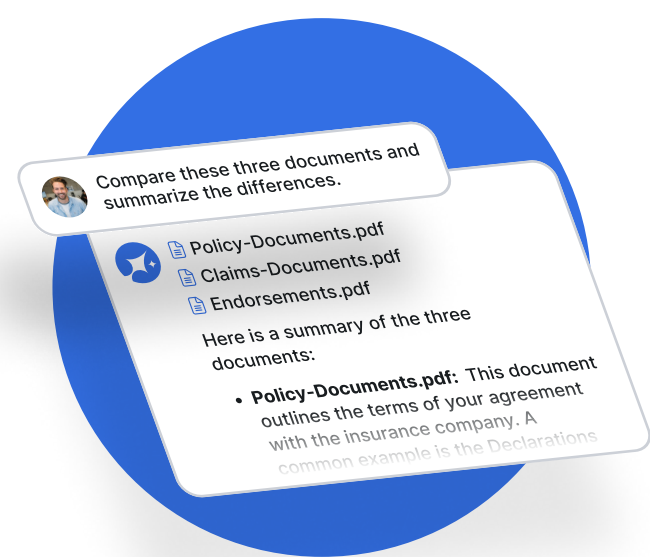# Generate Enterprise
by Iterate.ai

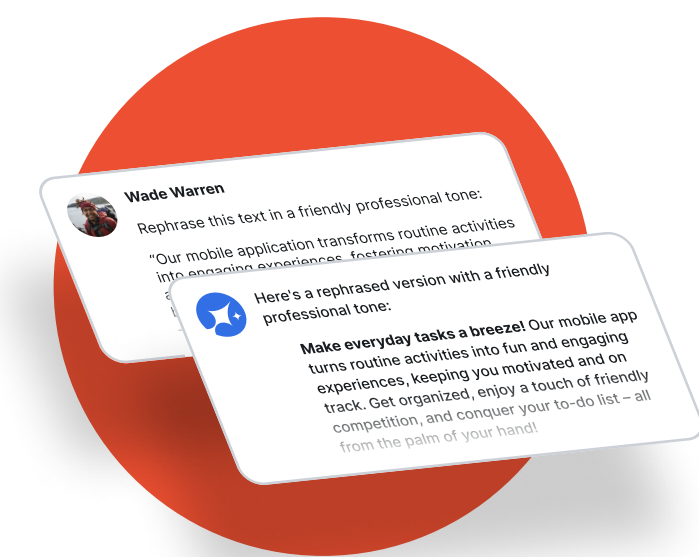## Manage LLMs & Generative AI for Your Enterprise

Enterprise Deep Document Search Powered by LLMs and Retrieval Augmentation Generation (RAG)

Interplay Generate Enterprise allows an organization to leverage multiple LLM applications and vector databases to search and retrieve complex structures and unstructured documents. The Text Augmentation feature can re-tone, rephrase, or rewrite the text. Service Pilot can summarize and suggest follow-up emails by cross-referencing internal data sources and private documents. Document Search will summarize data, analyze trends, and allow for LLM-powered Q&A. Generate Enterprise can operate on the cloud and on-premise and supports many hardware configurations such as Intel Xeon, Gaudi Accelerators, NVidia systems, and on Intel Meteor Lake (laptops).

Deep Document Search

Generate Enterprise advanced document search can ingest multiple documents and then perform complex smart searches on structured and unstructured data. Queries can be plain language prompts, configurable via the interface. Users can ask for comprehensive summaries, comparative differences between documents, unique or shared terminology and phrases. Generate Enterprise is unique in its ability to process large volumes of long (300+ page) documents rapidly and accurately.
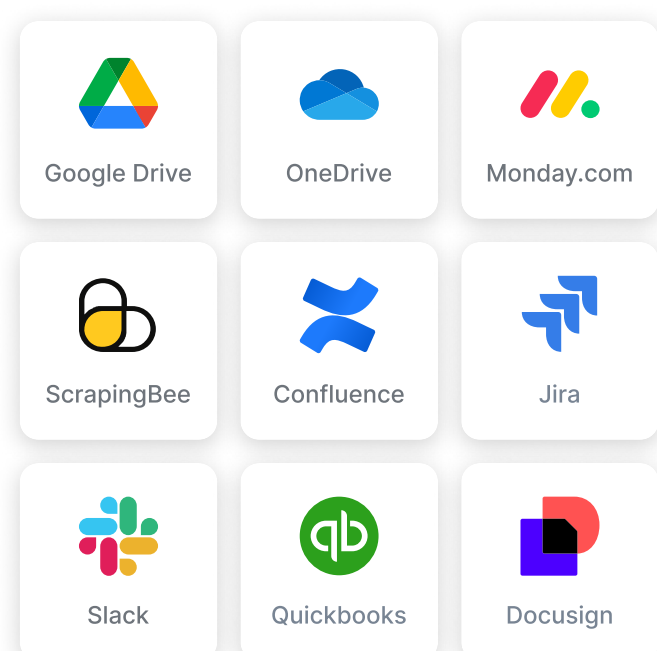
Service Pilot

Our Service Pilot can bring efficiency to support teams, customer relations, and research analysts with queries to multiple documents across multiple databases. This is all done without a single line of code: no advanced queries are required: all can be done with straightforward prompt engineering, allowing any advanced business user full leverage to the private LLM.

Enterprise System Integrations

Iterate Generate Enterprise integrates directly with a growing number of common enterprise platforms where documents, chat records, email, tickets, and contracts reside. Because Generate Enterprise is vendor-agnostic, there are no business restrictions on what can be sourced in the private LLM. Any platform with an API can be integrated.

Google Drive · OneDrive · Monday.com · ScrapingBee · Confluence · Jira · Slack · Quickbooks · Docusign

Enterprise Deployments

Intelligent Support

Private Documents

Vector Database

Document Search

Iterate.ai

Generate Enterprise provides Advanced LLM and RAG Functions Within Optimized Architectures

Iterate Generate for the Enterprise is built to be secure, flexible, and easily deployable. Built on top of the Interplay low-code application platform, Generate Enterprise is LLM agnostic, with the ability to support a number of foundation models. Generate runs completely locally within the data center or private cloud. Training documents can be uploaded with ChromaDB for vector database fine-tuning. Disparate data sources can be segmented into separate vector databases to improve accuracy across RAG generated responses. All referenced source documents are available with a simple click.

Secure Vector Database Architecture

Generate Enterprise has security incorporated at all levels for the highest security deployment requirements. This security-first approach is a leading advantage of Generate Enterprise compared to other mass-market AI Manager suites.

- Runs in private cloud environments or secure data centers
- Network firewalls, IP whitelists, existing IT security protocols are all respected
- Integrates with existing user roles and rights such as Auth0 and LDAP

Deployable in the Enterprise

Iterate Generate Enterprise is fully containerized within a docker, allowing standardized deployment and upgrade management using existing IT application management tools such as Kubernetes or Jenkins. With a relatively small and efficient footprint, Generate Enterprise can run on CPU+iGPU, or CPU+GPU+NPU architectures.

- Deployable as appliance, edge server, or cloud deployment
- Available from major cloud-vendor stores and data center integrations
- Compatable across CPU and GPU architectures

Flexible Data Integrations

Because Iterate Generate Enterprise is independent of any single walled-garden for data sources, the AI Manager is exceptionally flexible-- which is advantageous for large organizations with myriad data sources and differing needs across departments.

- Foundation model agnostic: pursue best-of-breed for rapidly evolving market
- Wide application integrations for documents, financial records, tickets, code
- Fine-tuning across data sets within separated vector databases
- 3rd party data source integrations
- Data vendor agnostic
- Trainable to match org voice

Iterate Generate AI Manager Platform			
Featured Technologies	• Document Search RAG • Image Tagging and Search	• Foundation Model Agnostic	• Robust Security Layers
Integrations	• Google Drive • One Drive • Slack	• Confluence / Jira • Quickbook	• Monday.com • Docusign
Deployment	• Single Executable for Windows • Appliance	• Edge	• Cloud Agnostic
Available LLMs	• GPT4 / 3.5 (API Key) • Bison (API Key) • Mistral 7B (Local option)	• Llama 8B (Local option) • Gemma 2B (Local option) • Phi (Local option)	• BLIP (Local option)
Hardware Compatibility	• CPU Intel	• CPU+iGPU (Intel)	• CPU+GPU+NPU (Meteor Lake)
Footprint	• Size: 150MB (without LLM)	• Ram: 16GB	
Markets Served	• Consumer • Small, Medium Businesses	• Enterprise	• Existing Intel Business Customers

Technical Requirements

Footprint:

150MB without LLM
16GB RAM

Models:

GPT4 / 3.5 (API)
Bison (API)
Mistral 7B (Local)
Llama 8B(Local)
Gemma 2B (Local)
Phi (Local)
BLIP (Local)

Deploy Options:

On-Premise
Appliance
Cloud-Agnostic
CPU+GPU

Iterate.ai