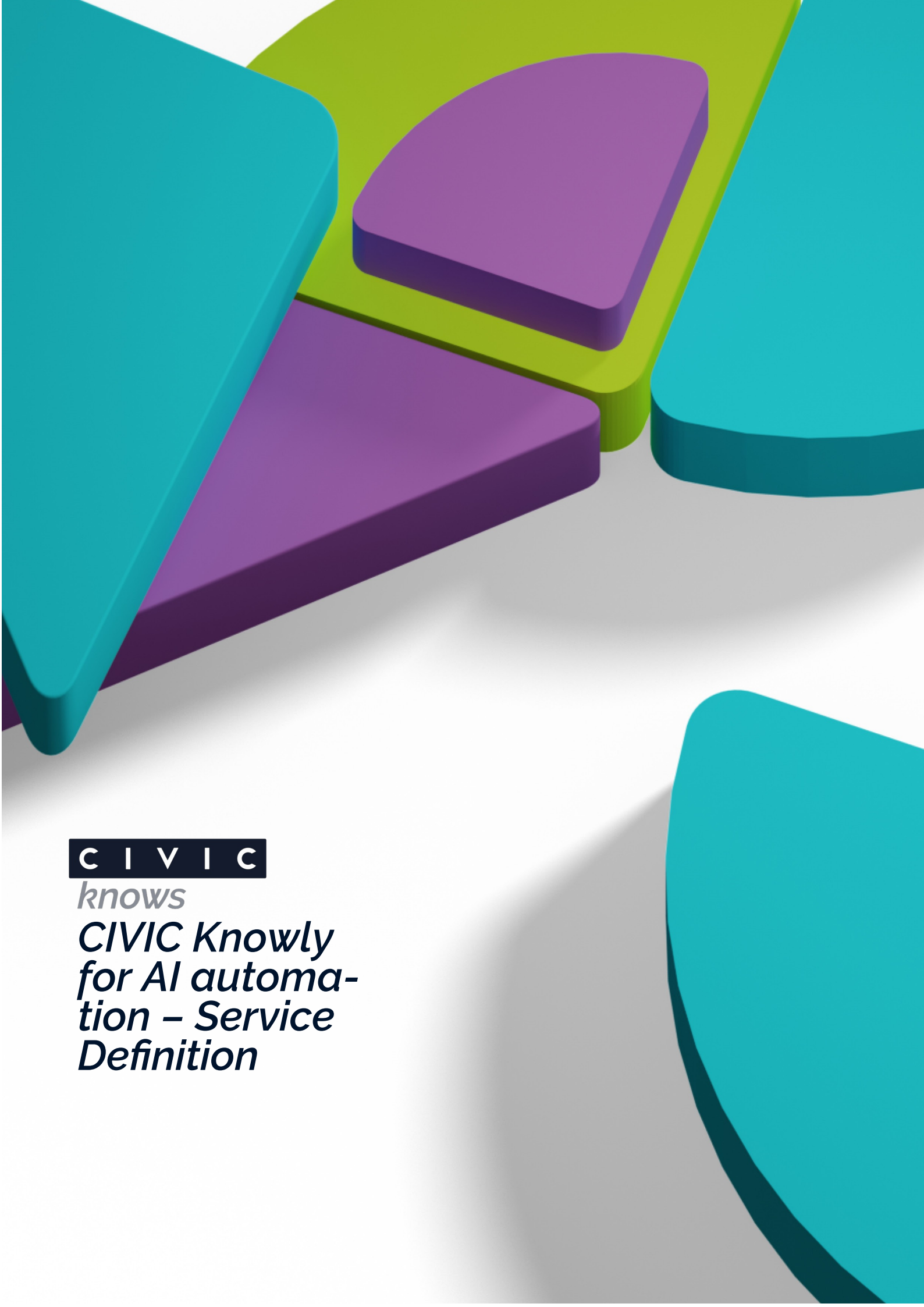

CIVIC

knows

CIVIC Knowly for AI automation – Service Definition

CIVIC

Contents

CIVIC Knowly for AI automation - Service Definition 3

Key features: **3**

Benefits **4**

Who uses Knowly **5**

Managed Hosting 6

Contact 7

CIVIC

CIVIC Knowly for AI automation - Service Definition

Knowly is our AI agent application bridging the gap between a vast amount of information and accessing the right information fast. Using Large Language Models (LLMs) by employing the Retrieval-Augmented-Generation (RAG) approach, Knowly achieves natural conversation access to documentation. All persistent components, such as databases, vector stores, document repositories, and embeddings, are managed within CIVIC's controlled and secure infrastructure, ensuring consistent data management and privacy safeguards. All inference operations occur entirely on-premises using a locally hosted model. If the remote configuration is used, only the composed request, which includes the user's prompt and the retrieved context (e.g., relevant documentation excerpts), is transmitted securely to an external model endpoint for processing. This design allows seamless switching between local and remote models while maintaining uniform data handling practices, privacy controls, and application logic.

Key features:

Front-end

- Chat Widget: Component that provides the chat interface for users to ask questions and view responses. Handles conversation history, user input, and displays AI-generated answers in real-time.
- Administration Area: Dedicated space to manage the core intelligence that drives Knowly. It offers the ability to add and organize relevant documentation and links, define user access permissions, monitor system performance, manage the entire system and ensure the knowledge base remains current and accurate.
- User Management: Create/Edit user accounts and manage individual access levels. Track user actions. Organize teams.
- Knowledge Base Management: Add resources via links and/or upload documents and manage the information which eventually forms Knowly's knowledge base.
- Parameterization and Monitoring: Performance monitoring, audit logs, tool usage tracking, global configurations, parameters adjustment (e.g., LLM selection, LLM's settings, temperature, topK, topP, documentation split strategy etc.), LLM's system message instructions and other administration tools.
- Character, Language, personalisation

Server Side Modules

- Chat: Endpoints and services which receive the user questions via HTTP, validate the requests, and initiate the RAG pipeline. Returns structured responses with answers and source references.
- Document: Endpoints and services for document lifecycle operations. Handles upload, update, and deletion requests, and coordinates document processing workflows.
- User Management: Endpoints and services for managing user accounts
- System Management: Endpoints and services which handle all available settings
- Monitoring: Endpoints and services to track usage and user actions, performance indicators.
- RAG Orchestrator: Central coordination service that manages the complete inference

pipeline. Sequences retrieval, prompt building, and answer generation steps while handling errors and fallbacks.

- Data ingestion and preprocessing: This stage involves extracting and transforming the raw data from the Knowledge Base into a usable format for Knowly.
- Embedding Model: Converts text chunks into numerical representations, which are called embeddings. These embeddings are then stored into a vector database.
- Hybrid Search: Combines results from both vector and keyword searches. Merges, ranks, and de-duplicates chunks from both systems to provide comprehensive context retrieval.

Storage

- Vector Database (vectors): Stores text chunk embeddings and enables fast semantic similarity search. Maintains vector indexes for efficient nearest-neighbour queries.
- Index Repository (clear text): Search engine that indexes chunk text for full-text search capabilities. Supports advanced querying, filtering, and relevance ranking based on keyword matching.
- MariaDB (metadata): Relational database that stores document metadata, chunk relationships, and system configuration. Maintains referential integrity and supports complex queries.

LLM APIs

- Local Ollama Service: External process that runs local LLM and embedding models. Provides HTTP API for text generation, managing GPU/CPU resources for AI inference.
- External API: Provisioning of an API to securely communicate with an external API, like OpenAI

Benefits

Using an AI module to retrieve information from documentation offers several valuable benefits, especially compared to traditional keyword-based search or manual lookup. Here are the key advantages:

- Faster Information Discovery. Knowly understands natural-language questions, meaning users don't have to know exact keywords or file locations. This dramatically reduces time spent searching through large or complex documentation.

- Improved Accuracy and Relevance. Knowly models interpret context, intent, and semantics, returning results that match what the user means, not just what they type. This leads to fewer irrelevant results, more precise answers and better handling of ambiguous queries.

- Summarization and Explanation. Knowly summarizes long sections of documentation or explain technical details in simpler terms. This is useful for onboarding new team members, quickly understanding complex APIs or configurations, generating step-by-step guides from raw documents.

- Cross-Document Linking. Knowly pulls information from multiple documents and presents a unified answer, even when the relevant details are spread across different pages, PDFs, documents, wikis and internally linked knowledge bases.

- Context-Aware Responses. Knowly tailors answers based on what's specifically asked by the user, taking into consideration their role or level and the system/environment in use.
- Reduced Cognitive Load. Users don't need to remember the details to ask the question. Knowly handles retrieval and interpretation, allowing users to focus on problem-solving.
- Automation and Integration. Knowly can integrate with IDEs, CI/CD pipelines, support tools, chatbots and more to enable automated troubleshooting, documentation lookup, and contextual suggestions.
- Documentation house-keeping. Knowly can generate missing links, rewrite unclear sections, flag outdated or inconsistent information and improve the overall quality of documentation over time.
- Scalable and Always Available. Knowly supports thousands of users simultaneously, providing instant responses without human intervention—ideal for enterprise documentation or large ecosystems.
- Enhanced User Experience. Overall, Knowly creates a more intuitive and conversational way to interact with documentation, similar to asking an expert rather than searching a static knowledge base.

Who uses Knowly

It's widely used by:

- Large and medium companies use Knowly to allow employees or customers to ask questions and instantly get answers from internal documentation, knowledge bases, or customer-facing content.
- Customer-service platforms and help-desk providers integrate Knowly so that website visitors can get answers 24/7 without needing a human rep.
- Companies using AI chat widgets as lead-gen / marketing tools: bots that answer product or service questions, filter leads, collect contact info - all while backed by the company's documentation or website content.
- Internal teams in organizations use Knowly to manage and retrieve info from documentation, meeting notes, project docs, wikis etc
- Tech companies, SaaS providers, or any organization producing extensive documentation (APIs, user manuals, internal standards) use Knowly to help users find relevant info quickly.
- Help-desk / support-oriented businesses that deliver online customer support, combining Knowly with their knowledge base, live chat and ticketing
- Small to mid-sized businesses - including those without big support staff - use Knowly to handle frequently asked questions (FAQ), manage leads, and provide 24/7 basic support rather than hiring full-time staff.
- Researchers, engineers, or teams handling complex, evolving documentation use Knowly to retrieve and synthesize information across multiple documents - especially when knowledge is scattered across formats (PDFs, docs, wikis, meeting notes).
- Help-desk agents or support staff: rather than memorizing everything or manually searching docs, they rely on Knowly to quickly fetch the right article or answer from the knowledge base, speeding up response time and reducing errors.

Managed Hosting

It is important that your hosting platform is scalable and secure to allow for peak demand traffic and also provide the peace of mind of knowing your website is in safe hands. At CIVIC, we offer our own managed hosting services: scalable cloud hosting, dedicated servers, and multi-server infrastructures from our secure ISO27001 accredited Tier 3 datacentres in Edinburgh and Glasgow. Our main hosting environment is based at Pulsant in South Gyle, Edinburgh, which is currently the Scottish Government's approved hosting supplier.

To ensure scalability, resilience and security of your website, we recommend that you consider hosting your site on our general purpose clustered hosting platform. Benefits of this service include:

- Operational management, monitoring and backup procedures
- Service Level Agreement (SLA) for network uptime, service response time, support and cover, in and out of office hours
- Physical security, resilience, availability and security of the network
- High bandwidth

Our end-to-end hosting service proposition and the Service Level Agreement which underpins it enables us to offer managed infrastructure services to support both current and future requirements in delivering your web systems. We have built an enterprise level hosting environment to support our clients' business critical systems and ensure maximum 'up time' for their websites and applications. This environment has been designed to create a robust and resilient infrastructure and conforms to current IT best practice.

Our highly secure and robust hosting infrastructure serves both generalist and specialist hosting needs. We also operate a public cluster suitable for websites and applications with fairly common requirements, plus private single purpose platforms and customised infrastructures for those who require it. There are no single points of failure and business critical hardware is replicated so that it can be deployed immediately if production systems fail. We deliver and take responsibility for a fully managed infrastructure, supported 24/7 from multiple sites in Europe, offering a combined service-desk which incorporates Technical Support, Service Delivery Management, and continuous proactive Quality Improvement.

Contact

CIVIC

12 South Charlotte Street
Edinburgh
EH2 4AX
E: info@civicuk.com

www.civicuk.com
Registered in Scotland SC221925
VAT Registration No 774906100
CIVIC Computing Ltd