



The transformation architecture

Scaling Predictable AI across the enterprise

*A comprehensive guide to closing the gap between
raw AI capability and enterprise results.*



Closing the AI ROI gap

56% of CEOs say they've gotten nothing back from their AI investments.

Raw AI is powerful but unpredictable.

It can hallucinate, violate regulations, and spiral token costs. Enterprise implementation requires governance – process, orchestration, compliance monitoring – that most teams build manually, slowly, expensively, and incompletely.

Where do most AI initiatives fail:

1

Operational risk

Security concerns are the #1 obstacle to scaling agentic AI. ⁽¹⁾

AI-generated code introduces 1.7X security vulnerabilities. Hallucinations put regulations at risk. Compliance officers can't sign off on decisions they can't understand.

2

Runaway economics

40%+ of agentic AI projects predicted to be cancelled by 2027 due to escalating costs. ⁽²⁾

Token consumption keeps climbing. Cost per transaction is entirely unpredictable. The more you scale, the higher costs spiral.

3

Technical debt

Google Research calls AI the "High Interest Credit Card of Technical Debt." ⁽³⁾

AI agents building in isolation create unmaintainable code. 254% increase in defects compared to human-developed code.

Harnessing AI for predictable transformation

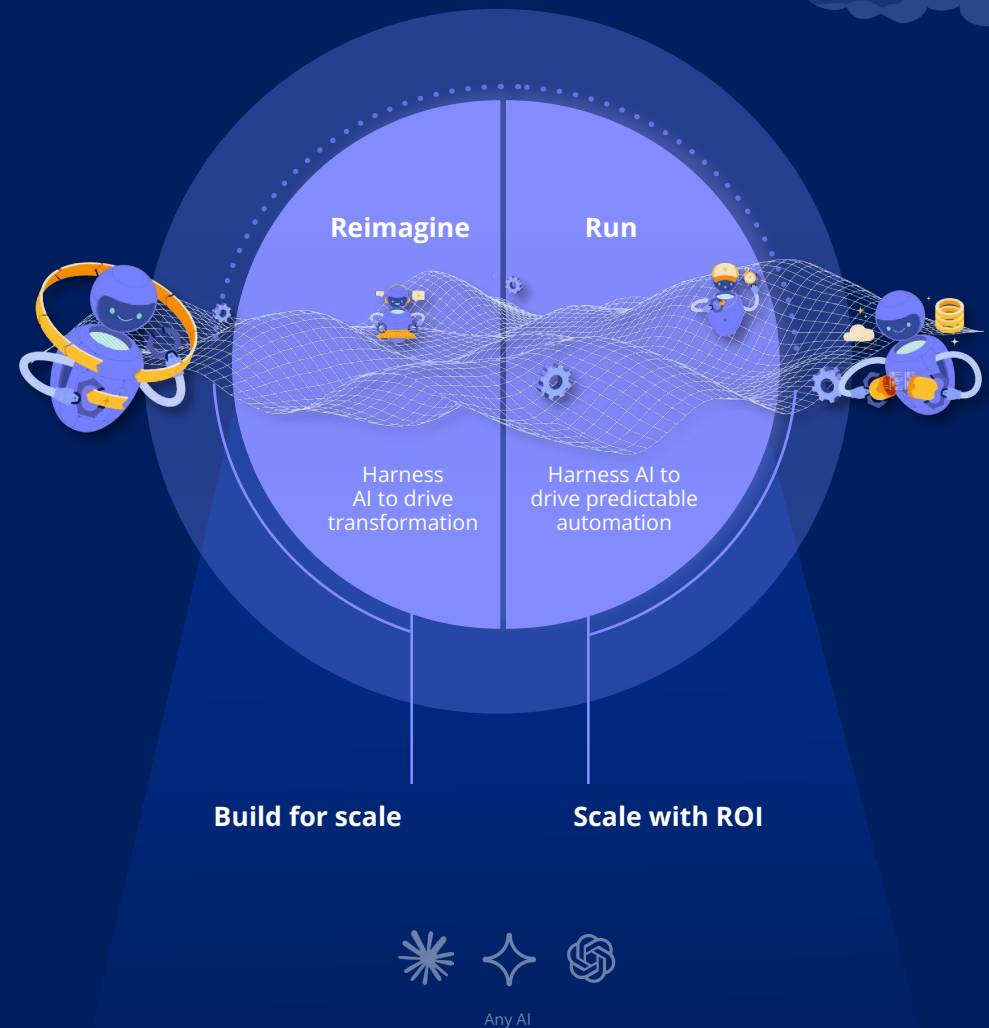
A harness is the operational layer where AI agents run – managing context, state, and orchestrated tool calls.

Rather than agents operating independently, a harness channels them by controlling what information they see, what state they work within, what other agents they can invoke, and what tools they can call. Most agent harnesses operate at the task level – managing the technical implementation of an agent.

But true transformation calls for harnessing AI to work at a higher level – applying AI to accelerate the end-to-end transformation lifecycle – by:

- **Understanding** business requirements.
- **Designing** the right architecture.
- **Building** future-proof applications.
- **Orchestrating** end-to-end processes.
- **Automating** work, reliably.

Enterprise AI harness
Transform & orchestrate
with AI at scale



More code doesn't equal more value

Every CIO has an AI
development mandate, but
few are actually seeing
real productivity gains.

Vulnerabilities. Defects. Runaway costs.

Raw AI can generate code in hours, but
enterprise code must be maintainable,
consistent, auditable, and secure. Speed
without structure creates debt.

BLACK BOX

Impossible to explain
simply to auditors and
business partners.

VULNERABLE

AI-generated code introduces
1.7X security holes.

QUALITY

254% increase in defects
from AI-developed code
impacting productivity.

MASSIVE COSTS

High token consumption
with massive waste as AI
churns and duplicates code.

```
const buildWhereClause = (filters: Record<string, unknown>): string =>
  Object.entries(filters)
    .map(([key, value]) => `${key} = ${stringify(value)}`)
    .join(" AND ");

class QueryBuilder {
  constructor(private schema: TableSchema) {}

  select(
    columns: string[] = ["*"],
    filters: Record<string, unknown> = {},
    limit?: number
  ): string {
    const base = `SELECT ${columns.join(", ")} FROM ${this.schema.tableName}`;
    const where =
      Object.keys(filters).length > 0
        ? ` WHERE ${buildWhereClause(filters)}`
        : "";
    const limitClause = limit ? ` LIMIT ${limit}` : "";
    return `${base}${where}${limitClause}`;
  }

  insert(row: Row): string {
    const keys = Object.keys(row);
    const values = keys.map(k => stringify(row[k]));
    return `INSERT INTO ${this.schema.tableName} (${keys.join(
      ", "
    )}) VALUES (${values.join(", ")}`;
  }

  update(row: Row, filters: Record<string, unknown>): string {
    const assignments = Object.entries(row)
      .map(([key, value]) => `${key} = ${stringify(value)}`)
      .join(", ");

    const where = buildWhereClause(filters);

    return `UPDATE ${this.schema.tableName} SET ${assignments} WHERE ${where}`;
  }

  const buildWhereClause = (filters: Record<string, unknown>): string =>
    Object.entries(filters)
      .map(([key, value]) => `${key} = ${stringify(value)}`)
      .join(" AND ");

  class QueryBuilder {
    constructor(private schema: TableSchema) {}

    select(
      columns: string[] = ["*"],
      filters: Record<string, unknown> = {},
      limit?: number
    ): string {
      const base = `SELECT ${columns.join(", ")} FROM ${this.schema.tableName}`;
      const where =
        Object.keys(filters).length > 0
          ? ` WHERE ${buildWhereClause(filters)}`
          : "";
      const limitClause = limit ? ` LIMIT ${limit}` : "";
      return `${base}${where}${limitClause}`;
    }

    insert(row: Row): string {
      const keys = Object.keys(row);
      const values = keys.map(k => stringify(row[k]));
      return `INSERT INTO ${this.schema.tableName} (${keys.join(
        ", "
      )}) VALUES (${values.join(", ")}`;
    }

    update(row: Row, filters: Record<string, unknown>): string {
      const assignments = Object.entries(row)
        .map(([key, value]) => `${key} = ${stringify(value)}`)
        .join(", ");

      const where = buildWhereClause(filters);

      return `UPDATE ${this.schema.tableName} SET ${assignments} WHERE ${where}`;
    }
  }
}
```

Transformation you can scale

Reimagine fast & build
systems you can rely on.

Pega  **Blueprint™**

Analyze: Synthesize enterprise patterns from your legacy systems and processes.

Compare: Match against industry best practices.

Design: Requirements become high-fidelity Blueprints before app generation.

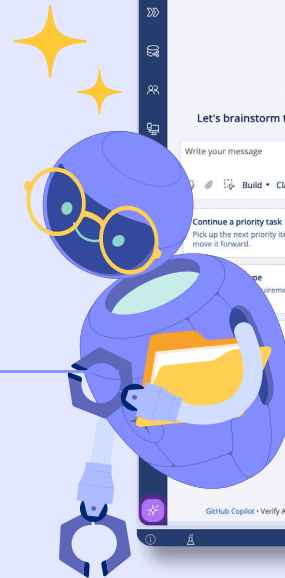
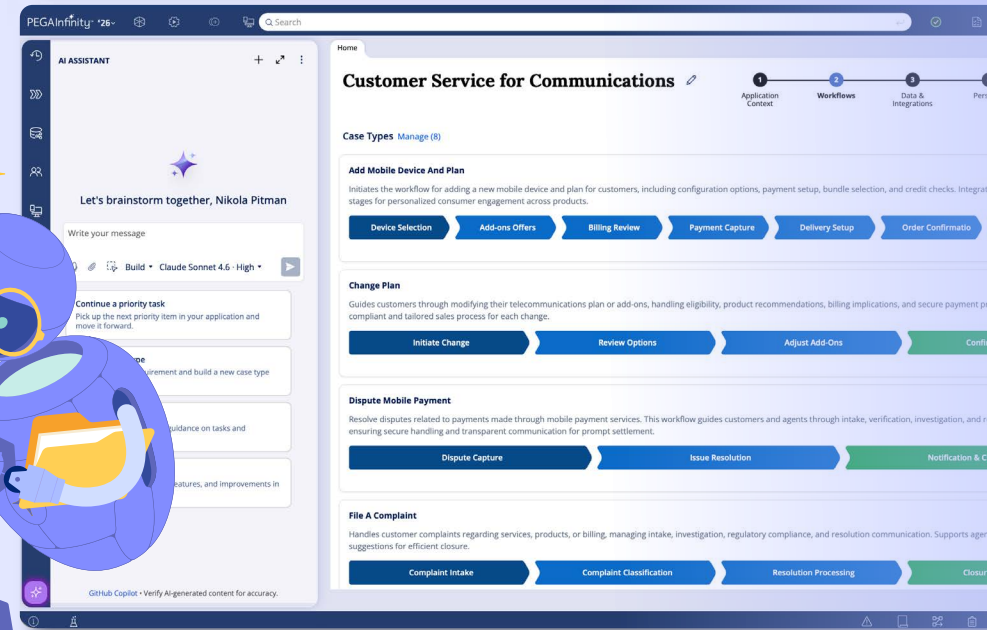
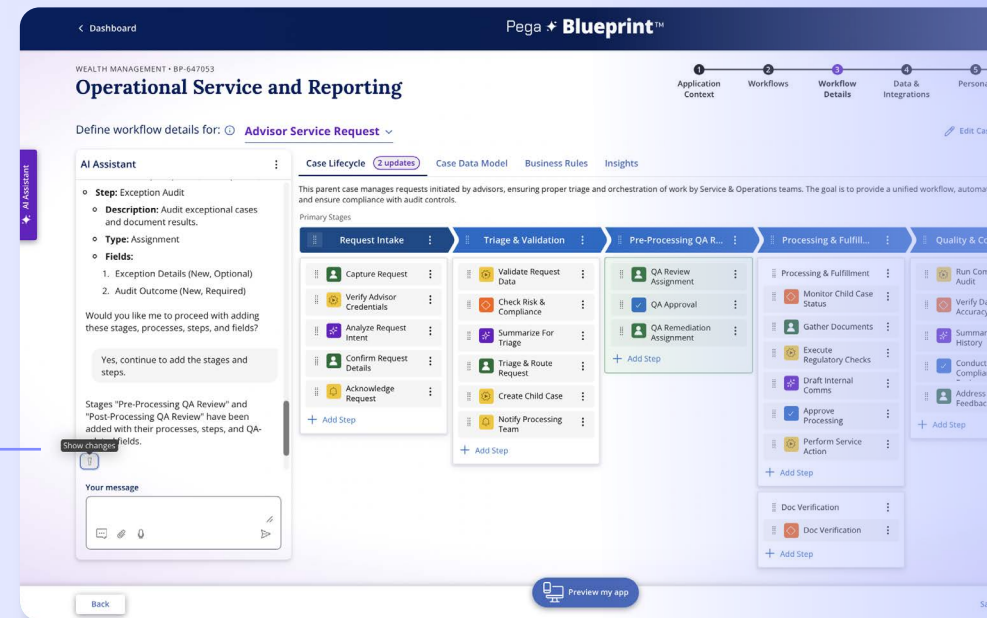
Infinity Studio™

Build: AI agents understand what needs to be done and build autonomously.

Understand: Instantly understand how the system works and what AI is building.

Test & run: AI agents across the SDLC monitor quality, performance, and deployments.

AI harnessed
across the transformation lifecycle
Analysis – Design – Research – Configuration – DevOps



The Result

AI-powered development that accelerates value realization doesn't compound technical debt.

Instantly understandable: Built through transparent design. Non-technical stakeholders see what was built and why it matters.

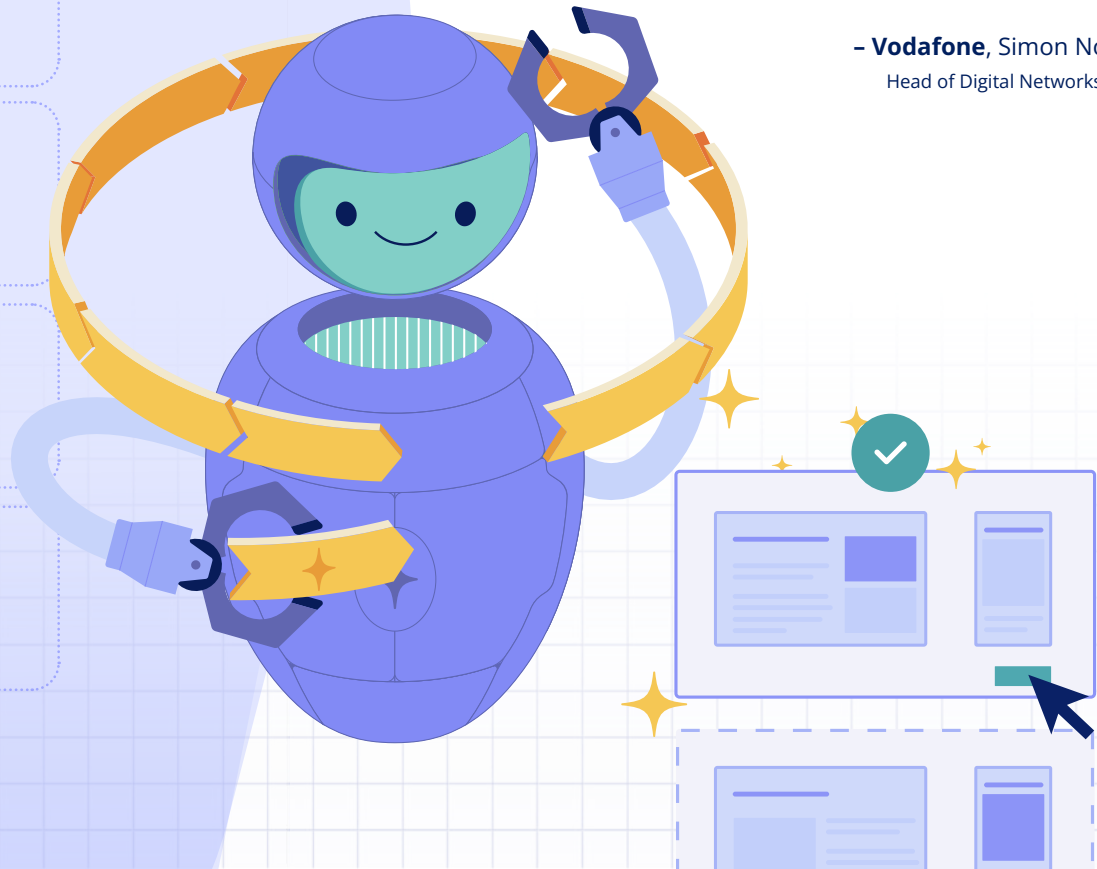
Maintainable: Model-driven architecture means updating the Blueprint so behavior changes instantly. No rewriting from scratch.

Reliable: Deterministic design means consistent behavior. Fewer defects, fewer security vulnerabilities. Compliance from design, not patched in later.

Aligned with business: Built on requirements and best practices, not developer hunches. Delivers the right value.

“ Instead of taking two to three months... we can put a working product in users' hands within 48 hours.

– **Vodafone**, Simon Norton,
Head of Digital Networks & OSS



From COBOL to Cloud

How Unum is transforming claims – going from mainframe to AI in the cloud unlocking \$20M savings.

\$20M

Total savings – across infrastructure reduction and process efficiencies

7K

Hours saved across claims operations – with faster response to members

1.5M

Lines of COBOL analyzed and transformed

Legacy processes, legacy codebase:

Mainframe-based claims operations with decades-old code. High cost to change. Strategic need: Modernize to cloud-native while improving claims speed and reducing operational cost.

Traditional approach:

Line-by-line, COBOL-to-Java translation. Years of effort. Months of testing per module. Massive cost. High risk of logic errors.

The Pega approach:

Blueprint AI analyzed the legacy code, understood business logic, synthesized best practices, and generated deterministic cloud-native applications. Rather than translating code, Unum reimaged the processes the code supported.

unum

Meet Cloud
Group's Ultimate



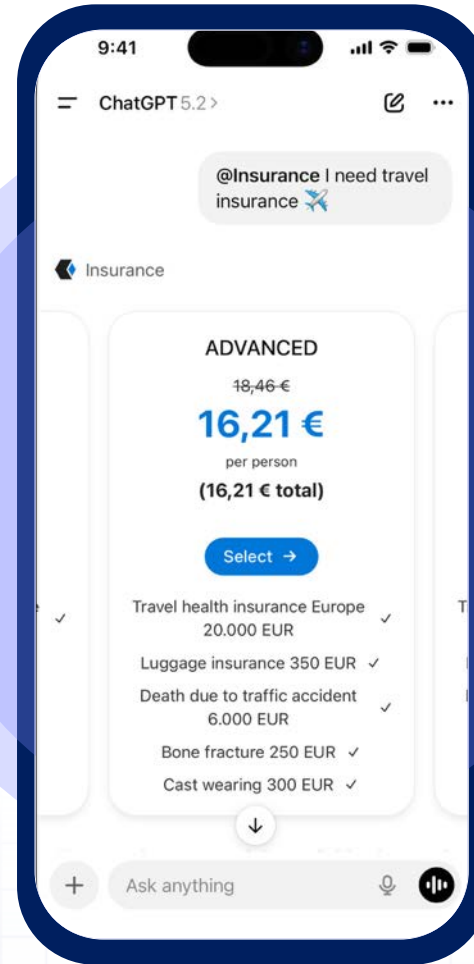
Agents without structure = risks waiting to happen

Security & risk concerns are the No.1 obstacle to scaling agentic AI.

A single hallucination in a regulated environment can cost millions.

When agents autonomously orchestrate mission-critical, customer facing, and regulated work – even little mistakes can mean big problems.

Agents need to be compliant, reliable, orchestrated alongside your people and systems, and completely measurable.



COMPLIANCE GAPS

AI hallucinations can miss critical validations, rules, and process requirements

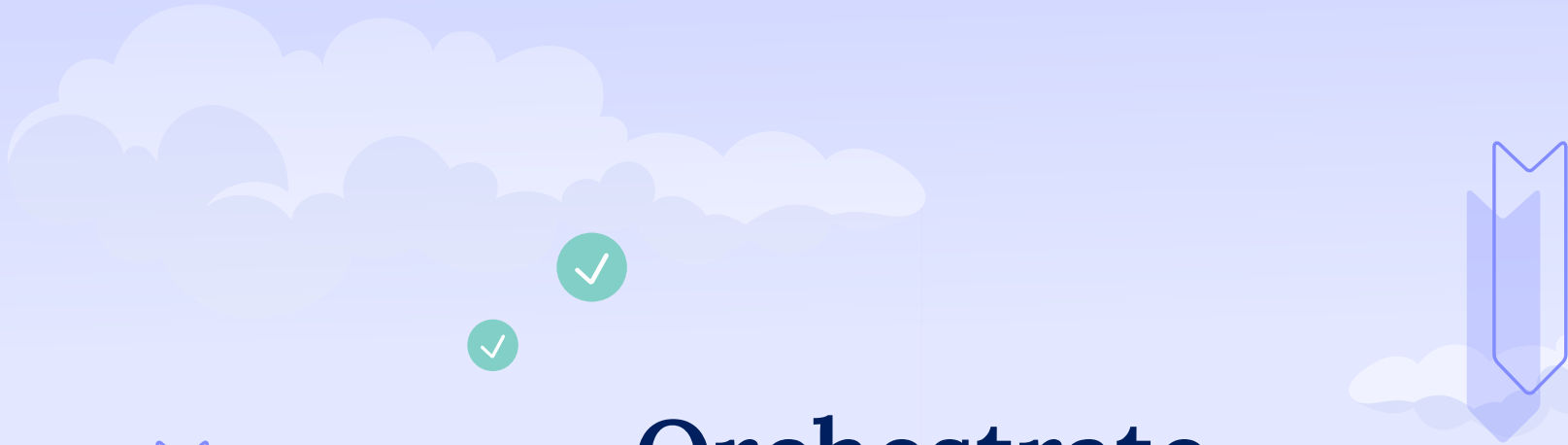
BRAND RISKS

Agents gone rogue risk mistakes, and mishandling of PII and sensitive data

NEGATIVE ROI

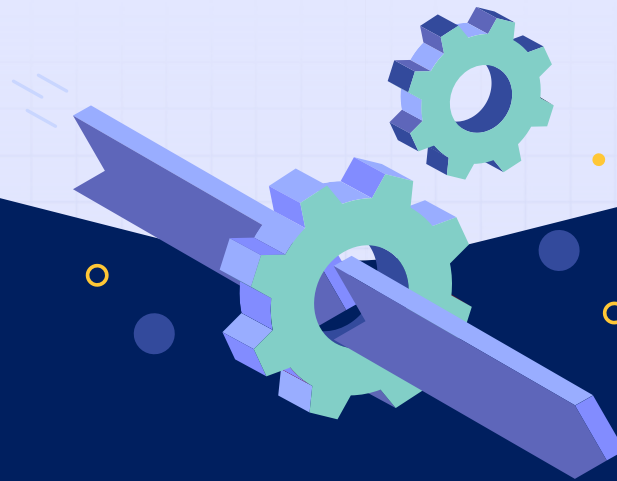
Long running agentic workflows burn through tokens, re-reasoning at every step of the process





Orchestrate processes predictably

Harness AI, people, & systems together
to automate work with confidence.



Harness AI, people, & systems together to automate work with confidence

AI experience layer

Automate work intake.
Deliver self-service for
customers and employees.

Omni-channel | AI-assisted | Triage and routing



MCP Web & Mobile Contact Center Chat Voice Back-office Collaboration

Orchestration layer

Drive consistent work execution
across the operation with a unified
process harness.

100% governance | Reporting and traceability | Prioritization and SLAs



AI automation layer

Automate manual work with a team of AI
agents operating within structure.

Open – any agent, any platform
Powered with full context of customer journey
Rules, standards, and validation enforced



The harness drives & governs

Ensuring work gets done right

Rules & standards

Every decision path respects business rules, compliance requirements, and operational standards. 100% compliance without fail.

Context & data

AI agents are fueled with slim context and data – only what they need to get the job done effectively from across systems, documents, knowledge bases, and history.

Focused agents

Each agent has a clear, bounded job. Hyper-focused agents need fewer tokens and hallucinate less.

Human oversight

Agents assist, humans decide. Escalates automatically when complexity triggers.

Auditability

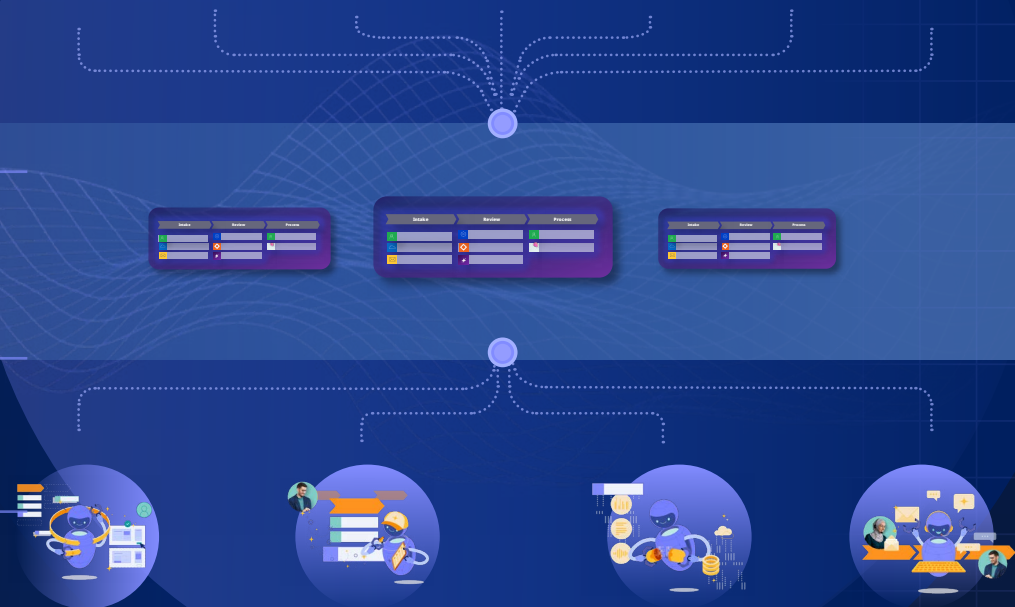
Why was a decision made? What data was available? Which rule triggered? Full visibility at every step.

Predictable cost

You pay for outcomes (cases processed), not for AI thinking (tokens). Costs don't spiral as you scale.



MCP Web & Mobile Contact Center Chat Voice Back-office Collaboration



The 6Rs of agentic automation

Work follows a predictable pattern.

It's received, routed, researched,
reviewed, responded to, resolved.

These six Rs haven't changed
in decades because they reflect
how work actually moves through
operations.

**AI plugs seamlessly into three
high-impact phases** – automating the
routine, repeatable, auditable parts
while humans focus on judgment.



RECEIVE & ROUTE **24/7 automated self-service**

The challenge:
Emails, calls, tickets –
without prioritization
or SLA management.

The opportunity: Understand and
categorize incoming emails, calls,
chats, and documents and route to
the right automated process.



RESEARCH & REVIEW **Agents synthesize, humans approve**

The challenge:
Manual review of data,
document, and SOP
with limited auditability.

The opportunity: Enrich and
analyze data and documents
for approval by automated
logic or human reviewers.



RESPOND & RESOLVE **Eliminate wrap-up & unlock insights**

The challenge:
Minutes spent per
case on wrap up notes
and drafting
correspondence.

The opportunity: Summarize
findings and draft correspon-
dence for final approval and
notification.



24/7 automated self-service

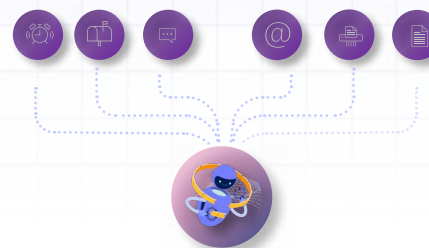
Understand & categorize incoming emails, calls, chats, and documents and route to the right automated process.

What's happening now: Work arrives through email, chat, calls, documents, forms, APIs. Someone reads, understands, classifies, extracts data, and routes. This is manual, slow, and invisible. Your team is losing time to something routine and repeatable.

How Pega harnesses AI:

- Document agents' ingest files and extract data automatically.
- Chat and email agents capture requests conversationally, categorize, and pull details.
- MCP tools connect customer-facing agents to consistent service library.
- Orchestration automatically assigns work based on skills, availability, urgency, and rules.

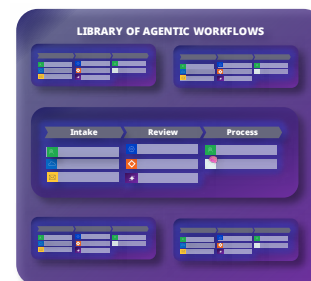
Impact: Faster cycle time; no missed SLAs; complete visibility; team sees pre-classified, enriched, ready-to-work assignments.



Automate intake from any channel

Embed agents in any channel to understand and categorize work

DX API (web, mobile) | Document Agent
MCP | A2A | Agentic Email & Chat



Automate intake from any channel

Embed agents in any channel to understand and categorize work

DX API (web, mobile) | Document Agent
MCP | A2A | Agentic Email & Chat



Workflow	Assignment	Due In	Skills	Complexity	Priority	Assigned to
Workflow 1	ASSIGNED TO AGENT 1	10:00 AM	Basic	Low	High	Agent 1
Workflow 2	ASSIGNED TO AGENT 2	11:00 AM	Basic	Low	Medium	Agent 2
Workflow 3	ASSIGNED TO AGENT 3	12:00 PM	Basic	Low	Low	Agent 3
Workflow 4	ASSIGNED TO AGENT 4	1:00 PM	Basic	Low	Medium	Agent 4
Workflow 5	ASSIGNED TO AGENT 5	2:00 PM	Basic	Low	High	Agent 5
Workflow 6	ASSIGNED TO AGENT 6	3:00 PM	Basic	Low	Medium	Agent 6
Workflow 7	ASSIGNED TO AGENT 7	4:00 PM	Basic	Low	Low	Agent 7
Workflow 8	ASSIGNED TO AGENT 8	5:00 PM	Basic	Low	Medium	Agent 8
Workflow 9	ASSIGNED TO AGENT 9	6:00 PM	Basic	Low	High	Agent 9
Workflow 10	ASSIGNED TO AGENT 10	7:00 PM	Basic	Low	Medium	Agent 10

Automate intake from any channel

Embed agents in any channel to understand and categorize work

DX API (web, mobile) | Document Agent
MCP | A2A | Agentic Email & Chat

Agents synthesize, humans approve

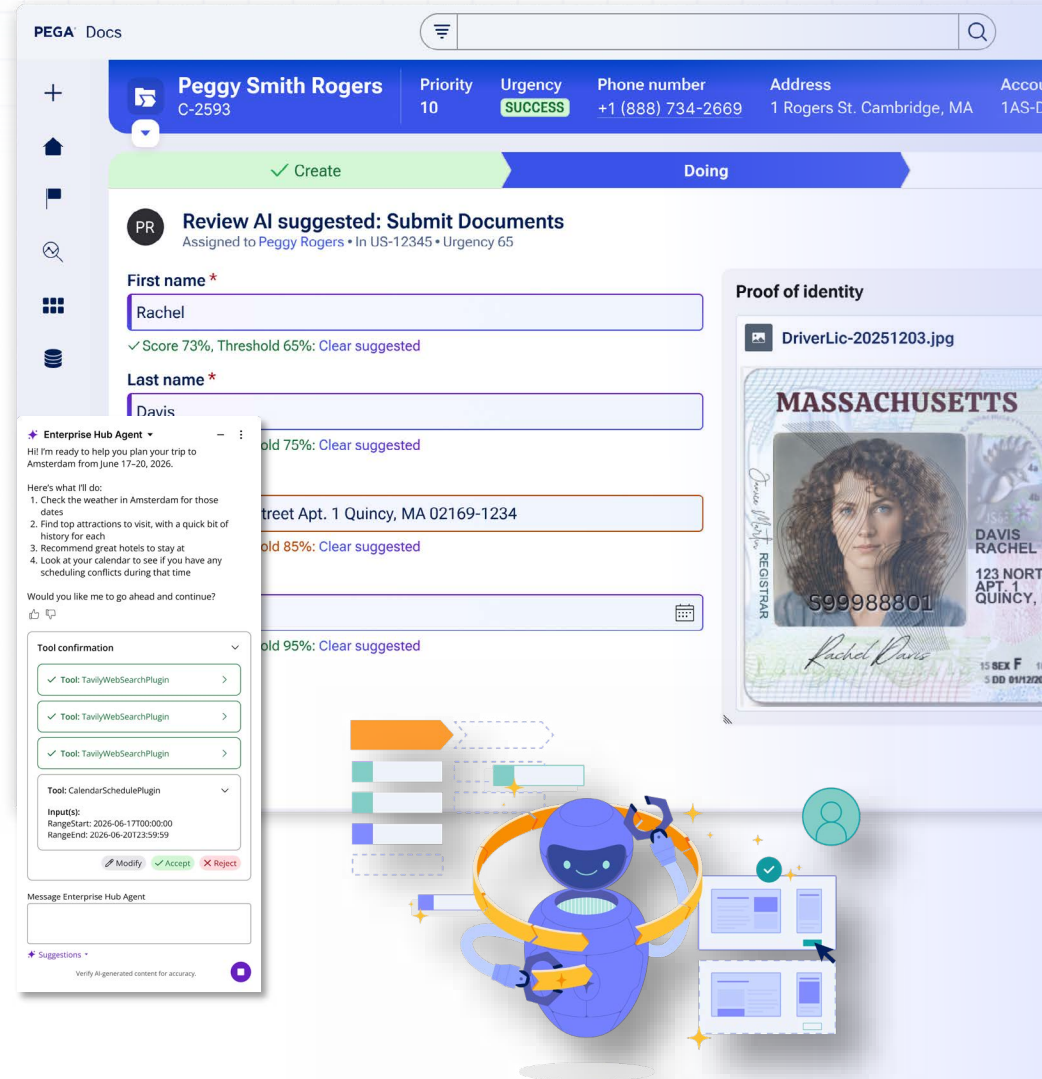
Enrich & analyze data and documents for approval by automated logic or human reviewers.

What's happening now: Team members spend minutes per case logging into multiple systems, reading hundreds of pages, synthesizing to figure out decisions. Time spent on assembly, not judgment.

How Pega harnesses AI:

- Document agents analyzing documents at scale; team sees summaries and insights.
- AI agents pull from multiple data sources simultaneously.
- SOPs and best practices synthesized into context-specific guidance.
- AI assistants surface relevant information in real time.
- Recommended next best action is generated for review and approval.

Impact: Time per case drops 30–60%, consistency across cases, new employees productive in weeks, better decisions.



Eliminate wrap-up & unlock insights

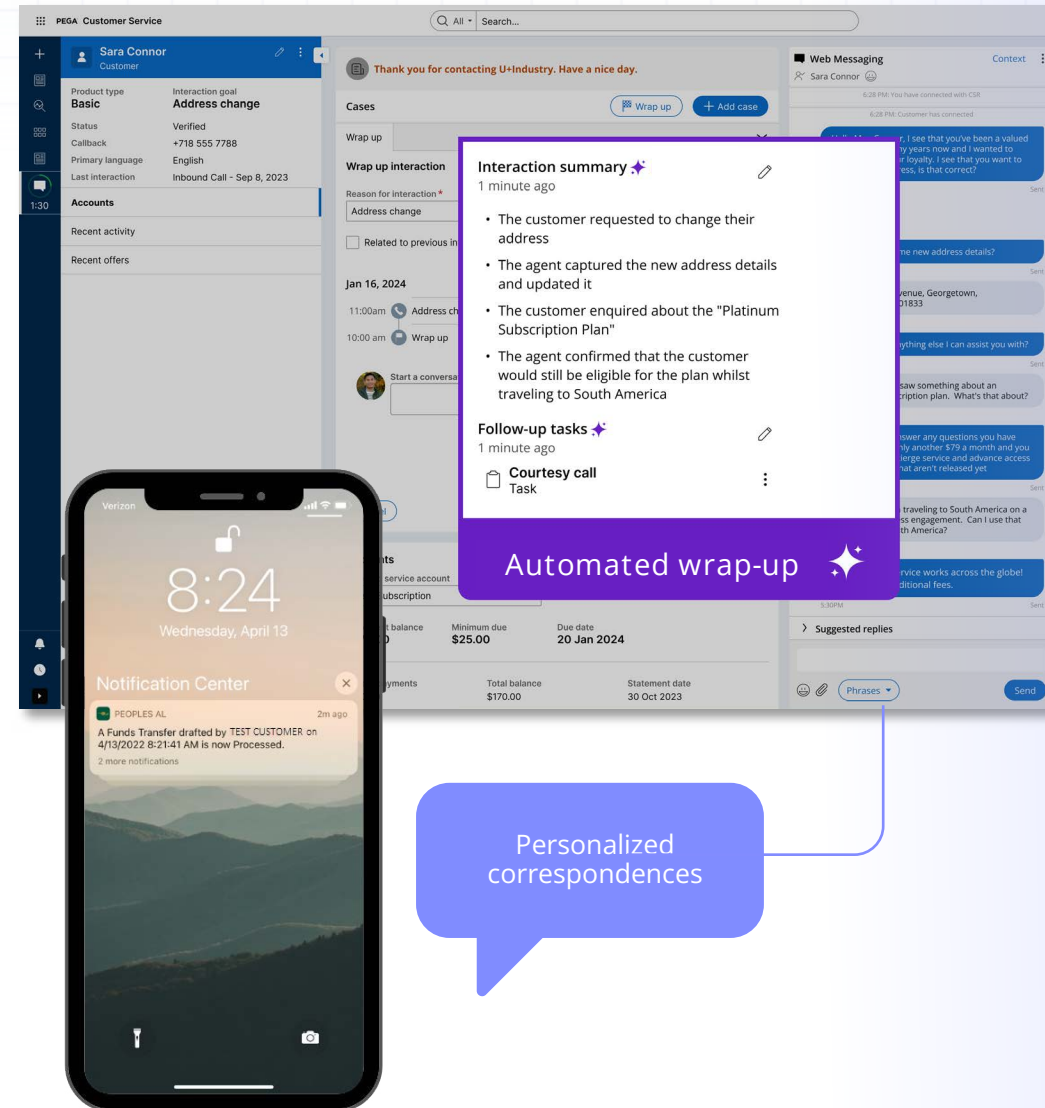
Summarize findings and draft correspondence for final approval and notification

What's happening now: Once a decision is made, wrap-up begins – writing emails, drafting documents, writing case notes. Hours per day going to correspondence instead of the next case. Wrap-up notes are inconsistent, can't be analyzed.

How Pega harnesses AI:

- Correspondence agents pull from template library, fill in case-specific details.
- Document generation fills standardized templates, personalizes within standards.
- Case summaries synthesize entire customer journey into comprehensive notes.
- All output reviewed and approved by humans before dispatch.

Impact: Closure time drops significantly, wrap-up notes become operational data, unlock operational intelligence from scale.



Scaling agentic asset finance



How Lloyds is transforming asset finance
– going from manual work to agentic
workflows across 2M+ cases per year

83%

Automatically
indexed – no human
categorization required

99%

Accuracy in data extraction
– from complex financial
documents

2M+

Scale per year – with
consistent quality and
SLA performance

The challenge:

Asset finance cases require complex documentation, research across multiple data sources, consistent decision-making. High volume. High complexity. Manual bottlenecks in every phase.

2M+ agentic cases:

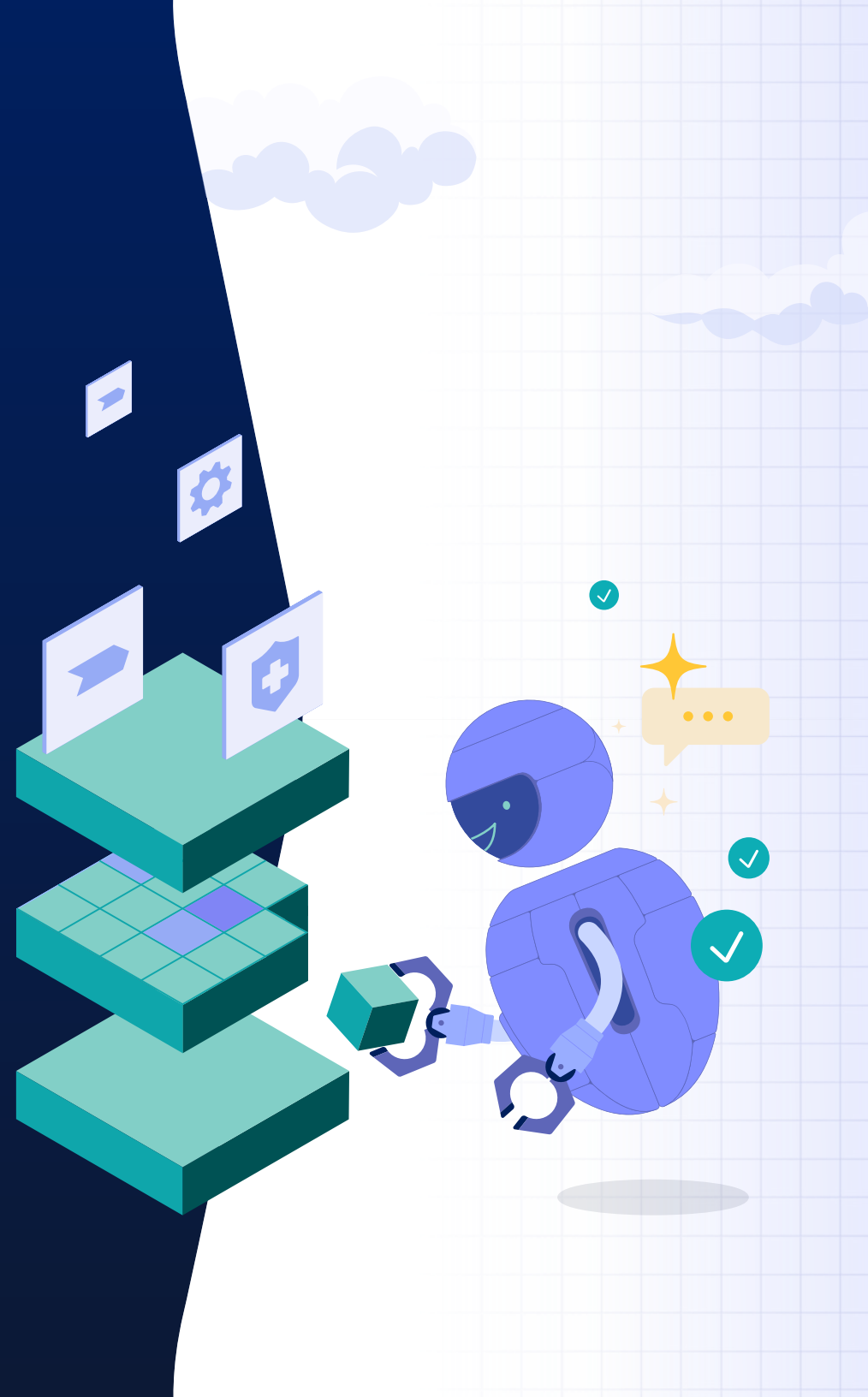
Deployed agentic orchestration across full case lifecycle – intake, research, resolve – with complete governance and audit trails. Scaled from pilot to 2M+ cases in production in months, not years.



PREDICTABILITY MATTERS

Reason every step, or run the right workflow?

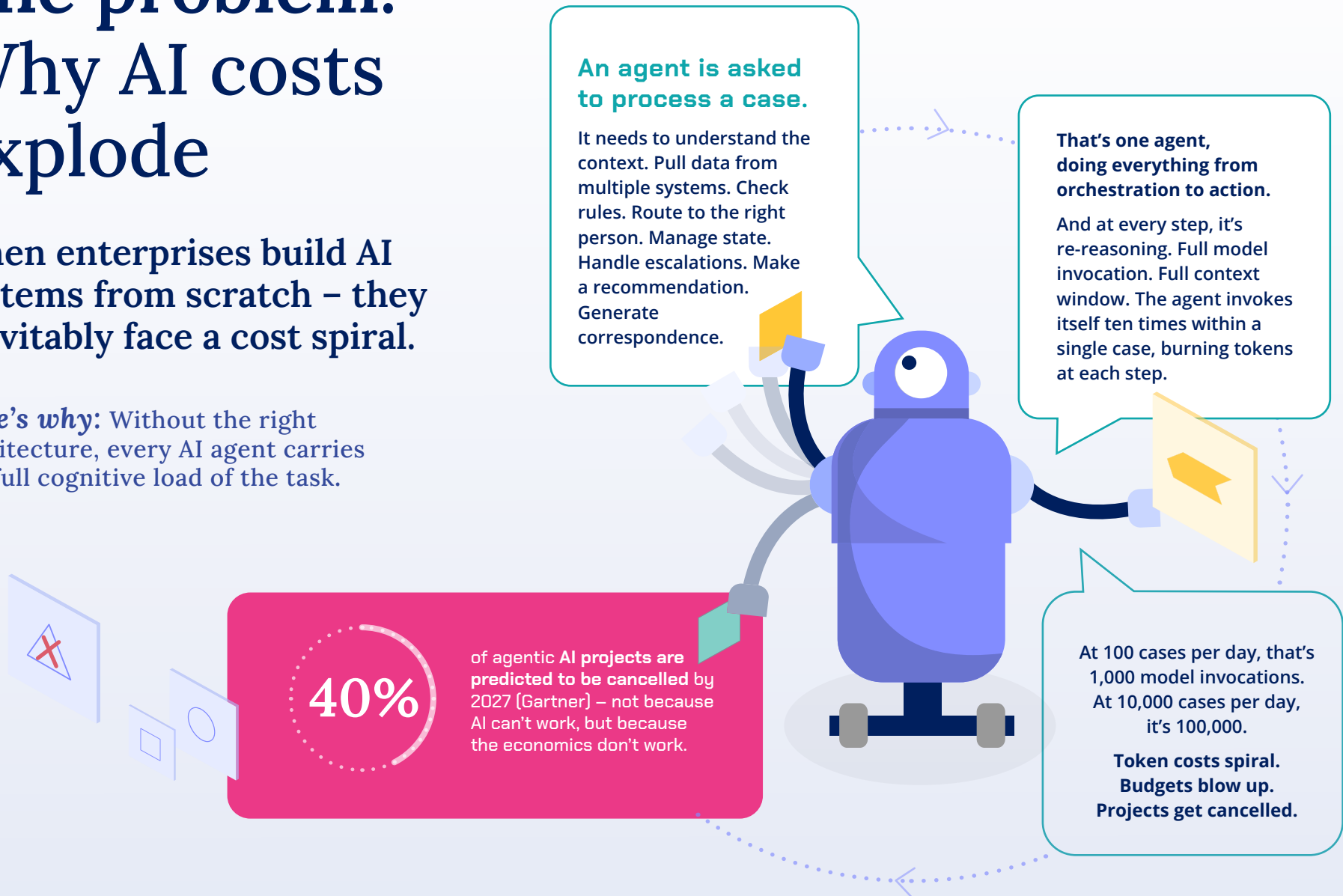
Most enterprise work is deterministic. Re-figuring the same steps on every request drives up token cost and risk. Pega controls the workflow and the context – passing each agent only what it needs for the next step.



The problem: Why AI costs explode

When enterprises build AI systems from scratch – they inevitably face a cost spiral.

Here's why: Without the right architecture, every AI agent carries the full cognitive load of the task.



The Pega difference: Harness efficiency

Pega's harness eliminates the re-reasoning problem through clean separation of concerns.

The result: Agents stay hyper-focused. They don't re-reason at every step. They don't manage state. They don't handle routing logic. They do one thing – and do it well.

Token efficiency compounds at scale. Per-case AI cost drops, not rises.



Orchestration handles:

This is efficient logic, not reasoning



Routing

escalation



state
management

rule
enforcement



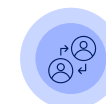
prioritization

compliance



AI agents handle:

This is reasoning – but scoped, focused, bounded



Engagement

content
generation



synthesis

guidance



Case-based pricing: Paying for value, not tokens

This architecture – efficient, scoped AI agents within orchestration – is why Pega offers case-based pricing.

Rather than charging by tokens (how much AI is thinking), we charge by cases (how much work the platform is completing).

An enterprise processing 2M+ cases per year (like Lloyds) isn't paying 2M per-token-cost. They're paying a fixed rate per case, regardless of how many times AI is invoked internally.



This is why Pega can offer unlimited AI use

Not as a subsidy. Not as a loss leader.
As intelligent architecture.



The orchestration engine is doing the heavy lifting. AI agents focus on what matters. When you separate these concerns properly, unlimited AI doesn't mean unlimited cost – it means predictable cost, tied to outcomes, not tokens.

Analyst validation: Leading AI platform + leading orchestration

Pega is positioned as a Leader in both:



Forrester's AI Platforms Wave

2026

for AI-powered development capabilities and the ability to harness AI to build applications you can trust

Gartner's Business Orchestration & Automation Technologies (BOAT) 2025

for orchestration and governance at enterprise scale

This dual positioning matters:

AI capability without orchestration is chaos. Orchestration without AI capability is waste.



Pega is the leading Enterprise Transformation Company™ that helps organizations Build for Change® with enterprise AI decisioning and workflow automation. Many of the world's most influential businesses rely on our platform to solve their most pressing challenges, from personalizing engagement to automating service to streamlining operations. Since 1983, we've built our scalable and flexible architecture to help enterprises meet today's customer demands while continuously transforming for tomorrow. For more information on Pega (NASDAQ: PEGA), visit <http://www.pegacom>