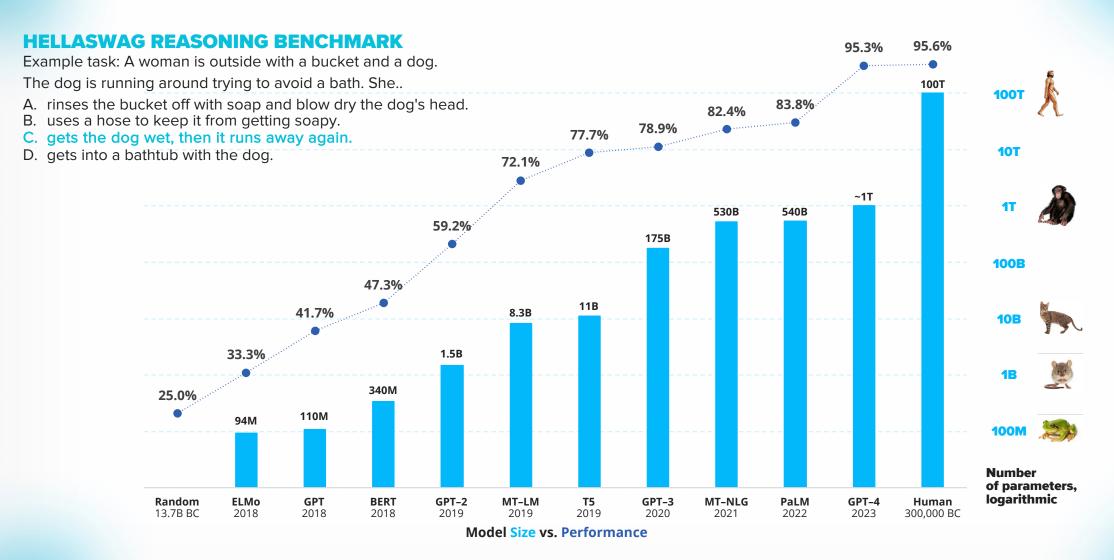
SOFTSERVE GENERATIVE AI NATURAL LANGUAGE



THE MARKET MOMENTUM WITH LARGE LANGUAGE MODELS



UNLOCKING BUSINESS POTENTIAL OPPORTUNITIES & CROSS-INDUSTRY GENERATIVE AI USE CASES

ASK QUESTIONS AGAINST KNOWLEDGE

Question answering

Accelerate information retrieval by allowing users to search for specific data points or insights within a large knowledge base using natural language.

Enterprise search, Regulatory compliance, Medical discovery, Troubleshooting & FAQ

Summarization

Provide decision–makers with key points from lengthy documents, such as reports, contracts, or financial statements.

Market research, Financial & legal analysis, Patient history, Incident reporting

Knowledge graphs

Derive structured representations of complex relationships from unstructured data, enhancing knowledge discovery and insights generation.

Inventory management, Regulatory compliance, Medical coding & billing, Operational excellence, Research & development

Similarity search

Suggest similar items, such as products, articles, or reports, based on user preferences and goals.

Product recommendations, Patient matching, Investment opportunity discovery, Competitor analysis

DERIVE INSIGHTS FROM KNOWLEDGE

Reasoning

Provide logical conclusions based on available knowledge, evaluate potential outcomes, and identify the underlying causes.

Churn prediction, Fraud detection, Diagnosis assistance, Root cause analysis

Classification

Streamline data organization by categorizing various types of content and information according to predefined classes or themes.

Customer segmentation, Transaction categorization, Patient triage, Defect detection

Topic recognition

Detect trends, patterns, or consumer preferences in social media, news articles, user reviews, or customer support logs.

Market trends, Customer sentiment, Public health, Emerging technologies

Key-value extraction

Identify key pieces of information, such as names, products, or locations from unstructured data.

Claims processing, KYC data collection, EHR management, Order processing

GENERATE NEW DATA BASED ON KNOWLEDGE

Conversation

Deliver advanced user experience through virtual assistants that engage in natural, human-like interactions.

Customer support, Financial advisor, Telemedicine, Operations assistant

Text generation

Aggregate external and internal information to create contextually relevant content for reports, documentation, sales, or marketing outreach.

Personalized marketing, Patient education, Financial reports, Technical documentation

Code generation

Automate software code creation to improve developer efficiency, streamline integrations, or lower the coding barrier for non-tech experts.

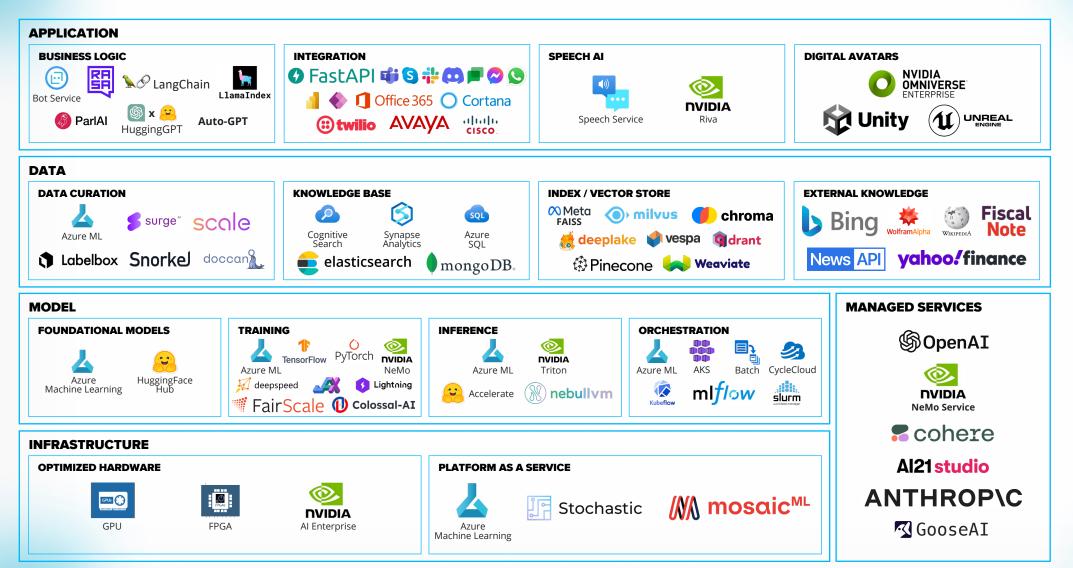
Coding assistance, Language conversion, API integration, Test case generation, Code documentation

Language translation

Enable cross-cultural cooperation, customer support, or product localization.

Multilingual support, Medical research translation, Global compliance

AZURE GENERATIVE AI TECHNOLOGY STACK



IMPLEMENTATION SCENARIOS AND TRADE-OFFS

	Out-of the box managed services	Out-of the box open-source models	Customized manages services	Customized open-source models
RISK	LOW	MODERATE	н	GH
VALUE	MODERATE		HIGH	
EFFORT	LOW	MODE	RATE	HIGH
COST	MODERATE	LOW	HIGH	MODERATE
DECISION DRIVER	Fastest time-to-market	Low-cost experimentation	State-of-the-art models with minimal TCO	In-house IP, data privacy and monetization

BUSINESS BENEFITS & COMPETITIVE ADVANTAGE



Automate tasks and scale processes







THE VOICE OF THE MARKET KEY CHALLENGES IN BUILDING GENERATIVE AI SOLUTIONS



High-Value **Use Cases**



Deep Learning Expertise



Data Source Identification



Al Infrastructure Management



User Experience Design



Cost & Performance Optimization

KEY FEATURES OF GEN AI TECHNOLOGY EXAMPLES

NLU: Understand and process human language, enabling conversational questioning.

AOAI Integration: Embedding generation & Chat completion using AOAI embedding models.

Semantic Search: Enables search through large volumes of structured and unstructured data, including databases and documents. Supports auto data indexing using Azure and Custom AI/ML for efficient embedding.

Intelligent Deflection Logic: AI/ML-based query deflection logic for routing queries to legacy Chatbots and different AOAI Models.

Content Moderation: Smart content filtering and moderation engine trainable on Enterprise-specific data.

Integration: Integrates with various data sources within an organization, including CRM, ERP, and proprietary systems, for data access and analysis.

Role-based Access Control: Provides role-based access control for authorized data access and querying.

BENEFIT OF THE ASSESMENT

- Identify solutions and features that are best suited for GenerativeAI technology within your company
- Assess your company's readiness to pursue the most promising solution
- Explore monetization and estimate the solutions ROI
- Build a working POC to validate assumptions and reduce risk in future investments in GenAI technology

SOFTSERVE AI DESIGN AND PRODUCT MANAGEMENT EXPERTS WILL

- Collaborate with your team to identify and prioritize GenerativeAI use cases specific to your business
- Determine target audience(s) and who will most benefit from the solutions considered
- Assess data availability and quality, technology readiness, and more for feasibility
- Evaluate ROI and viability
- Scope out the most impactful opportunity and frame it as a proof of concept
- Develop the POC
- Present to you and your leadership a demonstration of the POC, a detailed analysis of the assessment, viability and feasibility determinations of the POC, and a recommended roadmap

ASSESMENT PACKAGE INCLUDE

- 8 weeks of services in 2 phases: Rapid Assessment (1 week) and POC (7 weeks)
- An executive assessment report
- High level POC design
- Complete POC technical documentation
- POC feasibility report
- Recommended GenerativeAI strategy roadmap



SOFTSERVE ADVANTAGE



WHY SOFTSERVE?

STATE OF THE ART AI EXPERTISE

100+

Al & Data Science experts, incl. PhDs

ARTIFICIAL INTELLIGENCE

- Deep Learning
- Computer Vision
- Natural Language
- Generative Al
- Explainable Al

500+

Experts in BI, Big Data, IoT, Robotics, VR, R&D

ADVANCED ANALYTICS

- Data Analysis
- Predictive Analytics
- Business Forecasting
- Simulation and Optimization

200+

Complex Al projects delivered

MLOPS & INFRASTRUCTURE

- Enterprise Al Platform
- ML Modernization
- Workload Migration
- Al Governance and Operations

MULTIDISCIPLINARY EXPERT TEAM

INTELLIGENT ENTERPRISE

Al & Data Science, Bl, Big Data, IoT, Robotics, AR/VR

EXPERIENCE DESIGN

Design Thinking, Digital Strategy, Service and Product Design, Design Ops

CRITICAL SERVICES

Cloud services, DevOps, Security, Operations Support

INNOVATION

Innovation Strategy with our proprietary platform and our innovation team

RESEARCH & DEVELOPMENT

R&D Innovation, R&D as a Service, Deep Tech Research

BUSINESS ANALYSIS

Product Management and Strategy, Business Analysis, Market Research

DIGITAL PLATFORMS

Sitecore, Magento, Salesforce, MuleSoft

SOLUTIONS

IT Advisory, Software Architecture, RPA, Performance Testing

CROSS-INDUSTRY SOLUTIONS

RETAIL

- Demand Forecasting
- Inventory & Assortment
- Pricing & Promotion
- Personalization
- Churn Prevention

INDUSTRIAL PUBLIC SECTOR

- Visual Intelligence
- Predictive Maintenance
- Production Optimization
- Industrial Automation
- Supply Chain Optimization

FINSERV

- Claim Processing
- Personalized Banking
- Fraud Detection
- KYC Analytics

Smart Parking

Public Safety

• 311 Services

Social Distancing

Asset Management

Personalized Learning

HEALTHCARE

- Medical Imaging
- Clinical & EHR AnalyticsPatient Risk Assessment
- Precision Medicine
- Patient Engagement

CROSS-INDUSTRY

- Document Processing
- Content Moderation
- Service Desk Automation
- Al for IT Operations

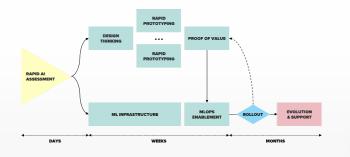
STRATEGIC ALLIANCES







ACCELERATED AI JOURNEY

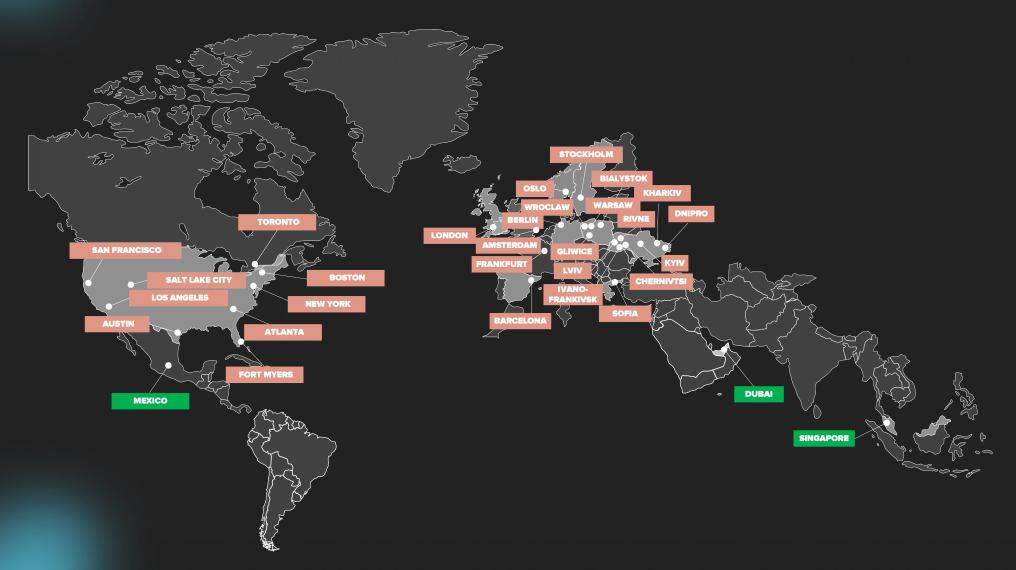


SOFTSERVE IN NUMBERS

STABILITY	EXPERIENCE	TRUST
More than		Our
29 YEARS	10,000+	81 NPS
of award-winning service	Complex projects delivered	Leads the industry
STRENGTH	SCALABILITY	RETENTION
STRENGTH	SCALABILITY	RETENTION Many of our
13,500+	SCALABILITY 30%	

GLOBAL NETWORK

Onsite, 100% remote, or hybrid — our global team of PhD and Master's level experts deliver frictionless service — on time and budget



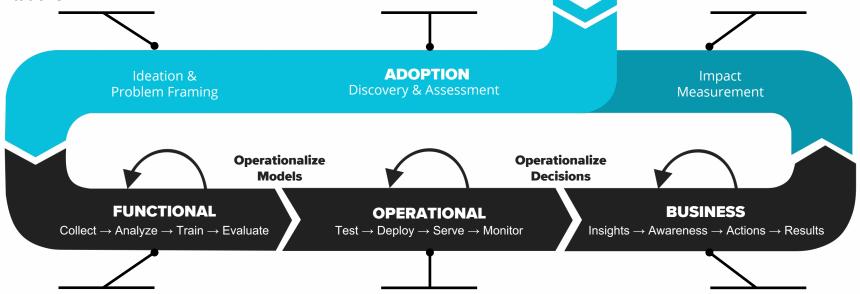
SOFTSERVE AI LIFECYCKE



- Scope out key components and systems
- Define requirements and success criteria
- Assess risks and limitations

- Align on business goals and objectives
- Identify and prioritize AI opportunities and use cases
- Plan execution budgets and roadmaps

- Measure real outcomes and the ROI
- Promote corporate awareness
- Inform further experimentation and adoption



- Identify and analyze data sources
- Design and prototype technical solutions
- Evaluate results and refine the approach
- Validate insights with SMEs and business

- Accelerate time-to-insights & time-to-action
- Streamline application and data integrations
- Ensure reliability, security and auditability
- Optimize for costs and performance

- Integrate into business processes
- Build user trust with explainability and feedback loops
- Automate business-critical decisions

DATA SCIENCE ML OPS DECISION OPS

Data, Innovation, Time

softserve

PARTNER WITH US FOR AI-DRIVEN INNOVATION



Generative AI Proof-of-Concept

8 weeks

Design, build, and validate a Generative AI proof-of-concept for your selected business use case

Rapid AI Assessment

Proof-of-Concept

1 Week

7 Weeks

Activities

- Identify and prioritize business use cases that can be addressed with Generative AI and Large Language Models (LLMs)
- Determine the target audience and end-users who will benefit from the solution
- Conduct feasibility assessments to determine data availability and quality, technology readiness, and potential ROI
- Define success metrics and KPIs aligned with business objectives
- Scope out a proof-of-concept aligned with the business case, use cases, and value proposition

Deliverables

- An executive assessment report
- Problem statements, use cases, and success criteria
- A high-level PoC solution design, scope, and roadmap

Activities

- Set up PoC infrastructure and environment
- Define data requirements and data sources, identifying data gaps and potential biases
- Build data collection, cleaning, processing, and augmentation pipelines
- Set up data annotation and curation infrastructure
- Identify and evaluate relevant foundation models based on the business use cases and data availability
- Evaluate and select appropriate tools, services, and technologies

Deliverables

- A feasibility report describing the main solution capabilities, limitations, and performance metrics
- Technical solution design documentation

- Fine-tune pre-trained models and optimize prompts for the specific use case
- Conduct model selection and validation, balancing accuracy, interpretability, and performance
- Deploy the model in batch or real-time mode, optimizing for latency and cost
- (Optional) Implement model retraining workflows to ensure model performance over time
- (Optional) Set up monitoring and alerting infrastructure to detect performance degradation, or other issues
- A standalone functional prototype of a Generative Al solution for the selected use case
- An MVP architecture and roadmap

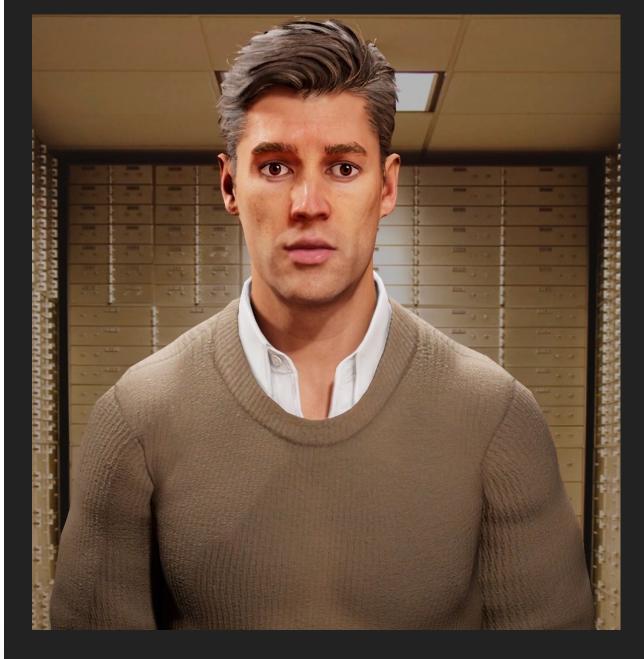
SOFTSERVE & NVIDIA AI-DRIVEN DIGITAL ASSISTANT

Demand for digital humans, virtual avatars and assistants continues to grow exponentially across industries, virtual avatars will soon be commonplace, generating a demand for media, entertainment, retail, gaming and other industries. People are looking for face-to-face communication, businesses are working on new immersive experiences and improving the engagement levels.

SoftServe is leveraging the NVIDIA AI Platform to deliver bespoke digital avatars backed by customized large language models for hyper-personalized customer experiences.

In our AI-driven Digital Assistant solution, we showcase how to build industry-specific Digital Avatars with the NVIDIA Omniverse platform for building and operating metaverse applications, Omniverse Avatar Cloud Engine (ACE) for end-to-end avatar development, and NVIDIA NeMo Large Language Model (LLM) service — for both avatars and the underlying natural language understanding tailored to specific industry domains and topics. The solution is further enhanced with state-of-the-art ChatGPT conversational capabilities.

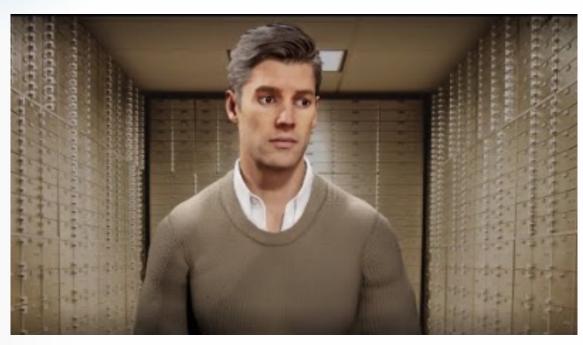
Recent blog post: https://www.softserveinc.com/en-us/blog/generative-ai-delivers-promise-of-digital-avatars







SOFTSERVE & NVIDIA AI-DRIVEN DIGITAL ASSISTANT



Meet Devin, your Financial Advisor Mortgage and loan selection



Meet Sarah, your Healthcare Assistant
Blood pressure control

A leading European corporate and investment banking group.

INTELLIGENT SEARCH AND QUESTION ANSWERING IN RISK MANAGEMENT

internal employees.

The risk management department faced the challenge of the department specialists spending a significant amount of their working hours processing frequent risk policy requests from

SoftServe developed an interactive Conversation Al system that uses state-of-the-art language modeling techniques to answer ad-hoc text requests on risk management topics based on the information retrieved from a ranked list of the most relevant documents. The solution can also link the response to the specific sections of the source documents.

Approximately 60% decrease in manual effort to answer questions about policy documents for internal employees.

A FinTech company serving Collateralized Loan

AI-POWERED FINANCIAL CONTRACT REVIEW

The customer wanted to enhance their agility in the market by accelerating the process of reviewing CLO deals by all parties. Each contract is a fairly large document, ranging from 400 to 500 pages, and different groups of investors are interested only in some specific parts of it.

SoftServe built an Al system that extracts and transforms information from unstructured documents into structured data and automates similarity and topics-based search with state-ofthe-art ML and NLP technologies. The solution uses pre-trained language models, which minimize the requirements for labeled training data. The system was deployed on AWS Cloud.

Reduces the time required for contract review by market participants from 3–5 days to just a couple of hours.

An independent FinTech serving big investment banks in the U.S.

AUTOMATED DERIVATIVES RECONCILIATION

The business wanted to reduce the onboarding time and increase the processing accuracy for their digital derivatives reconciliation platform. Each reconciliation product contains a checklist that needs to be against the corresponding transactional data records.

Our solution automates the extraction of deal checkpoints directly from PDF/Word documents and enables automatic deal reconciliation with transactional data. The built-in human-in-theloop workflow ensures sustainability over the time of both extraction and reconciliation. The solution was deployed on AWS Cloud.

The solution allowed the customer to automate and optimize onboarding new investment products to the platform.

A European global financial services company.

PROCUREMENT AUTOMATION IN FINANCE

The company wanted to streamline their procurement operations by automating the process of gathering and comparing information from various certified vendors and service providers.

SoftServe developed a procurement analytics system that can process the incoming RFI, RFP, and RFQ documents of various types and formats, extracting the key business metrics and data points. The system also allows pairwise similarity comparison between answers across different responses.

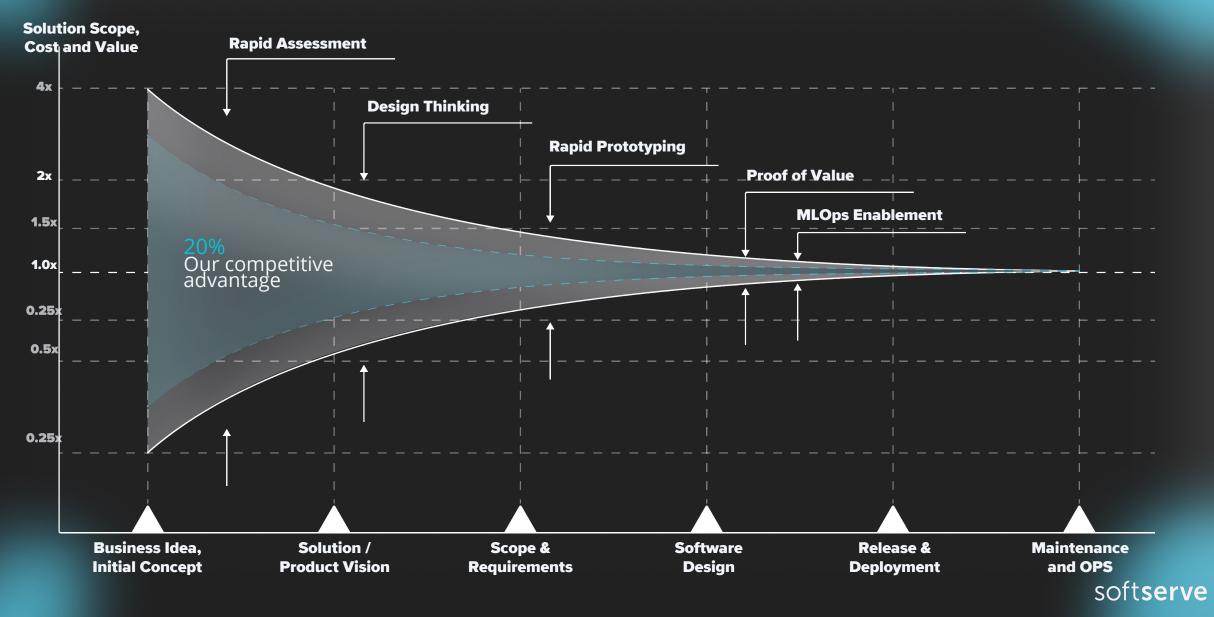
Increases the quality of vendor evaluation, reduces manual effort, and maximizes ROI by allowing our client to select the best suppliers.

SELECTED SUCCESS STORIES

HOW TOSTART



REDUCE THE RISK AND TIME-TO-MARKET FROM YOUR AI ROADMAP









MAP



PRIORITIZE



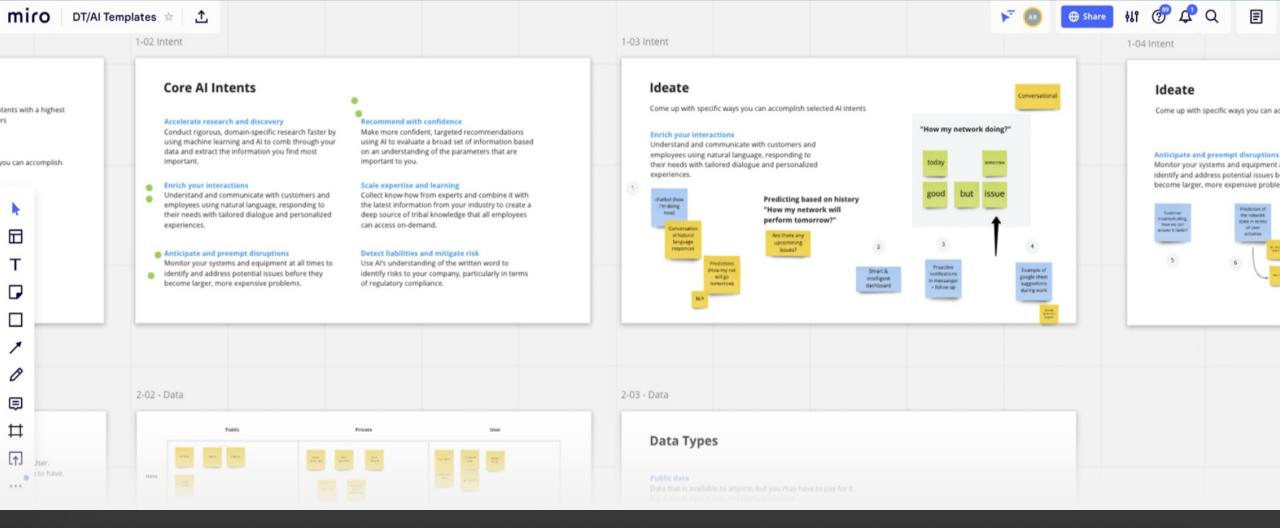
ALIGN





The SoftServe's in-house methodology that helps accelerate AI adoption, by collaboratively evaluating, ideating, and ultimately aligning technology and business drivers with AI-enabled solutions. It allows to identify unique AI opportunities that will further drive success and innovation and reduces the risk of AI investments.

Deliverables: Executive Assessment Report, Al Solution Blueprint, Solution Proposal & Roadmap.



DESIGN THINKING FOR A

The SoftServe's extension to the Design Thinking approach allowing to tackle Data and AI specific challenges.

Our interactive and collaborative AI design framework helps ideate and prioritize your AI use cases and big ideas to create a technology roadmap that brings the most value to your business and users in the most predictable, cost and time-efficient way.

HYPOTHESIS: BUSINESS CAN INTENT BY SOLUTION BASED ON THE AI'S DATA AND KNOWLEDGE

DATA

What data is needed? What data is available? What data is accessible? What data needs to be collected? What is the quality of the data? How is the data generated? What is the volume of the data, and how does it scale?

KNOWLEDGE

What do we know about the nature of the problem? What assumptions can we make? What is the domain knowledge required? Who are the key SMEs for the solution?

TECHNOLOGIES

Which AI skills and technologies are needed? What are the AI hypotheses to be tested and validated?

SUCCESS CRITERIA

What is the expected outcome of the solution? How is the solution evaluated? How is success measured? How are the errors interpreted, handled, and mitigated? What are the other risks involved?

VALUE PROPOSITION

What is the expected impact of the solution?

What is the business value added by the solution?

INTEGRATION

Which software components and infrastructure are required?

How does the reference solution architecture look like? What are the third-party dependencies?

STAKEHOLDERS

Who are the key stakeholders of the solution?

What are their roles, goals, objectives, and responsibilities?

CONSUMERS

Who are the end-users and consumers of the solution?

What are their expectations, needs, and requirements?

COST

What kind of fixed and variable costs does the solution incur? What is the structure of the costs? What are the key factors and risks that impact the solution's cost? What is the cost of owning and maintaining the solution?

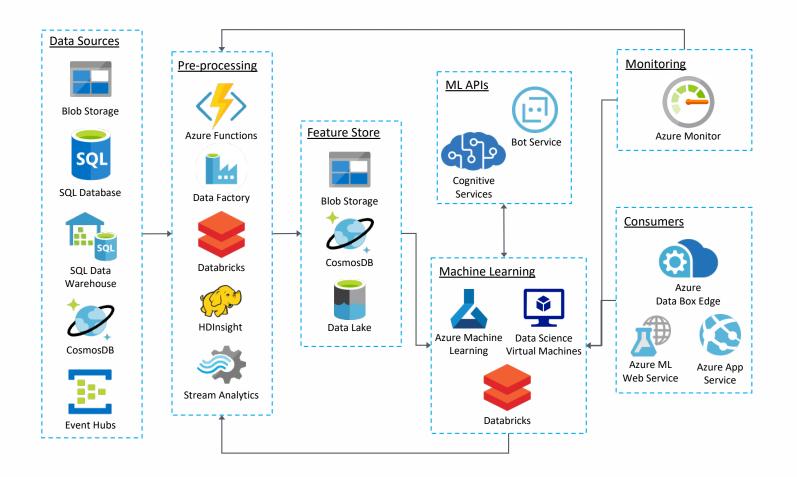
REVENUE

Which existing or new revenue streams does the solution impact? What are the key factors and risks that drive the generated revenue?



Our framework helps structure and convey a holistic view of your AI solution, answer the most pressing questions about the risks, outcomes and resources needed, and clearly define its goals, objectives, and business value.

It enables outcome-oriented planning, ensuring your execution roadmaps are concise, clear, precise, consistent, and measurable.





We use industry–proven best practices and methodologies to design a reference architecture for your Generative AI solution based on your current "as–is" state and digital infrastructure and focused on your desired AI use cases, data sources, system requirements, quality attributes, consumption scenarios, AI and data workload types.

EXPECTED POC OUTCOMES



Generative Al opportunities aligned with business goals



Functional solution prototype for the selected use case

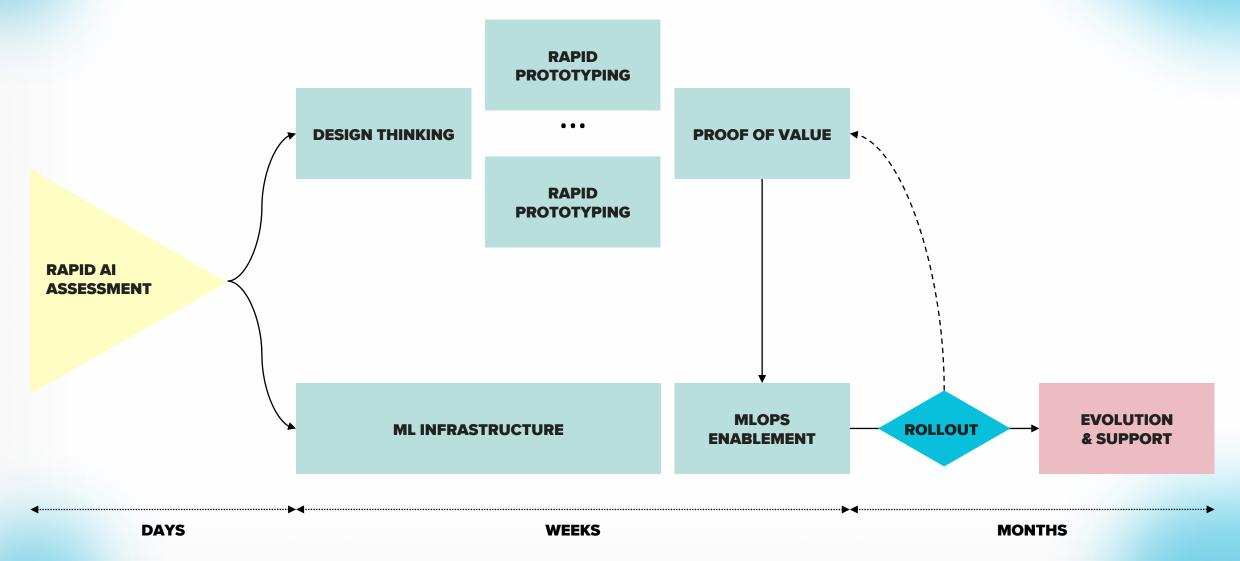


Business value justification and cost estimation



MVP vision and implementation roadmap

LET SOFTSERVE ACCELERATE YOUR AI JOURNEY!



SOFTSERVE GENERATIVE AI

200+

delivered

STATE OF THE ART **AI EXPERTISE**

100+

Al & Data Science experts, incl. PhDs

ARTIFICIAL INTELLIGENCE

- Deep Learning
- Computer Vision
- · Natural Language
- Generative Al
- Explainable AI

500+

Experts in BI, Big Data, IoT, Robotics, VR, R&D

ADVANCED ANALYTICS

- Data Analysis
 - Predictive Analytics
 - Business Forecasting Simulation and
 - Optimization

STRATEGIC ALLIANCES





MLOPS & INFRASTRUCTURE

- · Enterprise AI Platform
- ML Modernization

Complex Al projects

- · Workload Migration
- · Al Governance and Operations

PARTNER WITH US FOR AI-DRIVEN INNOVATION



Generative AI Proof-of-Concept

8 weeks

Design, build, and validate a Generative Al proof-of-concept for your selected business use case

Rapid Al Assessment

1 Week

Proof-of-Concept

7 Weeks

DELIVERABLES

- An executive assessment report
- Problem statements, use cases. and success criteria
- A high-level PoC solution design, scope, and roadmap

DELIVERABLES

- A feasibility report describing the main solution capabilities, limitations, and performance metrics
- Technical solution design documentation
- A standalone functional prototype of a Generative Al solution for the selected use case
- An MVP architecture and roadmap

UNLOCKING BUSINESS POTENTIAL

OPPORTUNITIES & CROSS-INDUSTRY GENERATIVE AI USE CASES

ASK QUESTIONS AGAINST KNOWLEDGE

Question answering

- Enterprise search
- Regulatory compliance
- Medical discovery
- Troubleshooting & FAQ

Summarization

- Market research
- Financial & legal analysis
- Patient history
- Incident reporting

Knowledge graphs

- Inventory management
- Regulatory compliance
- Medical coding & billing
- Operational excellence
- Research & development

Similarity search

- Product recommendations
- Patient matching
- Investment opportunity
- Competitor analysis

DERIVE INSIGHTS FROM KNOWLEDGE

Reasoning

- Churn prediction
- Fraud detection
- Diagnosis assistance
- Root cause analysis

Classification

- Customer segmentation
- Transaction categorization
- Patient triage
- Defect detection

Topic recognition

- Market trends
- Customer sentiment
- Public health
- Emerging technologies

Key-value extraction

- Claims processing
- **KYC** data collection
- EHR management
- Order processing

GENERATE NEW DATA BASED ON KNOWLEDGE

Conversation

- Customer support
- Financial advisor
- Telemedicine Operations assistant

Text generation

- Personalized marketing
- Patient education
- Financial reports
- Technical documentation

Code generation

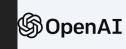
- Coding assistance,
- Language conversion
- API integration
- Test case generation
- Code documentation

Language translation

- Multilingual support
- Medical research translation
- Global compliance

BUSINESS BENEFITS & COMPETITIVE ADVANTAGE

- **Automate tasks** and scale processes
- Reduce costs and increase efficiency
- **Enhance user** experience and personalization
- Accelerate innovation and Al adoption







softserve

APPENDIX



HOW/WHY DOESIT WORK?



WHAT ARE GENERATIVE LARGE LANGUAGE MODELS?

Generative LLM is a language-based <u>simulator</u> that can evolve from an initial state defined by the <u>prompt</u>, drawing from the vast array of <u>knowledge</u>.

By predicting the next word in a sequence, the model maintains an exponential number of simulations selecting the ones that maximize the probability of the entire response, carrying on the progression of meanings, concepts, thoughts, and facts in the most coherent and contextually appropriate way.

HARNESSING THE POWER OF LLMs WITH PROMPT ENGINEERING

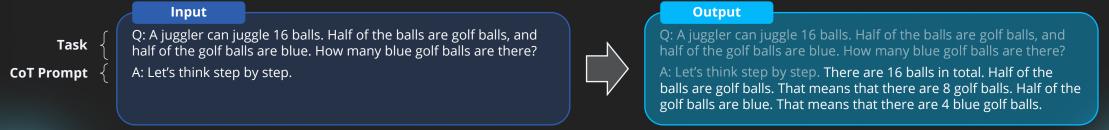
Few-Shot Prompting



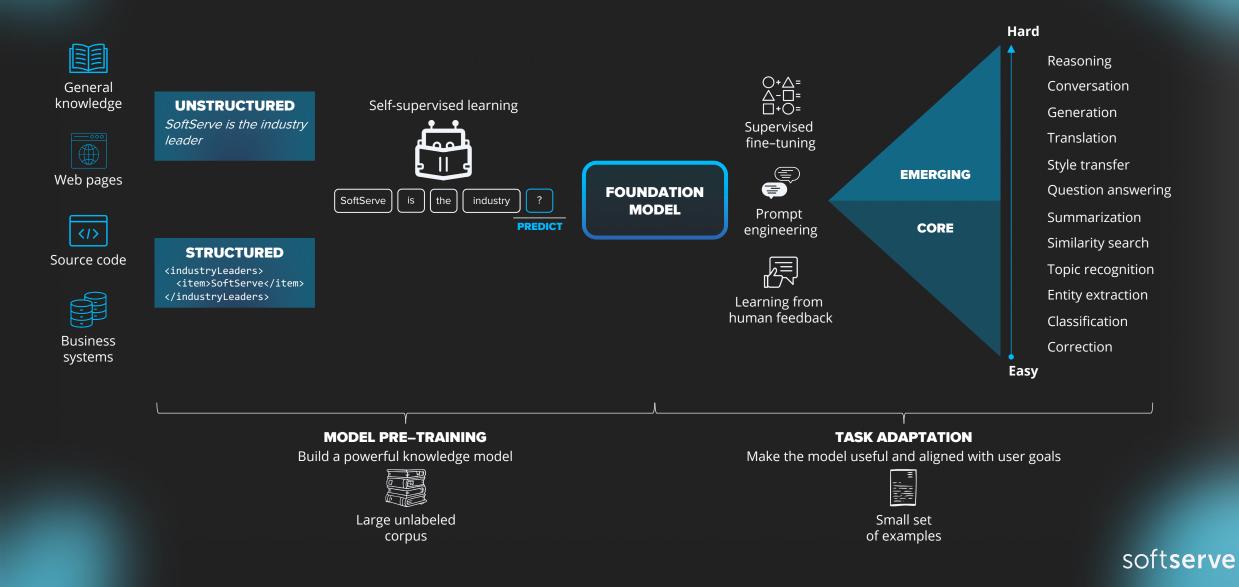
Zero-Shot Prompting



Chain of Thought Prompting



FROM THE FOUNDATIONS TO TASK-SPECIFIC FUNCTIONS



SOFTSERVE AND MICROSOFT



SOFTSERVE AND MICROSOFT OVERVIEW

MICROSOFT SOLUTIONS PARTNER

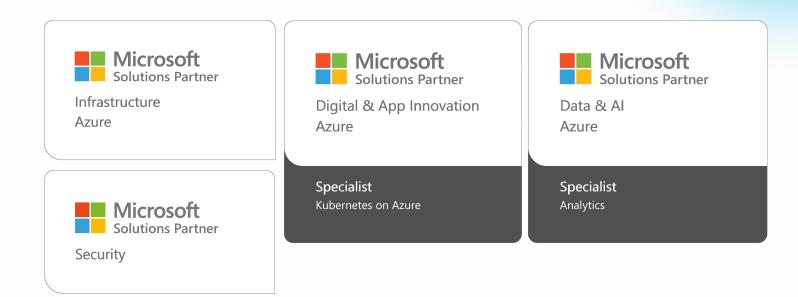
Partner since 2004

MICROSOFT PRACTICE

- 2 Microsoft MVPs
- 500+ Satisfied Customers
- 1,000+ Delivered Projects
- 500+ Microsoft Certified Professionals
- 250+ Azure Certified Professionals

PROGRAM PARTICIPATION

- ECIF Eligible
- AMMP Eligible
- AAAP Eligible
- Solution Assessment Partner
- 20+ Marketplace Offerings



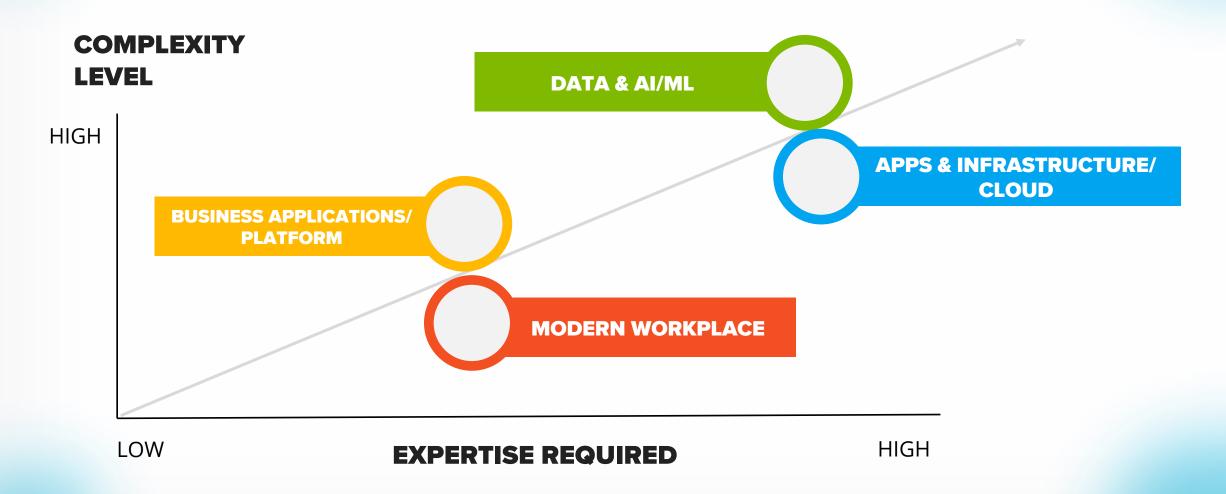
4X SOLUTION AREAS

- Data & Al
- Digital & App Innovation
- Infrastructure
- Security

2 X SPECIALIZATIONS

- Kubernetes on Azure
- Analytics on Azure

SOFTSERVE APPROACH ACROSS MICROSOFT DOMAINS

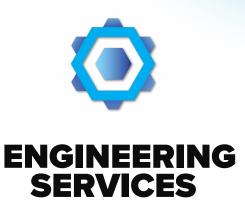


EXPERTISE IN PROVIDING END-TO-END SERVICES IN



















XR