating requirements.

PRODUCT ASSURANCE

- **Read-only data access**
- **Row-level security**
- **Permission-filtered schema**
- **Human approval**

PLAN THE ENGAGEMENT

01

Discover

Review your current process, source systems, users and constraints. Identify the operational problem the first phase should solve.

02

Demonstrate

Review the retrieved context, permitted data scope and approval boundary before enabling actions. Agree whether model hosting should remain on-prem or use a cloud provider.

03

Validate

Agree the applicable product scope, access controls, integration checks and acceptance criteria. Confirm deployment and support responsibilities.

04

Handover

Plan user guidance, operational ownership and support contacts. Confirm the reporting cadence and review process for the agreed scope.

START A CONVERSATION

Explore AI Layer with our team.

Bring your operational scenario. We will walk through the relevant workflow and discuss a suitable scope.

info@excelerongroup.com

www.excelerongroup.com · Lahore, Pakistan