

NeoAi

Your Vision, Realized with NeoAi^{✦✦}

YOUR GATEWAY TO THE AI ERA

About Us

Leading by Passion. Driven by Innovation.

Overview



4000+
Employees



10
Global Offices



9
Development
Centers



100k sq. ft.
Office Space

Collaborations



Clientele



1500+
Clients



50+
Countries



22+
Industry
Verticals



85%
Client Retention

Achievements



500+
AI Solutions
Delivered



70%
Faster
Go-To-Market



200+
Enterprise
Clients



24/7
Expert
Support

What We Do

Level up your business with our AI expertise.

AI Strategy & Consulting

We help enterprises define and accelerate their AI journeys with tailored roadmaps, platform selection, and responsible AI advisory. From ideation to ROI analysis, our experts guide innovation at every stage.

Unleash the power of AI with vision and clarity.

AI Engineering & Product Development

We build scalable, production-ready AI solutions—from custom LLM agents to generative copilots and intelligent assistants. Our engineering teams specialize in integrating AI into enterprise apps and services.

Turn bold AI ideas into high-impact digital products.

Data & AI Infrastructure

We design modern data pipelines and deploy AI-first architectures with vector stores, retrieval engines, and MLOps at the core. Our infrastructure ensures your models are fast, secure, and cloud-optimized.

Build smarter systems on solid AI foundations.

Generative AI & Automation

We transform business operations with advanced GenAI systems for content generation, smart document workflows, and automated decision-making. From voice to vision, we unlock multimodal intelligence at scale.

Reimagine productivity through intelligent automation.

The Tools in Our Belt

The technology behind our AI expertise.



Agentic Frameworks

Multi-agent coordination, tool-use, and autonomous reasoning

AutoGen CrewAI

LangChain LangGraph OpenAI

Agents SDK Guardrails AI

Google Vertex AI Agents



Models & APIs

Top-tier LLMs and APIs powering our core solutions

Claude Gemini Mistral

LLaMA GPT-4 / ChatGPT

GitHub Copilot Perplexity Sonar

Replit Ghostwriter



Data & Retrieval

Semantic search, vector storage, and document indexing

Elasticsearch Milvus

Pinecone Weaviate Qdrant

Opensearch PostgreSQL

Vespa Cursor



Voice & Speech

Speech synthesis, transcription, and conversational AI

Amazon Polly Azure Speech

Google TTS/STT Whisper

Kaldi PlayHT ElevenLabs



Synthetic Media

AI for video, image, and avatar generation

Runway Kaiber

Sora DALL-E 3 Midjourney

Leonardo HeyGen Pika

Jina DreamMachine

Stable Diffusion XL



Workflows

LLM workflows, DAGs, and automation frameworks

Langflow Airflow Dagger

Wandb Gumloop

Temporal NN Prefect

Zapier Wootric



Monitoring & Control

Eval frameworks, safety nets, and feedback tools

LangServe PromptLayer

Rebuff OpenAI Evals DeepEval

TruLens Guardrails Logfire



Dev Tools

AI copilots, autocompletion, and dev productivity boosters

Claude Cursor Codeium

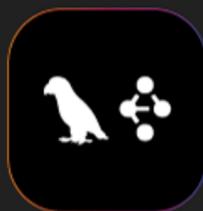
Ghostwriter Codex

Windsurf Bolt WO.dev

Our Proven AI Stack

Tools we trust to scale pilots into production.

Agentic Frameworks



LangGraph

Declarative multi-step flows with memory, retries, and branching.



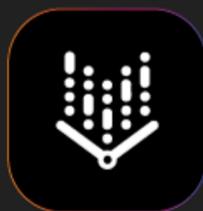
CrewAI

Modular, role-based agents for orchestrated tasks.



OpenAI Agent SDK

Lightweight agent setup, provider-agnostic.



Google Vertex

Enterprise-grade agent workflows with built-in compliance.

Retrieval & Memory



LlamaIndex

Fine-tuned RAG pipelines with schema-aware loaders.



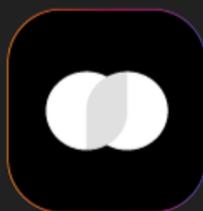
Weaviate

Semantic vector search for structured and unstructured data.



Pinecone

Low-latency, scalable vector DB for real-time retrieval.



ChromaDB

Fast, local-first indexing ideal for edge or offline scenarios.

Ocrhestration & Workflows



N8n

Visual automation builder with hundreds of built-in connectors.



LangChain

Rapid prototyping for LLM chains, agents, and tools.



Dagger

CI/CD-friendly automation that scales with DevOps needs.



ReAct Pattern

Tool-calling pattern for complex, interpretable agent workflows.

Crafting Robust AI Solutions

A curated ecosystem powering seamless AI delivery.



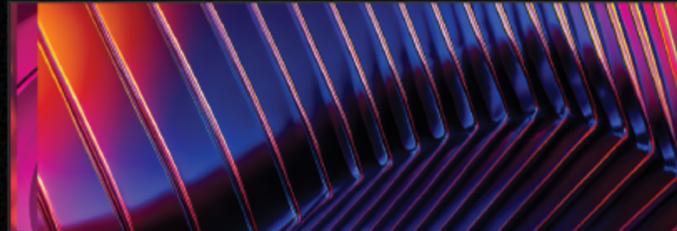
Embeddings

Fine-tuned DistilBERT.
Semantic comprehension and
search indexing



Search Strategy

Hybrid (semantic + keyword)
Balance recall and precision in
RAG



Reranking

Jina Reranker v2.
Relevance reordering based
on user context



Web Scraping

BeautifulSoup + Selenium +
Crawl4AI – Full-scale structured
data extraction



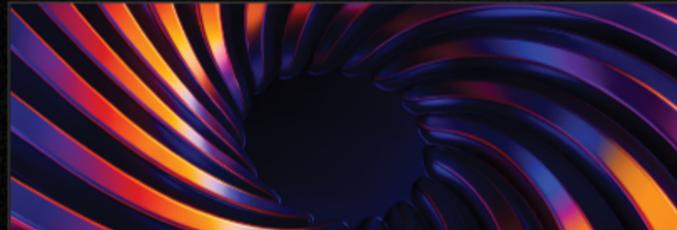
LLM Router

LiteLLM – Multi-model routing
with cost control



Document Parser

pyMuPDF – Extract content
from PDF, Word, PPT



Vector DB

MongoDB Atlas + Qdrant –
Scalable vector storage for
production RAG

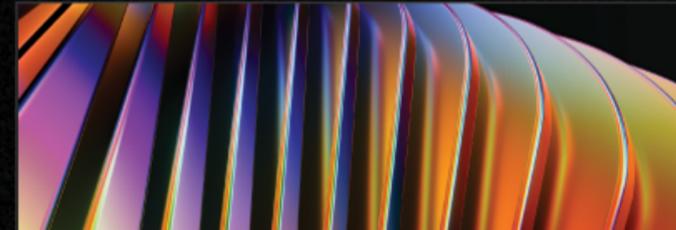
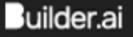
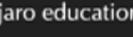
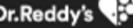
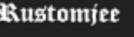


Image Processing

GPT-4o + CoIQwen – Multimodal
analysis (text, image, layout)

Selected Clientele

22+ Industries. 100+ Fortune 500 Brands.

| Financial Services | Media & Entertainment | Digital Products & Platforms | Telecom | Professional Services | Manufacturing | Travel & Hospitality | eCommerce & Retail | Healthcare | Education & Learning | Real Estate | Supply Chain & Logistics |
|---|---|---|--|---|---|---|---|---|---|---|---|
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |

Global Presence



Development Centers

- Mumbai (HO)
- Mumbai (Prabhadevi)
- Navi Mumbai
- Pune
- Bengaluru
- Ahmedabad
- Noida
- Indore
- Kolkata
- Chennai
- Hyderabad
- Dubai

International Offices

- USA
- UK
- UAE
- Australia
- Sweden
- Germany
- Singapore
- South Africa
- Philippines

Contact Us

✉ info@neosofttech.com

📞 +91-22-61500600

🌐 www.neosofttech.com