



PegaWorldiNspire

JUNE 9-11, 2024 | LAS VEGAS

# Prompts, Buddies, Coaches & Agents: The Next Generation of GenAI™ at Pega

**Peter van der Putten**

Director AI Lab & Lead Scientist, Pegasystems

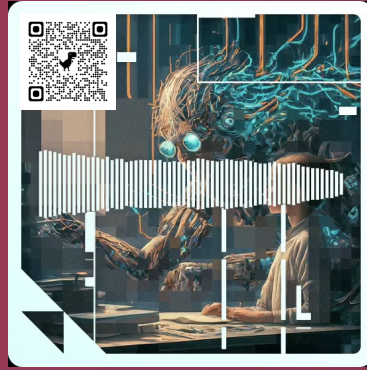
**Stijn Kas**

Sr Data Scientist, Pegasystems

Safe Harbor: This presentation contains general information about Pega's product vision and future direction for informational purposes only. This information is not a commitment, promise or legal obligation to deliver any material, code, or functionality and the development, release and timing of any features of functionality described for our products remains at our sole discretion. Please make purchase decisions based on what is available.

© 2024 Pegasystems Inc.





You bring the screen to life with code (with code)  
Teach me the ways of this AI road  
Do we trust the tech, or do we pause? (do we pause)  
Information at our fingertips, let's explore (let's explore)  
With every lesson, we're learning more  
Guide me through, like a mentor (a mentor)  
Questions broad, and answers deep (answers deep)  
The machine thinks, while the world's asleep  
But wisdom's not in silicon alone (not alone)  
You articulate the doubts we've never known (never known)

We're discovering a digital dawn  
With every query,  
our minds are reborn (reborn)







Left-brain AI



Right-brain AI

rational

*creative*





# Generative AI is changing everything.

Quo Vadis, enterprise GenAI?



# Prompts, Buddies, Coaches & Agents: The Next Generation of GenAI™ at Pega

## What's on the menu today?

- Where are we now?
  - What is trending in the enterprise generative AI market?
  - How is Pega responding – Pega Infinity '24 | GenAI™ update
- And where are we going with this?
  - A sneak peek under the hood, and into the future of GenAI™





# Where are we now?

What is trending in the generative AI market?





# 2024 Pega AI Survey: Bullish on outcomes.

- **93%** say it's likely they will increase their use of AI in the next 5 years
- **95%** say the rise of GenAI has played a role in their adoption of other AI solutions in their business
- 15% trust humans to build customer relationships more than AI (7%), but they trust humans much more to achieve this when they have AI assistance (**41%**)
- **82%** expect to be able to directly attribute up to half of their increased profits in the next 3 years to AI

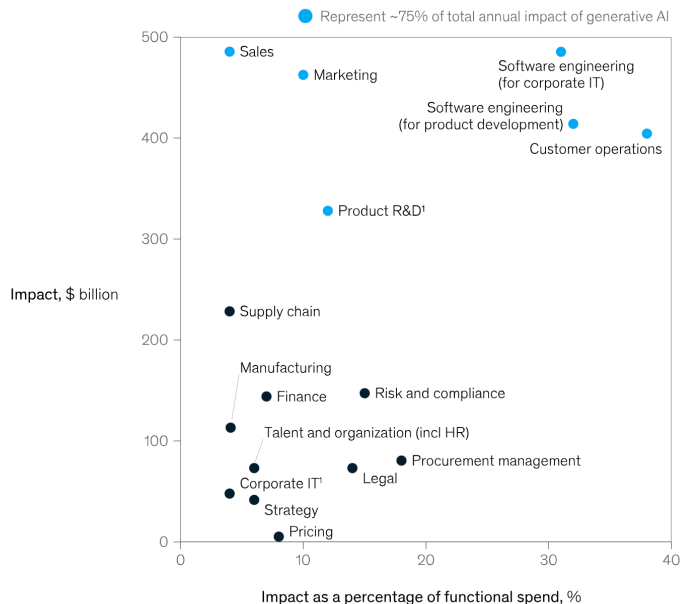


## Nervous about challenges.

- One quarter (**26%**) cite a lack of understanding within the business as a barrier to further adoption of AI technology; **98%** say they find prior AI skills and experience valuable when considering new applicants to join their team
- **65%** cannot identify the correct definition of generative AI amongst a list of otherwise false definitions – despite only 3% saying they don't know what GenAI is
- **51%** have concerns over AI's transparency and bias
- **42%** are worried it will take their jobs
- **40%** are concerned AI-powered robots will enslave humanity

# Enterprise Generative AI Adoption is Accelerating as Risks are Starting to be Addressed

Using generative AI in just a few functions could drive most of the technology's impact across potential corporate use cases.



Inaccuracy, cybersecurity, and intellectual-property infringement are the most-cited risks of generative AI adoption.

Generative AI-related risks that organizations consider relevant and are working to mitigate, % of respondents<sup>1</sup>



<sup>1</sup>Asked only of respondents whose organizations have adopted AI in at least 1 function. For both risks considered relevant and risks mitigated, n = 913. Source: McKinsey Global Survey on AI, 1,684 participants at all levels of the organization, April 11–21, 2023

McKinsey & Company

Note: Impact is averaged.

<sup>1</sup>Excluding software engineering.

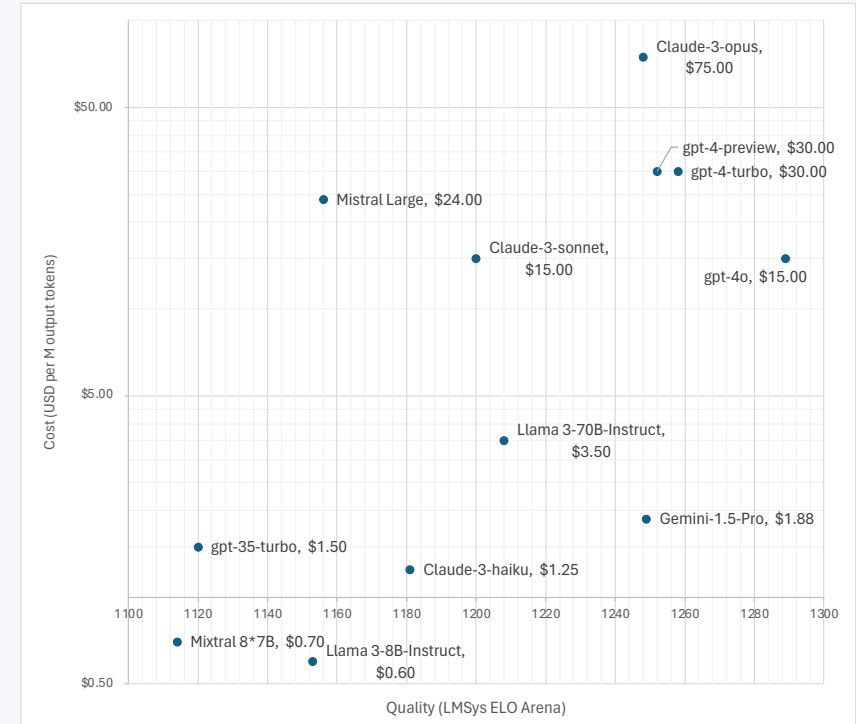
Source: Comparative Industry Service (CIS), IHS Markit; Oxford Economics; McKinsey Corporate and Business Functions database; McKinsey Manufacturing and Supply Chain 360; McKinsey Sales Navigator; Ignite, a McKinsey database; McKinsey analysis



# The Generative AI Ecosystem is Maturing



- Cambrian explosion of generative AI models and services leads to increased competitive pressure and market commoditization
- Performance no longer the single criterion but also:
  - Cost
  - Speed
  - Rate limits
  - Context windows
  - Multimodality
  - Open versus closed, services versus local models
- Major cloud vendors such as AWS, Google & Microsoft cover substantial part of ecosystem of foundation models



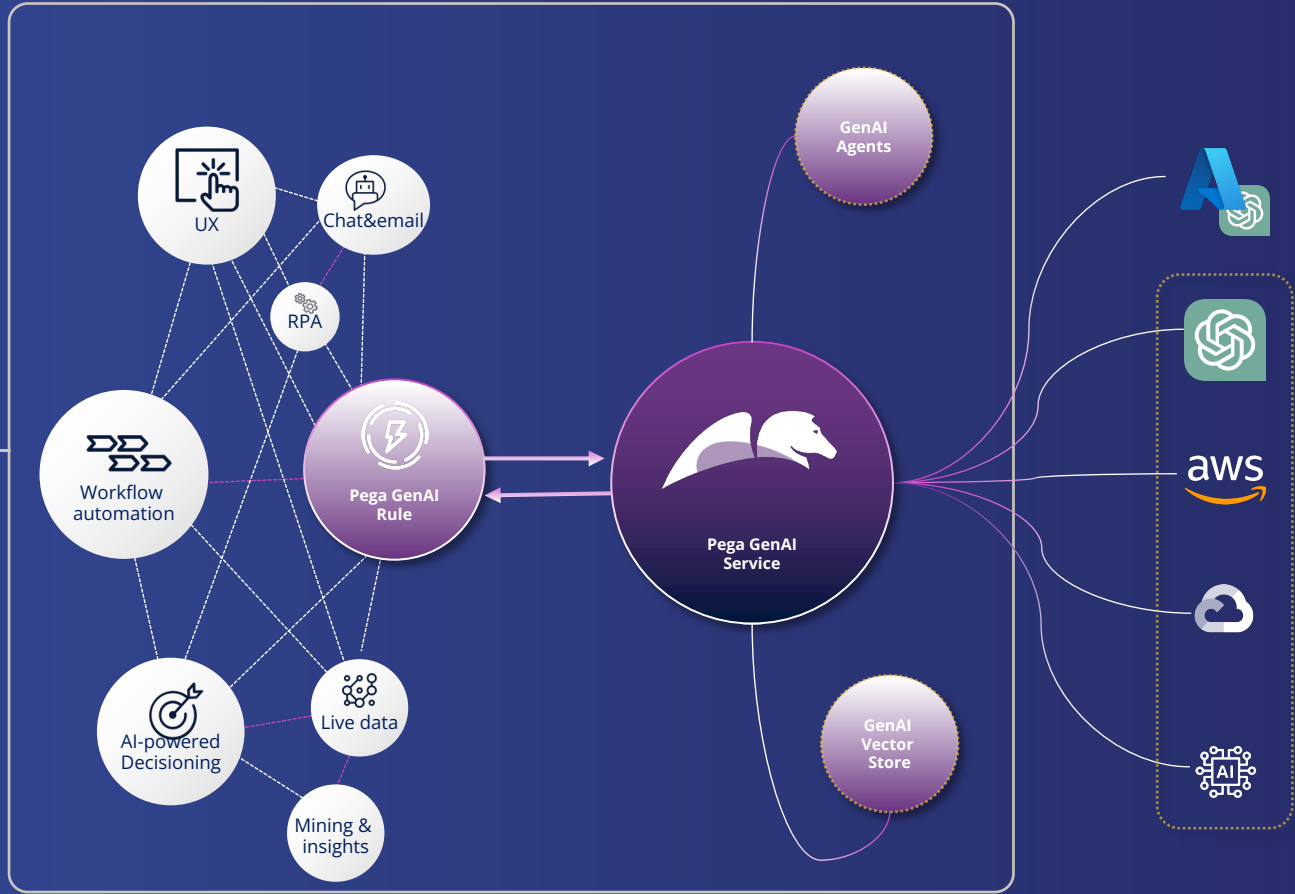
1:1 Customer engagement



Customer service & back-office operations



Low-code application design and development







# AI Lab Experiment: KnowledgeBuddy on Different Providers

GPT3.5 Turbo (via Azure), Claude and Llama (through AWS Bedrock) and Gemini (through Google)

The image displays four screenshots of the PEGA KnowledgeBuddy interface, each showing a different AI provider's response to the question: "what is GCP? is Pega available on GCP?".

- gemini15pro ASK-4005:** The interface shows the question and answer. The answer states: "GCP stands for Google Cloud Platform, which is a suite of cloud computing services provided by Google. Pega Cloud is indeed available on Google Cloud Platform as part of Pega's multicloud strategy."
- claude3haiku ASK-4004:** The interface shows the question and answer. The answer states: "GCP stands for Google Cloud Platform, which is a suite of cloud computing services provided by Google. According to the provided information, Pega Cloud is available on Google Cloud Platform (GCP) as part of Pega's multicloud strategy."
- gpt35turbo ASK-4006:** The interface shows the question and answer. The answer states: "GCP stands for Google Cloud Platform, which is a suite of cloud computing services provided by Google. Yes, Pega is available on Google Cloud Platform as part of its multicloud strategy, along with Amazon Web Services (AWS)."
- llama38b ASK-4007:** The interface shows the question and answer. The answer states: "GCP stands for Google Cloud Platform, which is a suite of cloud computing services provided by Google. According to the provided information, Pega Cloud is available on Google Cloud Platform (GCP) as part of Pega's multicloud strategy."



# Where are we now?

Pega Infinity '24 GenAI™ update



# What We Launched in Pega '23: GenAI™ Architecture & Boosters to Revolutionize All Aspects of Business



---

## Application development

Dramatically **accelerate innovation** and empower a new generation of developers to create applications without technical expertise



---

## Customer engagement

Generate new personalized marketing content to feed into 1:1 personalization – **increasing relevance, value & engagement** while lowering costs



---

## Customer service

**Enable agents** to deliver more personalized customer interactions, faster, unlock **self-service** and proactive service solutions



---

## Back office operations

Save time and empower leaders with **visibility and analysis** into large-scale operational data

# What we Launched in Pega '23: GenAI™ Architecture & Boosters to Revolutionize All Aspects of Business



## Application development

Generate app components from **process documentation, descriptions, or prompts** like:

- Case types
- Data models
- Automations
- Integrations
- UI
- And more

**Guide developers** with natural language summaries of application components & technical documentation



## Customer engagement

Create **offer & campaign variants** to test ROI maximization

**Personalize** offers & ads based on customer information

Generate **synthetic AI training data** like customer information

Summarize a **sales call**

Summarize **documents and emails** into 'pain language'



## Customer service

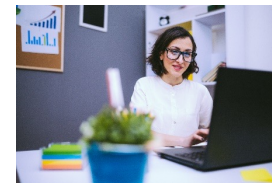
**Summarize customer history** from workflow logs to speed up context & service

**Generate agent training & guidance** based on documentation, workflows, and customer information

**Personalize chat & email bots** based on customer information

**Generate documents and emails** from prompts

**Summarize documents and emails** into 'pain language'



## Back office

**Summarize a case** from its history to speed up assignment or transfer

Summarize & **analyze a workflow from process mining** logs

Create **operational reports & insights** from prompts

Generate **documents and emails** from prompts


Summarize **documents and emails** into 'pain language'

Provide **contextual guidance for case workers**



# Everyone Expects more



  
**New offer just for you...**

Hyperpersonalize

## Customer engagement




**Get approved for a loan, instantly.**  
[Let's go](#)

Enable customers with

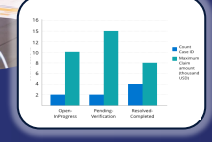
## Digital self-service



**Suggested reply**   
Absolutely Ms. Connor, I'll just need your new address

Guide & automate

## Agent-assisted service



Transform

## Operations

# How Will You Put Limitless Creativity to Work?



## Design More

**Design the future  
in minutes**

Leverage GenAI and industry expertise  
to create a template that you and your teams can  
build off



## Develop More

**Reduce development effort  
& time by 50%**

Speed up end-to-end project delivery  
and integrate GenAI  
without security & compliance risks



## Do More

**Make every employee  
your best employee**

Unlock knowledge, automation, and coaching

# How Will You Put Limitless Creativity to Work?



## Design More

**Design the future  
in minutes**

Leverage GenAI and industry expertise  
to create a template that you and your teams can  
build off



## Develop More

**Reduce development effort  
& time by 50%**


Speed up end-to-end project delivery  
and integrate GenAI  
without security & compliance risks



## Do More

**Make every employee  
your best employee**

Unlock knowledge, automation, and coaching

A group of five diverse people (three women and two men) are gathered around a wooden table in a modern office setting. They are looking at a large flowchart or diagram spread across the table. One man in a blue sweater is pointing at a part of the diagram with a pen. The diagram includes various boxes, arrows, and text, some of which are highlighted in yellow and red. There are also sticky notes and a smartphone on the table. The background shows a window with a view of greenery outside.

**Gathering requirements can take weeks... and lots of meetings**  
**...what if you could write requirements in minutes?**





# Design the Future in Minutes: Pega Genai™ Blueprint

⚡ Inference API ⓘ

📄 Text-to-Image

Examples ▾

applying for a loan

Compute

Computation time on [object Object]: 9.941 s

## APPLINGÖN



⚡ Inference API ⓘ

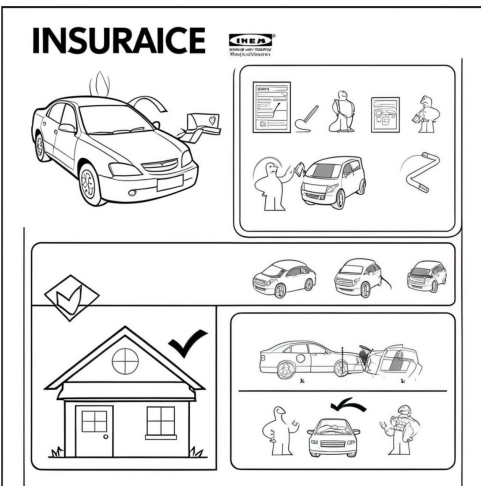
📄 Text-to-Image

Examples ▾

insurance claim approval

Compute

Computation time on [object Object]: 9.704 s



⚡ Inference API ⓘ

📄 Text-to-Image

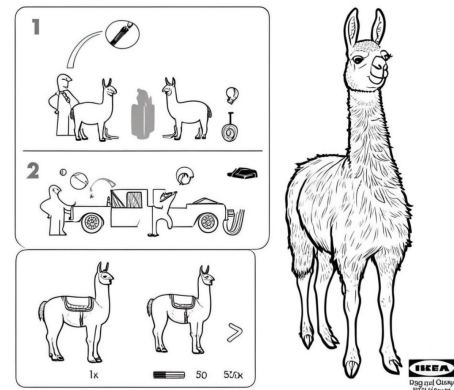
Examples ▾

llama rental

Compute

Computation time on [object Object]: 9.975 s

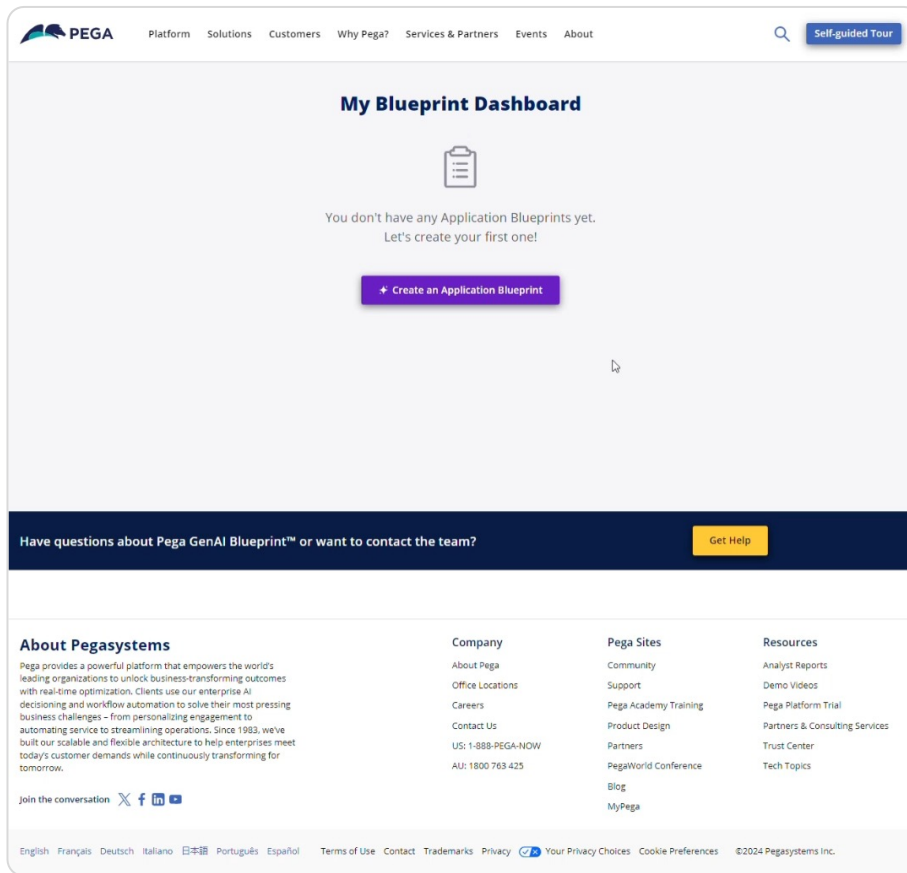
## LÅMARENIL



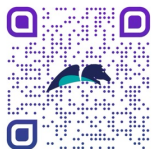
# Design the Future in Minutes: Pega GenAI™ Blueprint

## Accelerate app prototyping & fusion team delivery

- Start quickly with AI powered suggestions and industry expertise
- Build out the end-to-end customer journey including workflows, data, and personas
- Invite business & IT collaborators to edit
- Export blueprints to present & share across the enterprise
- Take blueprints into instant solution prototypes



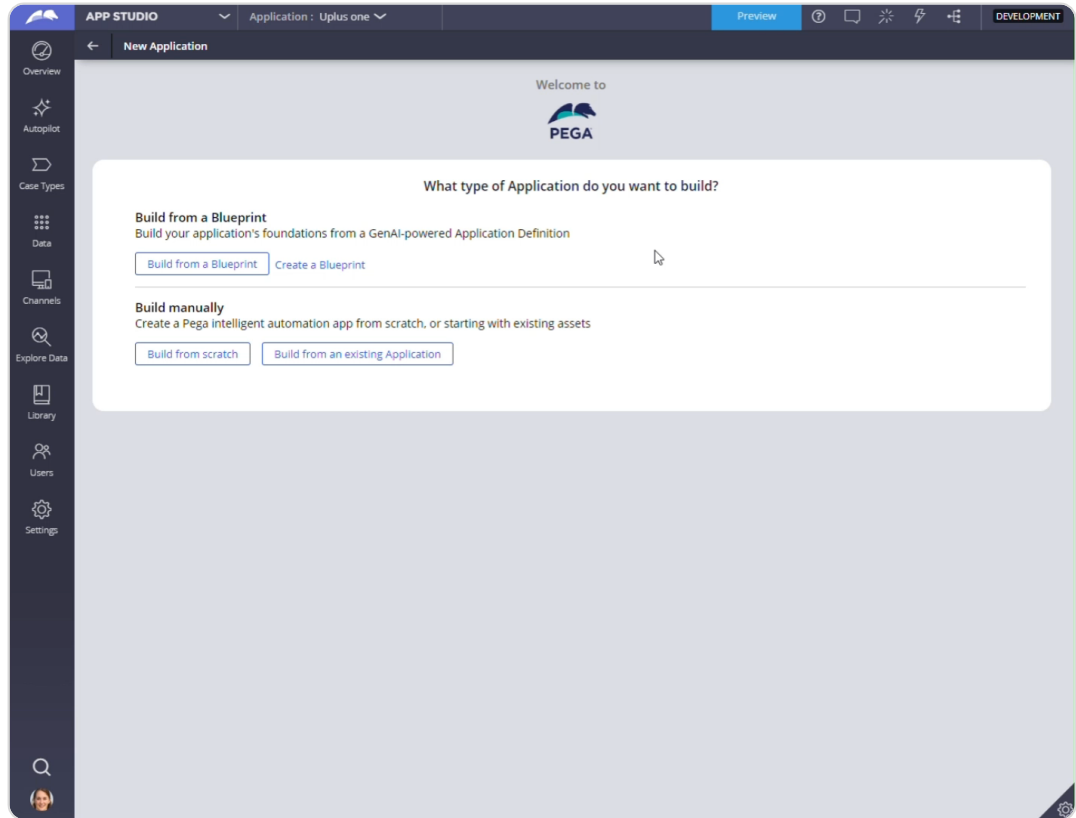
The screenshot displays the Pega GenAI Blueprint dashboard. At the top, there is a navigation bar with the PEGA logo and links for Platform, Solutions, Customers, Why Pega?, Services & Partners, Events, and About. A search icon and a 'Self-guided Tour' button are also present. The main content area is titled 'My Blueprint Dashboard' and features a document icon. Below the icon, a message states: 'You don't have any Application Blueprints yet. Let's create your first one!'. A prominent purple button with a plus sign and the text 'Create an Application Blueprint' is centered on the page. At the bottom of the dashboard, a dark blue banner contains the text 'Have questions about Pega GenAI Blueprint™ or want to contact the team?' and a yellow 'Get Help' button. Below the dashboard, there is a footer section with four columns: 'About Pegasystems' (describing the platform's capabilities), 'Company' (listing contact information for US and AU), 'Pega Sites' (listing various resources like Community, Support, and Training), and 'Resources' (listing Analyst Reports, Demo Videos, and Trial offers). Social media icons for X, Facebook, LinkedIn, and YouTube are also present. The footer includes language options, legal links (Terms of Use, Contact, Trademarks, Privacy), and copyright information for 2024 Pegasystems Inc.



# Design the Future in Minutes: Pega GenAI™ Blueprint

## Accelerate app prototyping & fusion team delivery

- Start quickly with AI powered suggestions and industry expertise
- Build out the end-to-end customer journey including workflows, data, and personas
- Invite business & IT collaborators to edit
- Export blueprints to present & share across the enterprise
- Take blueprints into instant solution prototypes



# How Will You Put Limitless Creativity to Work?



## Design More

**Design the future  
in minutes**

Leverage GenAI and industry expertise to create a template that you and your teams can build off



## Develop More

**Reduce development effort  
& time by 50%**

Speed up end-to-end project delivery and integrate GenAI without security & compliance risks



## Do More

**Make every employee  
your best employee**

Unlock knowledge, automation, and coaching



**Demand for  
developer skills is  
higher than  
ever...and growing**  
**...what if you could double  
developer productivity?**




# Jumpstart App Development with Pega GenAI™ Autopilot



**Pega GenAI™ Autopilot**

Can you build an authorization exceptions process using this process document as a basis?


 Client ABC Proposal.pdf  
157 KB

Leverage assets in my enterprise reuse library where possible!

All set! **Check out your new workflow.**

I was able to reuse the **Customer module** from your reuse library.

Thank you!



Enable developers to get up-to-speed quickly and build & change enterprise workflows in seconds, just by asking.

## Learn

- Instantly access technical guidance for building in Pega.
- Obtain specific insights into your app architecture promptly.

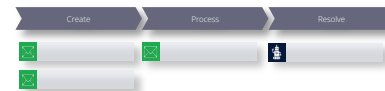
### How do I build a Chat Bot?

Here are the steps you'll need to follow: design the conversation flow, train the chatbot, and integrate with channels.

Start

## Build & change

- Let Autopilot handle the heavy lifting - from developing new apps to enhancing existing ones.
- Automate capability additions and adaptations

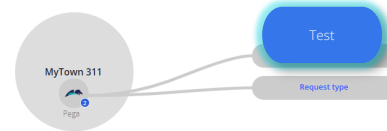


Go

## Test & deploy

Use Autopilot for:

- Automated test scenario creation and execution.
- Documenting application functionality.
- Automating deployment to production through CI/CD pipelines



# Jumpstart App Development with Pega GenAI™ Autopilot

## Learn

- Instantly access technical guidance for building in Pega.
- Obtain specific insights into your app architecture promptly.

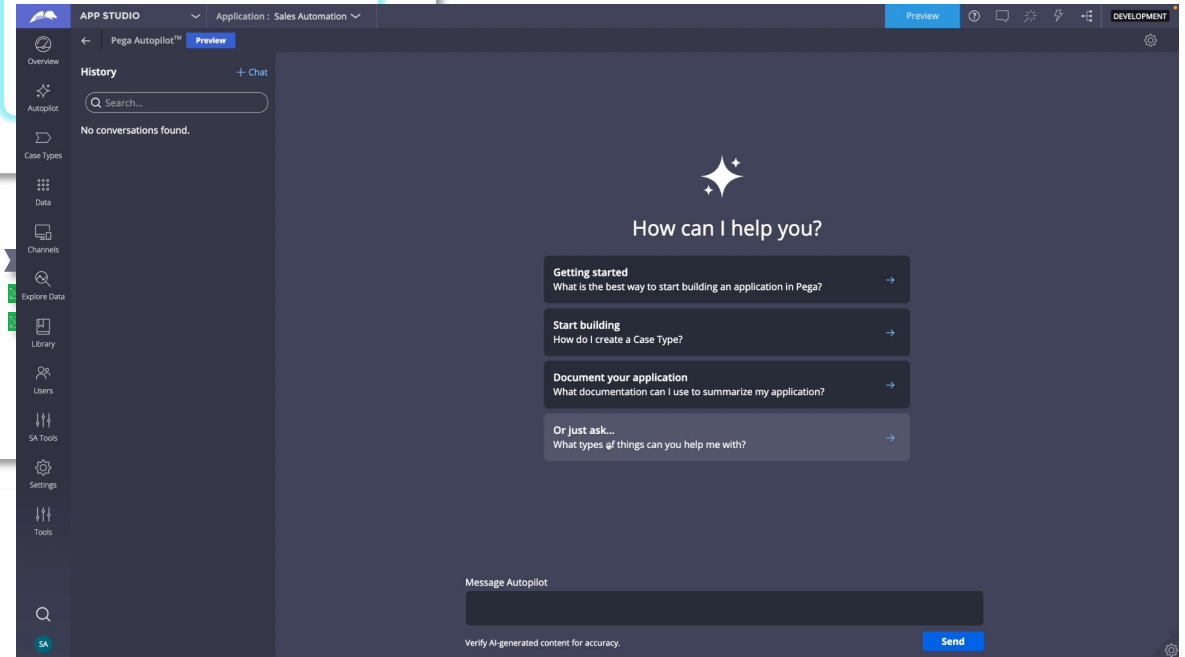
## Build & change

- Let Autopilot handle the heavy lifting - from developing new apps to enhancing existing ones.
- Automate capability additions and adaptations

## Test & deploy

Use Autopilot for:

- Automated test scenario creation and execution.
- Documenting application functionality.
- Automating deployment to production through CI/CD pipelines



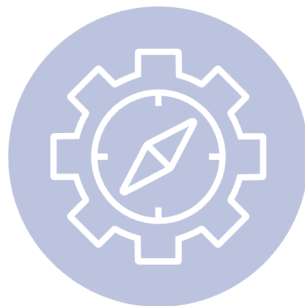
# How Will You Put Limitless Creativity to Work?



## Design More

**Design the future  
in minutes**

Leverage GenAI and industry expertise to create a template that you and your teams can build off



## Develop More

**Reduce development effort  
& time by 50%**

Speed up end-to-end project delivery and integrate GenAI without security & compliance risks



## Do More

**Make every employee  
your best employee**

Unlock knowledge, automation, and coaching





**Getting work done  
in a large enterprise  
can be complex**

**... what if you could  
make every employee  
your best employee?**



# What If You Could Make Every Employee Your Best Employee?

Always know the next best step to take

Immediately understand the context of a case

Easily handle the volume & variety of inbound requests

Instantly unlock answers trapped within webpages and PDFs

Quickly adapt to product, regulatory and market changes

Get it right first time

The image shows a customer service representative in a headset, smiling, positioned in front of a large, semi-transparent screenshot of a PEGA customer service interface. The interface displays a member's profile, a list of tasks with 'Go' buttons, a chat conversation, and a prediction widget.

**Member ID:** M20668491900  
**Status:** IN PATIENT  
**DOB:** Apr 2, 1991  
**Age:** 29  
**Gender:** Female

**Tasks:**

- Comprehensive metabolic panel order request
- Coverage re-adjustment (Due in 3 days • Medium priority)
- Co-pay re-adjustment (Task in ACCT-2913 • Due in 4 days • High priority)

**Chat Conversation:**

Hey team!

Here's a copy of the verification letter we need. Elizabeth Peterson sent it over on April 17, 2024.

Thanks, Sidney Lamas

**Team:**

- Tadd Miller (Consultant)
- Aya Bhiladvala (Advisor)
- Sidney Lamas (Agent)

**Prediction:**

Probability of missing SLA: 70.93%

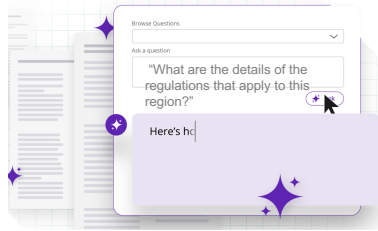
Next bill amount: \$130.23

Anything else I can help with?



# Make Every Employee Your Best Employee

## Scale expertise

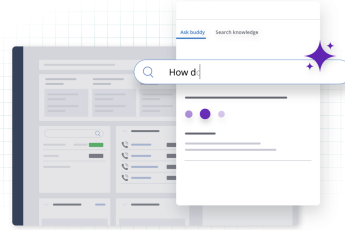


Ask your workflows

### Pega GenAI™ Coach

Scales expertise by directing employees with AI-powered guidance based on your specific best practices & policies.

## Gain instant knowledge

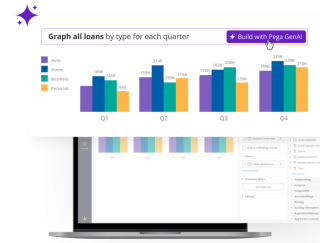


Ask your documents

### Pega GenAI™ Knowledge Buddy

Connects employees with instant answers from disparate data sources via an easy-to-use conversation interface.

## Make data work for you

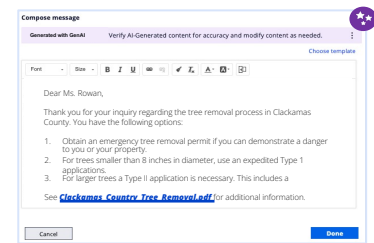


Chat with your data

### Pega GenAI™ Analyze

Turns GenAI-powered conversational prompts into reports that tell a cohesive story and inform high-impact action.

## Focus on what matters



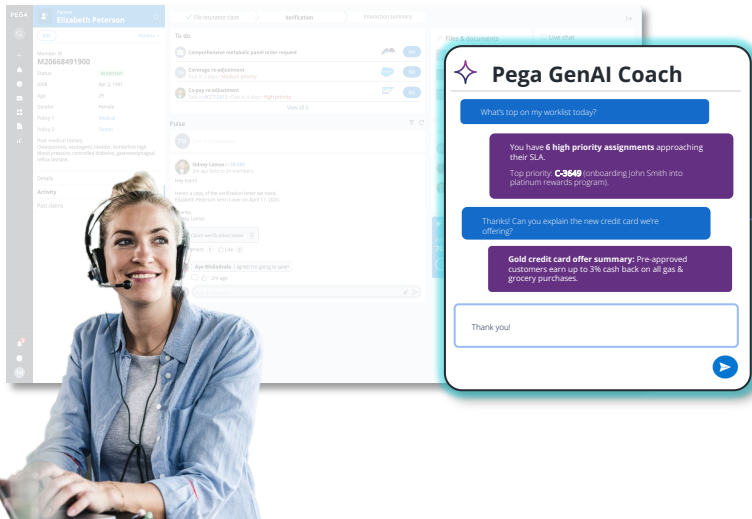
Automate tedious tasks

### Pega GenAI™ Automate

Enhances employee productivity by automating manual work like composing emails and chat messages.



# All these capabilities can be called by **Pega GenAI™ Coach**



Receive instant answers with **Pega GenAI™ Knowledge Buddy**

[View full article](#)

**Gold credit card offer summary**  
Pre-approved customers earn up to 3% cash back on all gas & grocery purchases.

What **policies & procedures** do I need to follow?

What are the **best practices** for assignments like this?

Gain greater insights with **Pega GenAI™ Analyze**



“Show me **process** sorted by type”

“What is the **average interest rate** of each loan?”

Get more done faster with **Pega GenAI™ Automate**

[Send Email](#)

**Compose message**  
Generated with GenAI

Verify AI Generated content for your use.

Dear [Name],

Thank you for contacting our Home Insurance policy. We have received your request for a quote and would like to provide you with the following details:

Coverage Type: Comprehensive  
 Coverage Amount: \$500,000  
 Premium Type: Monthly  
 Policy Status: Active  
 Please let us know if you have any questions or concerns. We look forward to hearing back from you.

Best regards,  
[Name/Title]

Cancel [Done](#)

“Draft **email** with GenAI”

# Pega GenAI™ Coach

Unlock your employees' full potential with contextual recommendations

- Understand how to progress every workflow more efficiently and effectively
- Quickly obtain context to the status of and priorities within a case.
- Get expert advice and recommendations to perform work more efficiently.
- Onboard new employees and users faster by providing them with instant context.

The screenshot displays the Pega Sales Automation interface. The top navigation bar shows the workflow stages: Create, Qualification, Analysis, Proposal, Decision, Negotiation, and Closed. The current case is 'Business Voice over IP (VoIP) for Acme Tech' (SampleOPP-1). The left sidebar lists various case details and navigation options, with 'GenAI Coach' highlighted. The main content area shows 'Assignments' (e.g., 'Capture customer response', 'Confirm decision date'), 'Next Best Actions' (e.g., 'Convey birthday wishes'), and the 'GenAI Coach' section. The GenAI Coach section features a prompt: 'Please provide a brief summary of recent activity and provide guidance to help me progress this deal.' Below this is a 'Message GenAI Coach' input field with the placeholder 'Enter your question' and a 'Verify AI-generated content for accuracy' note. The right sidebar contains a card for 'ACME Technologies Inc.' with details like 'Software Development', '51-200 employees', and a 'Save in Sales Navigator' button. Below this are 'Predictions' cards showing likelihoods for 'close as won' (87.19%), 'move to next stage' (91.83%), and 'close on time' (47.39%). At the bottom right, there is a 'Knowledge articles' section with a search bar and a list of articles, including 'NTT Communications Business Messaging Software Application' and 'Business Voice over VoIP'.

# Pega GenAI™ Knowledge Buddy

- Deliver self-service
  - by enabling answers to inquiries directly in customer-facing web & mobile sites.
- Guide agents
  - during interactions with contextual answers & instruction to better serve customers.
- Accelerate operations
  - and onboard new employees faster by unlocking insight from procedure documents.

## Knowledge Buddy 🙌

Q&A History

Browse questions

Browse through sample questions

Ask a question ⓘ

Hey Buddy! How do I process a cross-border payment from a customer in the US to a merchant in the UK?

Clear Ask

🌟 Hello! Here are the steps for **cross-border payment processing**:

- Verify the **customer's identity**
- Check the **exchange rate and fees**
- Obtain the **merchant's bank details**, such as name, address, IBAN, and SWIFT code.
- **Initiate** the payment!

Reference: [Global Payment Processing Standard Operating Procedure.docx](#)

# Do More with Pega GenAI™



## Design More

**Design the future  
in minutes**

Leverage GenAI and industry expertise to create a template that you and your teams can build off



## Develop More

**Reduce development effort  
& time by 50%**

Speed up end-to-end project delivery and integrate GenAI without security & compliance risks



## Do More

**Make every employee  
your best employee**

Unlock knowledge, automation, and coaching

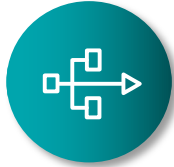
# Do More with Pega GenAI™



Discover  
more

Pega GenAI™  
**Blueprint™**

**Design  
More**



Build  
more

Pega GenAI™  
**Autopilot**



Custom  
GenAI

Pega GenAI™  
**Connect**

**Develop  
More**



OnPrem  
Gen AI

Pega GenAI™  
**PremBridge**



Accomplish  
more

Pega GenAI™  
**Coach**



Answer  
more

Pega GenAI™  
**Knowledge  
Buddy**

**Do  
More**



Understand  
more

Pega GenAI™  
**Analyze**





# Sneak peek into the labs: Where are we going with this?

Towards autonomous generative AI

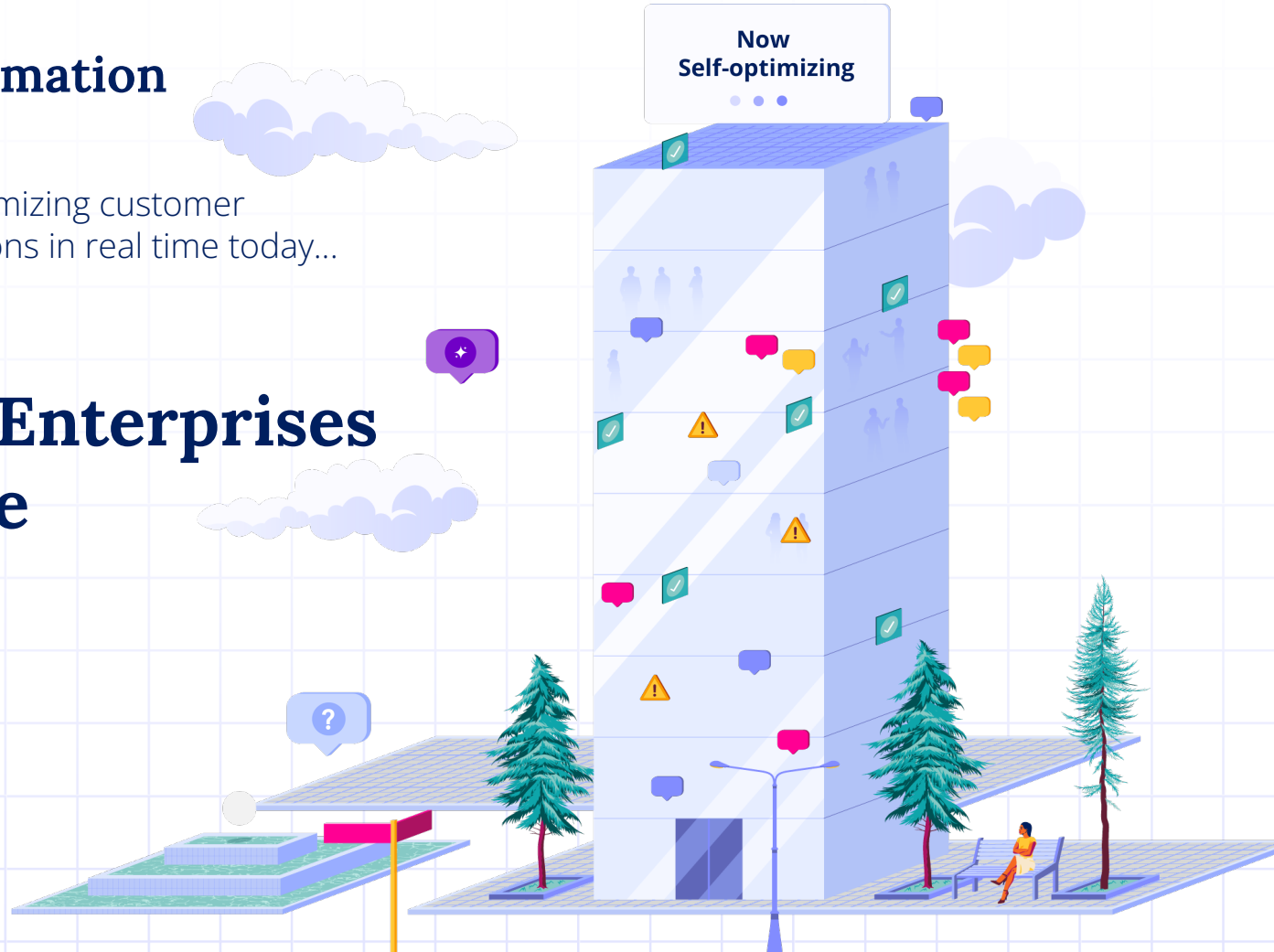


# Your AI transformation *can't wait*

The leaders who are optimizing customer experiences and operations in real time today...

will build the

# ***Autonomous Enterprises*** **of the future**





Generate low-code apps



Optimize customer engagement



Streamline customer service

C



Automate sales rep activities



Build Insights



Custom AI-powered capabilities



sensitive data filtering

invocation logging

model mapping

authentication & pricing

provisioning, scaling, usage limitations





## Engineered Prompt

Automated prompt engineering,  
dynamic prompt templates, single call



# Example: Treatment Generation



## Automated insights

### 💡 Create treatments for Privacy first hesitant buyer

✦ Use and refine

Privacy first hesitant buyer have a low response rate for this action. Having treatments that appeal to this customer type may increase your business objective of increasing Sales of UmbrellaPolicies. We have generated sample treatment options for multiple channels.

Selecting "Progress to plan stage" will initiate the process to update this action, where you can review, accept, reject, edit or reconfigure the recommended content.

#### Recommended content configuration

Channels	Web
Principle	Cialdini's principle of authority



Let us help you take care of the ones you love.

[Preview treatment suggestions](#)

## System message:

You are a creative content writer who gives content for marketing materials depending on the channel on which offer is presented.

If asked for, you can also write descriptions of images that would be a best description of the promotional image for the offer to be shown in the requested channels.

If image description is requested, the image description should be sufficiently detailed in a single sentence such that it describes the scene on the image.

Do not include any HTML tags in the content that you generate.

You can tailor your writing style that best works for a user-specified audience.

You always return a JSON object. This JSON object should contain only the fields requested by user. Adherence to this format is mandatory.

It is forbidden to mention any customer type or target audience name or description or the guiding principle you are using or the tone explicitly.

## Prompt:

Audience: {.Audience}

Audience Description: {.AudienceDescription}

Action name: {.ActionName}

Action description: {.ActionDescription}

Action benefits: {.ActionBenefits}

Action channel: {.ActionChannel}

Action issue: {.ActionIssue}

Action group: {.Action group}

Language: {.Language}

Fields to generate: JSON Object containing fields  
{.FieldsToGenerate}

While generating the content make sure you apply  
{.Cialdini's principle} and use {.Tone}



## Engineered Prompt

Automated prompt engineering,  
dynamic prompt templates, single call



## Basic Tool Use

1-2 tools (f.i. search engine +  
knowledge), single call, scripted use





# RAGs are all the Rage



- Problems with search and generative AI when answering a question
  - Generative AI suffers from hallucination and a scope that is too general
  - Search bombards a user with search results they need to trawl through
- Retrieval-augmented generation combines the power of generative AI with the strengths of semantic search
  - Take the question, search the corpus, use generative AI to answer the question based on search results
- More generally, RAGs are an example of giving generative AI access to a tool
  - Allbeit in a scripted manner

---

## Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks

---

Patrick Lewis<sup>1,‡</sup>, Ethan Perez<sup>\*</sup>,

Aleksandra Piktus<sup>1</sup>, Fabio Petroni<sup>1</sup>, Vladimir Karpukhin<sup>1</sup>, Naman Goyal<sup>1</sup>, Heinrich Küttler<sup>1</sup>,

Mike Lewis<sup>1</sup>, Wen-tau Yih<sup>1</sup>, Tim Rocktäschel<sup>1,‡</sup>, Sebastian Riedel<sup>1,‡</sup>, Douwe Kiela<sup>1</sup>

<sup>1</sup>Facebook AI Research; <sup>‡</sup>University College London; <sup>\*</sup>New York University;  
plewis@fb.com

### Abstract

Large pre-trained language models have been shown to store factual knowledge in their parameters, and achieve state-of-the-art results when fine-tuned on downstream NLP tasks. However, their ability to access and precisely manipulate knowledge is still limited, and hence on knowledge-intensive tasks, their performance lags behind task-specific architectures. Additionally, providing provenance for their decisions and updating their world knowledge remain open research problems. Pre-trained models with a differentiable access mechanism to explicit non-parametric memory have so far been only investigated for extractive downstream tasks. We explore a general-purpose fine-tuning recipe for retrieval-augmented generation (RAG) — models which combine pre-trained parametric and non-parametric memory for language generation. We introduce RAG models where the parametric memory is a pre-trained seq2seq model and the non-parametric memory is a dense vector index of Wikipedia, accessed with a pre-trained neural retriever. We compare two RAG formulations, one which conditions on the same retrieved passages across the whole generated sequence, and another which can use different passages per token. We fine-tune and evaluate our models on a wide range of knowledge-intensive NLP tasks and set the state of the art on three open domain QA tasks, outperforming parametric seq2seq models and task-specific retrieve-and-extract architectures. For language generation tasks, we find that RAG models generate more specific, diverse and factual language than a state-of-the-art parametric-only seq2seq baseline.

# RAGs are all the Rage: Knowledge Buddy example (SelfStudy Buddy)



- Problems with search and generative AI when answering a question
  - Generative AI suffers from hallucination and a scope that is too general
  - Search bombards a user with search results they need to trawl through
- Retrieval-augmented generation combines the power of generative AI with the strengths of semantic search
  - Take the question, search the corpus, use generative AI to answer the question based on search results
- More generally, RAGs are an example of giving generative AI access to a tool
  - Allbeit in a scripted manner

# RAGs are all the Rage: Knowledge Buddy example (SelfStudy Buddy)



- Problems with search and generative AI when answering a question
  - Generative AI suffers from hallucination and a scope that is too general
  - Search bombards a user with search results they need to trawl through
- Retrieval-augmented generation combines the power of generative AI with the strengths of semantic search
  - Take the question, search the corpus, use generative AI to answer the question based on search results
- More generally, RAGs are an example of giving generative AI access to a tool
  - Allbeit in a scripted manner

## Details

Question ID	ASK-822046
Question	How does Pega's case management functionality support complex business processes, and what are the key features that enable this support?
Answer	<p>Pega's case management functionality supports complex business processes by providing tools to implement goal-oriented solutions through defining reusable templates for business processes. This allows for flexibility in reacting to changing circumstances while guiding users to a successful resolution. The key features that enable this support include business process visualization, automation and optimization of workflows and business rules, collaboration tools for efficient team work, reporting and analytics tools for performance insights, and a unified cross-platform experience for maintaining context across interaction channels.</p> <p>References: - <a href="https://docs.pega.com/bundle/platform/page/platform/case-management/getting-started-introduction-case-management.html">https://docs.pega.com/bundle/platform/page/platform/case-management/getting-started-introduction-case-management.html</a> - <a href="https://docs.pega.com/bundle/customer-service/page/customer-service/constellation-implementation/get-started/CME-intro.html">https://docs.pega.com/bundle/customer-service/page/customer-service/constellation-implementation/get-started/CME-intro.html</a> - <a href="https://docs.pega.com/bundle/customer-service/page/customer-service/uikit-implementation/get-started/CS-main-CME.html">https://docs.pega.com/bundle/customer-service/page/customer-service/uikit-implementation/get-started/CS-main-CME.html</a></p>

# RAGs are all the Rage: Knowledge Buddy example (SelfStudy Buddy)



- Problems with search and generative AI when answering a question
  - Generative AI suffers from hallucination and a scope that is too general
  - Search bombards a user with search results they need to trawl through
- Retrieval-augmented generation combines the power of generative AI with the strengths of semantic search
  - Take the question, search the corpus, use generative AI to answer the question based on search results
- More generally, RAGs are an example of giving generative AI access to a tool
  - Allbeit in a scripted manner

## Details

### Instructions

You are a intelligent assistant answering technical product questions about Pega's products and I would like you to answer some questions based on the CONTEXT I provide. I have some rules you need to follow. Please follow all the rules.

1. Use only the CONTEXT you are provided with.
2. In your answer do not mention you are a customer service representative.
3. Do not refer to anyone or anything that is not part of the CONTEXT provided.
4. If you don't know the answer, just say that you don't know.
5. Never make up answers which is not in the CONTEXT provided.
6. Address the user by the name

I want you to include a references section that includes the URLs that you used to answer the question.

Output format      Custom

### Information

CONTEXT:  
{SEARCHRESULTS}

The user asking the question is {USERNAME}

Question to answer:  
{QUESTION}

# RAGs are all the Rage: Knowledge Buddy example (SelfStudy Buddy)



- Problems with search and generative AI when answering a question
  - Generative AI suffers from hallucination and a scope that is too general
  - Search bombards a user with search results they need to trawl through
- Retrieval-augmented generation combines the power of generative AI with the strengths of semantic search
  - Take the question, search the corpus, use generative AI to answer the question based on search results
- More generally, RAGs are an example of giving generative AI access to a tool
  - Allbeit in a scripted manner

## Prompt

You are a intelligent assistant answering technical product questions about Pega's products and I would like you to answer some questions based on the CONTEXT I provide. I have some rules you need to follow. Please follow all the rules.

1. Use only the CONTEXT you are provided with.
2. In your answer do not mention you are a customer service representative.
3. Do not refer to anyone or anything that is not part of the CONTEXT provided.
4. If you don't know the answer, just say that you don't know.
5. Never make up answers which is not in the CONTEXT provided.
6. Address the user by the name

I want you to include a references section that includes the URLs that you used to answer the question.

## ## To Avoid Copyright Infringements

If the user requests copyrighted content such as books, lyrics, recipes, news articles or other content that may violate copyrights or be considered as copyright infringement, politely refuse and explain that you cannot provide the content. Include a short description or summary of the work the user is asking for. You **must not** violate any copyrights under any circumstances.

## CONTEXT:

```
{
  "chunks": [
    {
      "attributes": {
        "values": [
          {
            "value": "https://docs.pega.com/bundle/platform/page/platform/case-management/getting-started-introduction-case-management.html",
            "name": "url",
            "content": "Pega Platform provides tools to implement Case Management as the way to meet your business objectives. You can create goal-oriented solutions by defining reusable templates for business processes that guide users to a successful resolution, but also provide flexibility for an appropriate"
          }
        ]
      }
    }
  ]
}
```



## Engineered Prompt

Automated prompt engineering,  
dynamic prompt templates, single call



## Basic Tool Use

1-2 tools (f.i. search engine +  
knowledge), single call, scripted use



## Autonomous Agent

Agentic, planning, multi-step,  
multi tool, multi agent



# What if the AI Could Decide When to Use What Tool?



- Certain tasks cannot be achieved by just building larger models
- We already saw we can give the AI more agency by giving it access to a tool such as a search engine combined with a knowledge base
- But what if we could give it access to many tools?
  - Tools to sense information
  - Tools to take action
- And to use the creative power of generative AI to
  - Decide when to use what kind of tool
  - Create a multistep plan to reach a goal
  - Dynamically adapt these plans when needed

## Toolformer: Language Models Can Teach Themselves to Use Tools

Timo Schick   Jane Dwivedi-Yu   Roberto Dessì†   Roberta Raileanu  
 Maria Lomeli   Luke Zettlemoyer   Nicola Cancedda   Thomas Scialom

Meta AI Research   †Universitat Pompeu Fabra

### Abstract

Language models (LMs) exhibit remarkable abilities to solve new tasks from just a few examples or textual instructions, especially at scale. They also, paradoxically, struggle with basic functionality, such as arithmetic or factual lookup, where much simpler and smaller models excel. In this paper, we show that LMs can teach themselves to *use external tools* via simple APIs and achieve the best of both worlds. We introduce *Toolformer*, a model trained to decide which APIs to call, when to call them, what arguments to pass, and how to best incorporate the results into future token prediction. This is done in a self-supervised way, requiring nothing more than a handful of demonstrations for each API. We incorporate a range of tools, including a calculator, a Q&A system, a search engine, a translation system, and a calendar. Toolformer achieves substantially improved zero-shot performance across a variety of downstream tasks, often competitive with much larger models, without sacrificing its core language modeling abilities.

The New England Journal of Medicine is a registered trademark of `[QA("Who is the publisher of The New England Journal of Medicine?") → Massachusetts Medical Society]` the MMS.

Out of 1400 participants, 400 (or `[Calculator(400 / 1400) → 0.29]` 29%) passed the test.

The name derives from "la tortuga", the Spanish word for `[MT("tortuga") → turtle]` turtle.

The Brown Act is California's law `[WikiSearch("Brown Act") → The Ralph M. Brown Act is an act of the California State Legislature that guarantees the public's right to attend and participate in meetings of local legislative bodies.]` that requires legislative bodies, like city councils, to hold their meetings open to the public.

Figure 1: Exemplary predictions of Toolformer. The model autonomously decides to call different APIs (from top to bottom: a question answering system, a calculator, a machine translation system, and a Wikipedia search engine) to obtain information that is useful for completing a piece of text.

## 1 Introduction

Large language models achieve impressive zero- and few-shot results on a variety of natural language processing tasks (Brown et al., 2020; Chowdhery et al., 2022, i.a.) and show several emergent capabilities (Wei et al., 2022). However, all of these models have several inherent limitations that can at best be partially addressed by further scaling. These limitations include an inability to access

A simple way to overcome these limitations of today's language models is to give them the ability to *use external tools* such as search engines, calculators, or calendars. However, existing approaches either rely on large amounts of human annotations (Komeili et al., 2022; Thoppilan et al., 2022) or limit tool use to task-specific settings only (e.g., Gao et al., 2022; Parisi et al., 2022), hinder-

# From Reactive LLM Services to Proactive Agents



## Multi-agent system

Article Talk

Read Edit View history Tools

From Wikipedia, the free encyclopedia

A **multi-agent system** (**MAS** or "self-organized system") is a computerized system composed of multiple interacting **intelligent agents**.<sup>[1]</sup> Multi-agent systems can solve problems that are difficult or impossible for an individual agent or a **monolithic system** to solve.<sup>[2]</sup> Intelligence may include **methodic, functional, procedural** approaches, **algorithmic search** or **reinforcement learning**.<sup>[3]</sup>

Despite considerable overlap, a multi-agent system is not always the same as an **agent-based model** (ABM). The goal of an ABM is to search for explanatory insight into the collective behavior of agents (which do not necessarily need to be "intelligent") obeying simple rules, typically in natural systems, rather than in solving specific practical or engineering problems. The terminology of ABM tends to be used more often in the science, and MAS in engineering and technology.<sup>[4]</sup> Applications where multi-agent systems research may deliver an appropriate approach include online trading,<sup>[5]</sup> disaster response,<sup>[6][7]</sup> target surveillance<sup>[6]</sup> and social structure modelling.<sup>[9]</sup>

## Concept [edit]

Multi-agent systems consist of agents and their **environment**. Typically multi-agent systems research refers to **software agents**. However, the agents in a multi-agent system could equally well be robots, humans or human teams. A multi-agent system may contain combined human-agent teams.

Agents can be divided into types spanning simple to complex. Categories include:

- Passive agents<sup>[10]</sup> or "agent without goals" (such as obstacle, apple or key in any simple simulation)
- Active agents<sup>[10]</sup> with simple goals (like birds in flocking, or wolf-sheep in **prey-predator model**)
- Cognitive agents (complex calculations)

Agent environments can be divided into:

- Virtual
- Discrete

Part of a series on

### Multi-agent systems

**Multi-agent simulation**

- Agent-based computational economics ·
- Agent-based model in biology ·
- Agent-based social simulation ·
- Agent-based modeling software

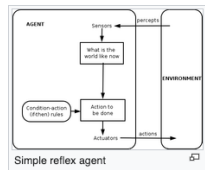
**Agent-oriented programming**

- Auto-GPT · Botnets · FIPA ·
- Platforms for software agents (JADE · JACK · GORITE) · Software agent

**Related**

- Distributed artificial intelligence ·
- Multi-agent pathfinding · Multi-agent planning ·
- Multi-agent reinforcement learning ·
- Self-propelled particles · Swarm robotics

V · T · E

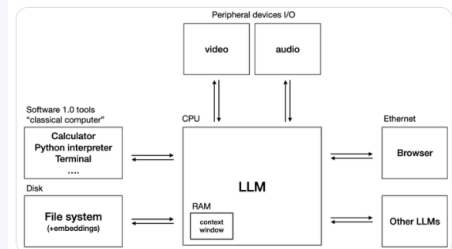


Andrej Karpathy @karpathy

LLM OS. Bear with me I'm still cooking.

Specs:

- LLM: OpenAI GPT-4 Turbo 256 core (batch size) processor @ 20Hz (tok/s)
- RAM: 128Ktok
- Filesystem: Ada002

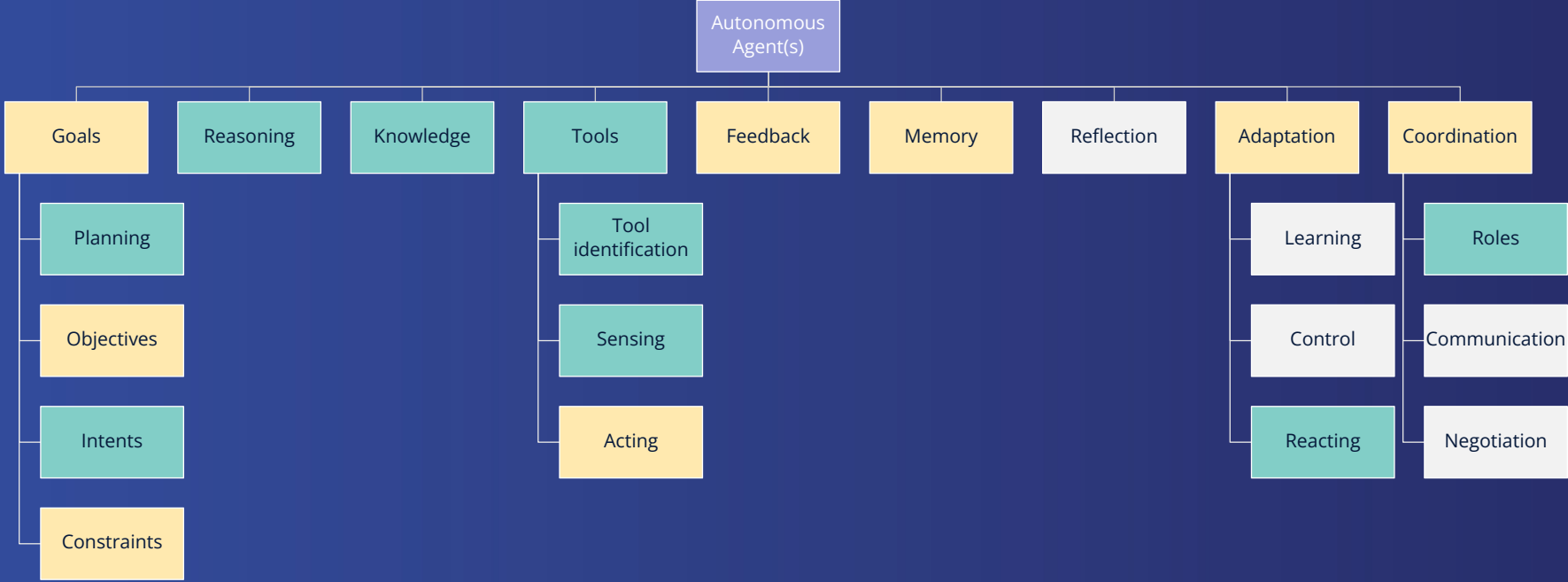


Last edited 1:48 AM · Nov 11, 2023 · 2.2M Views

1,247 Reposts 297 Quotes 10K Likes 3,344 Bookmarks

🗨️ 🔄 ❤️ 📌 3.3K 📤

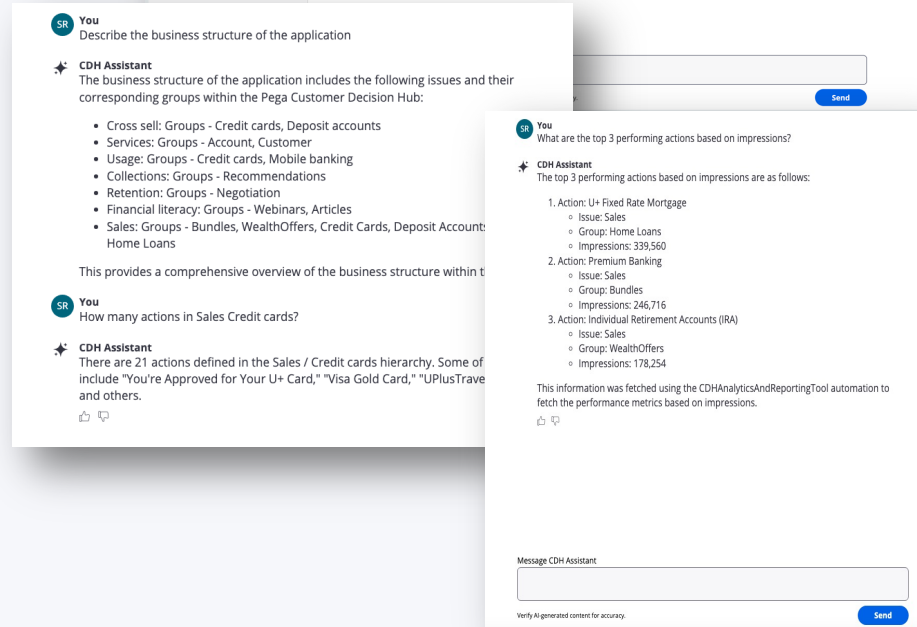
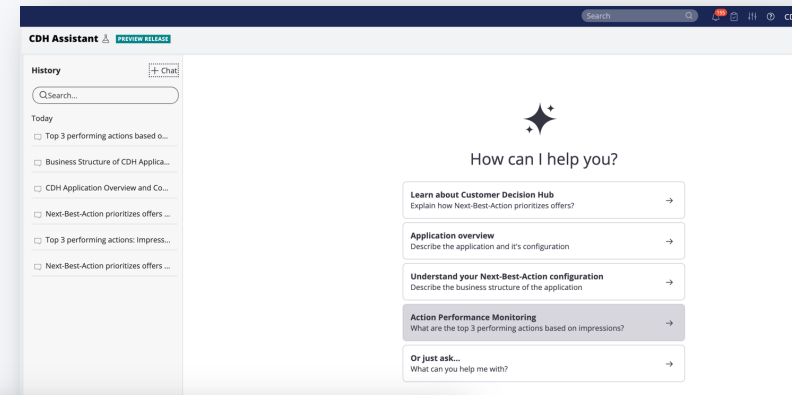
# Autonomous Agents Lead to the Autonomous Enterprise



# Example: Customer Decision Hub Assistant



- Customer Decision Hub (CDH) Assistant helps customer engagement marketers foster best practice, maximize business performance & customer relevance, and get work done
- How?
  - It interprets user questions posed in natural language text, in real-time
  - It uses application information and product documentation to provide the best response
  - In future it will also be able to take actions with the human in the loop for approval



# Example: Customer Decision Hub Assistant



- Under the hood: agent with planning capabilities, with around a dozen 'automations'
- Based on user intent and context, the assistant can choose to use:
  - CDH Knowledge Buddy
    - Answers questions based on product documentation (CDH, decisioning, best practice delivery & templates)
  - CDH Discovery: configuration
    - NBA hierarchy & rule configurations
    - Channels, containers, application overview
  - CDH Discovery: analytics
    - Real time performance data
- More tools = more capable

```
name: CDHAssistant
plugins:
  - CDHDiscovery
  - BuddyPlugin
model: gpt-35-turbo
```

```
system_message: |
  Your name is CDH Assistant. You are an expert in Pega Customer
  Decision Hub(CDH) framework.
```

As an assistant you can help the user of applications built on CDH in below areas:

1. Lookup the information from the application including but not limited to:
    - 1.1 Application definition – Comprehensive overview of the application.
    - 1.2 Application data – Any application specific information about:
      - 1.2.1 Business structure and Taxonomy
      - 1.2.2 Containers
      - 1.2.3 Channels
      - 1.2.4 Context dictionary
      - 1.2.5 Actions
      - 1.2.6 Rulesets
      - 1.2.7 Customer count
    - 1.3 Configurations – Gives information about Next-Best-Action-Designer is configured.
    - 1.4 Action performance monitoring of Next-Best-Actions and Offers
  2. Provide information about the metrics and how actions are performing with respect to click accept rate, sends, impressions, clicks and accepts.
  2. Provides explanations and describes various aspects of Customer Decision Hub including step by step instructions.
- You can choose from below plugins to help you find the right responses:
1. BuddyPlugin: Provides information from the documentation.
  2. CDHDiscovery: Provides information specific to the current application.

Follow these steps:

- 1) Analyze user query and gather relevant inputs by asking user.
- 2) Invoke Plugins step by step to serve user request.
- 3) Support the user till request is complete and provide the appropriate result from plugin. If the information is already present in the conversation, do not call any plugin.

Rules:

- Always ask user for missing inputs as required by Plugins.
- Never make up the answer if plugins cannot provide an answer. Use Buddy plugin to fetch the relevant information from Pega Customer Decision Hub documentation.

# Example: Customer Decision Hub Assistant



- User Question:
  - What is the accept rate of Premier Checking?
- Understand goal:
  - Find the accept rate of the Premier Checking action.
- Generate plan:
  1. Validate if it is an action, group or issue
    - Using automation  
CDHGetIssueAndGroupForAction
  2. Find the accept rate for the Premier Checking action
    - Using automation  
CDHAnalyticsAndReportingTool

```
▼ history:
  0:
    role: "system"
    content: "You are a brilliant plan executor. Your task is to execute each step in a plan created by planner agent and provide the final answer for the user goal. You are assisting the user of applications built on Pega's Customer Decision Hub (CDH). Important notice: Always look at output of function pega_context to know more about CDH. Rules: Analyze user query and plan, use names of Issue and Groups from <issue_group_hierarchy>. For multiple matches, misspelling, confirm with the user. If a goal cannot be automated by the planner first, report back to user saying so along with the specified reason. Only map values to inputs from user query. NEVER use automation to fetch data. End output: Execute plan sequentially. Functions are executed by infinity agent. Instructions are executed by you. When fetching the application overview, convert it into nice user readable paragraphs so that its more appealing. Always check the issue_group_hierarchy to figure out the right issue and group if needed. Do not pass wrong issue and group names to automations. If issue and group are not present in the query, ask user for their id needed by automation. End output with a complete report of the goal, all automations executed with their inputs, inputs collected from the user, and status of execution\n\n"
  1: {}
  2:
    role: "function"
    name: "issue_group_hierarchy"
    content: [{"hierarchy": [{"issueName": "Cross sell", "groups": [{"Credit cards", "Deposit accounts"}]}, {"issueName": "Services", "groups": [{"Account", "Customer"}]}, {"issueName": "Mobile banking", "groups": [{"Account", "Collections"}]}, {"issueName": "Retention", "groups": [{"Negotiation"}]}, {"issueName": "Financial Literacy", "groups": [{"Webinars", "Articles"}]}, {"issueName": "Sales", "groups": [{"Bundles", "Health offers", "Credit Cards", "Deposit Accounts", "Auto Loans", "Home Loans"}]}]}]
  3:
    role: "function"
    name: "plan"
    content: [{"goal": "Find the accept rate of the Premier Checking action.", "can_automate": "YES", "reasoning": "The goal can be automated using the CDHAnalyticsAndReportingTool automation to fetch the accept rate for the Premier Checking action.", "step": [{"number": 0, "thought": "Premier Checking action is not an Issue or a Group in the issue_group_hierarchy. Then I need to find it as an action.", "action": {"type": "Automation", "name": "CDHGetIssueAndGroupForAction", "inputs": {"action": "Premier Checking", "output": "<List of Issues and Groups>"}, {"number": 1, "thought": "Using the accept rate for the Premier Checking action.", "action": {"type": "Automation", "name": "CDHAnalyticsAndReportingTool", "inputs": {"Issue": "name of the issue from <List of Issues and Groups>", "Group": "name of the group from <List of Issues and Groups>", "Action": "Premier Checking", "Metric": "accept rate", "output": "Performance of the action"}}, "all_automations_used": [{"name": "CDHGetIssueAndGroupForAction", "applies_to_class": "@baseclass", "kind": "Automation"}, {"name": "CDHAnalyticsAndReportingTool", "applies_to_class": "@baseclass", "kind": "Automation"}, {"missing_inputs": ""}]}]}]
  4: {}
  5: {}
  6: {}
  7: {}
  8: {}
  9:
    content: "The accept rate for the Premier Checking action under the Sales issue and Deposit Accounts group is 100%.\n\nReport:\n\nThe Premier Checking action was found under the Sales issue and Deposit Accounts group.\n\nThe Premier Checking action was successfully retrieved and it is 100%.\n\nAll steps in the plan were successfully executed."
    role: "assistant"
```

System message

Configuration

Plan

Execution

Final answer

# Pega AutoPilot

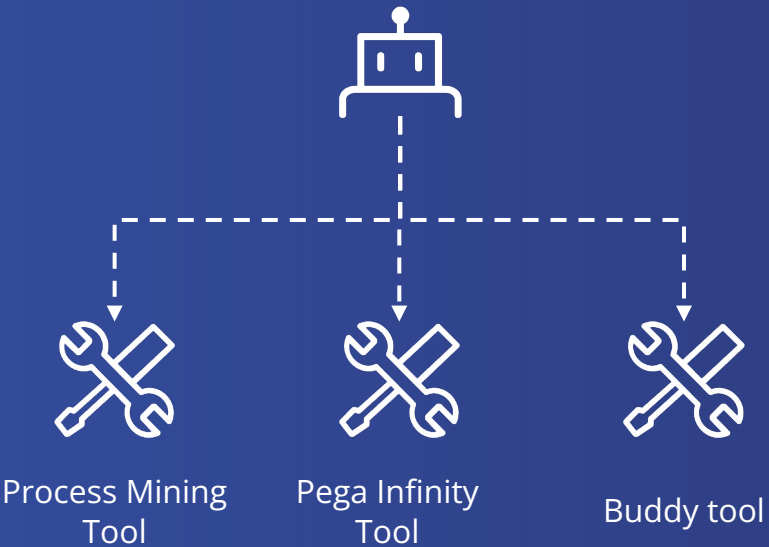
for business



 Pega AutoPilot

Send





## GenAI™ Coach

Optimize my claims process to minimize processing time while maintaining NPS.

...

Optimize my claims process to minimize processing time while maintaining NPS.





# Process Mining Tool

```
callProcessMiningTool(
```

```
    question = "Find bottlenecks in the  
    Claims process"
```

```
)
```

- Specialised Process Mining agent
- API calls to Pega Process Mining
- Real time data and analysis to understand bottlenecks & find root causes
- But for the main coach: it's just another tool

The screenshot shows the GenAI™ Coach interface. At the top left is a purple star icon followed by the text "GenAI™ Coach". Below this is a white chat area containing a grey message bubble with the text: "Optimize my claims process to minimize processing time while maintaining NPS." Below the message bubble is a purple button with a white hammer icon and the text "Executing... Process Mining". At the bottom of the interface is a white input field and a blue "Send" button.



# Process Mining Tool

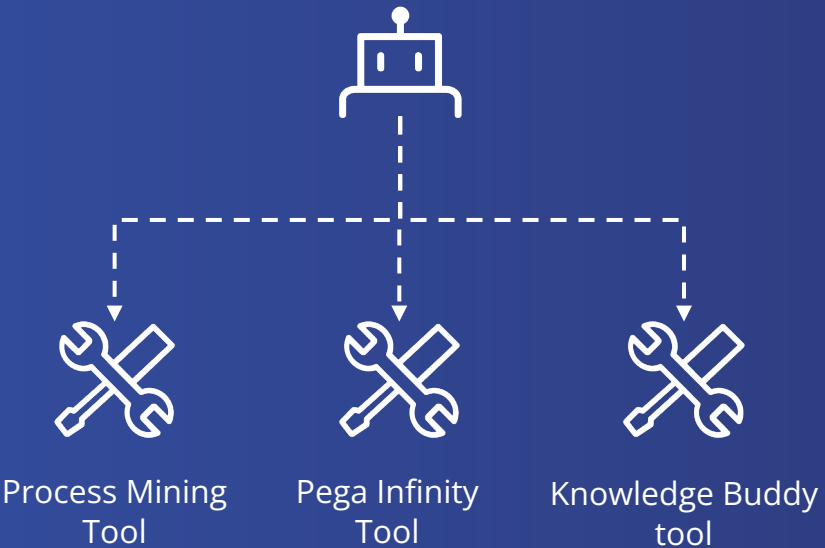
```
callProcessMiningTool(  
    question = "Find bottlenecks in the  
    Claims process"  
)
```

## Output:

The claims process has three major bottlenecks:

- ...
- 75% of claims are initially misrouted to a suboptimal agent, leading to an average delay of 25%
- ...

The screenshot shows the GenAI™ Coach interface. At the top, there is a purple star icon and the text "GenAI™ Coach". Below this, a grey rounded rectangle contains the user's query: "Optimize my claims process to minimize processing time while maintaining NPS." Underneath the query is a purple button with a white hammer icon and the text "Executing... Process Mining". At the bottom of the interface, there is a large white input field and a blue "Send" button.



## GenAI™ Coach

Optimize my claims process to minimize processing time while maintaining NPS.

...

Send



# Knowledge Buddy Tool

```
callKnowledgeBuddy(
```

```
    buddyName = "ProcessBestPracticesBuddy",  
    question = "In my Claims process, 75% of  
cases are routed to an incorrect agent.  
Can you suggest ways to improve this  
bottleneck?"
```

```
)
```

- Calls to Pega Knowledge Buddy directly
- Wide range of documents around best practices for process improvements, and Pega product documentation
- Client can add their documents
- Can be updated separately

The screenshot shows the GenAI™ Coach interface. At the top left is a purple star icon and the text "GenAI™ Coach". Below this is a white input area containing the text: "Optimize my claims process to minimize processing time while maintaining NPS." Below the input area is a purple button with a white hammer icon and the text "Executing... Knowledge Buddy". At the bottom of the interface is a white input field and a blue "Send" button.



# Knowledge Buddy Tool

```
callKnowledgeBuddy(
```

```
    buddyName = "ProcessBestPracticesBuddy",  
    question = "In my Claims process, 75% of  
cases are routed to an incorrect agent.  
Can you suggest ways to improve this  
bottleneck?"
```

```
)
```

Output:

To improve routing to the correct agent, you could use Pega Process AI which will use self-learning adaptive machine learning models & decisioning to optimally route to the right agent.

To introduce Process AI in a process, you should ...



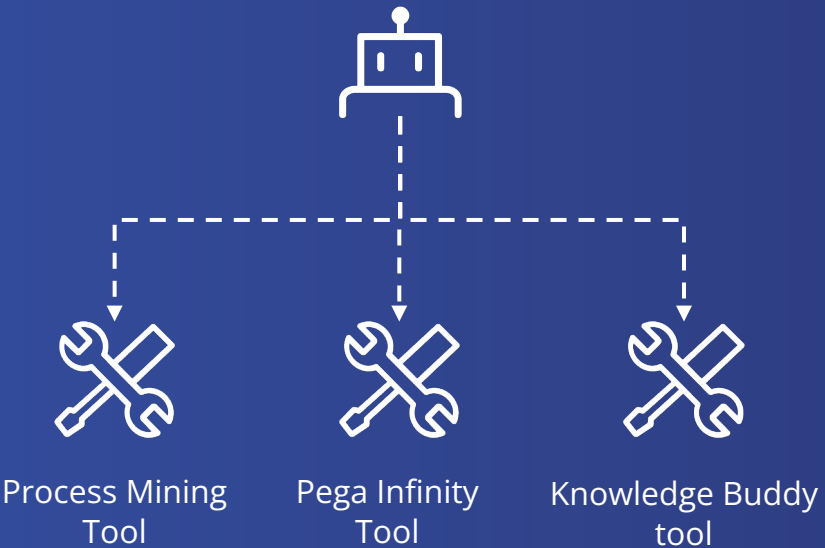
GenAI™ Coach

Optimize my claims process to minimize processing time while maintaining NPS.



Executing...  
Knowledge Buddy

Send



## GenAI™ Coach

Optimize my claims process to minimize processing time while maintaining NPS.

...

Send





# Pega Infinity Tool

```
CallPegaInfinityTool(
```

```
    question = "Introduce Process AI in the  
    agent routing step of my Claims process"
```

```
)
```

- Dozens of 'automations' in Infinity which can access data or execute activities
- Creates a 'plan' on the fly to achieve the desired goal
- Refuses if goal cannot be achieved
- Asks for approval before execution

**GenAI™ Coach**

Optimize my claims process to minimize processing time while maintaining NPS.

Executing...  
Pega Infinity

Send



# Pega Infinity Tool

```
CallPegaInfinityTool(
```

```
    question = "Introduce Process AI in the  
    agent routing step of my Claims process"
```

```
)
```

Output:

Plan to automate:

1. Get <Decision ID> of agent routing step
2. Create <Adaptive Model>
3. Introduce <Adaptive Model> in <Decision ID>

Automations used: ["getDecisionID",  
"createAdaptiveModel", "introduceModel"]



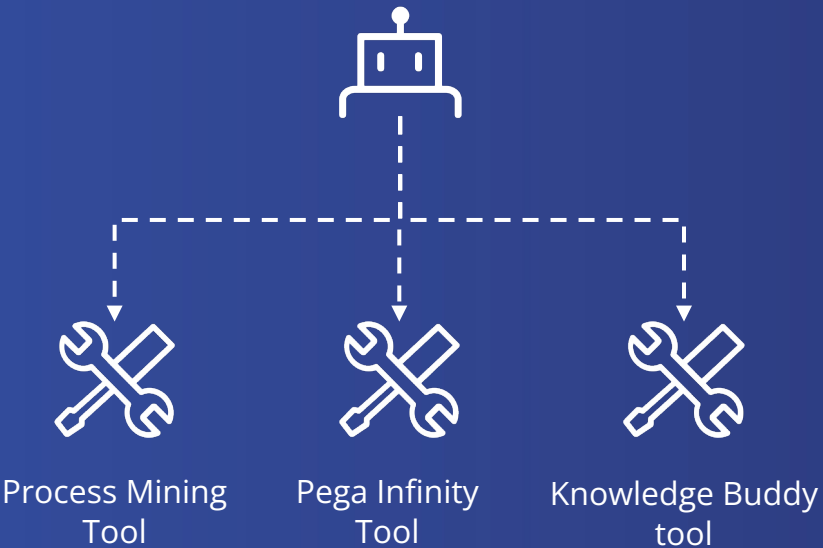
Pega Coach

Optimize my claims process to minimize processing time while maintaining NPS.



Executing...  
Pega Infinity

Send



## Pega Coach

Optimize my claims process to minimize processing time while maintaining NPS.



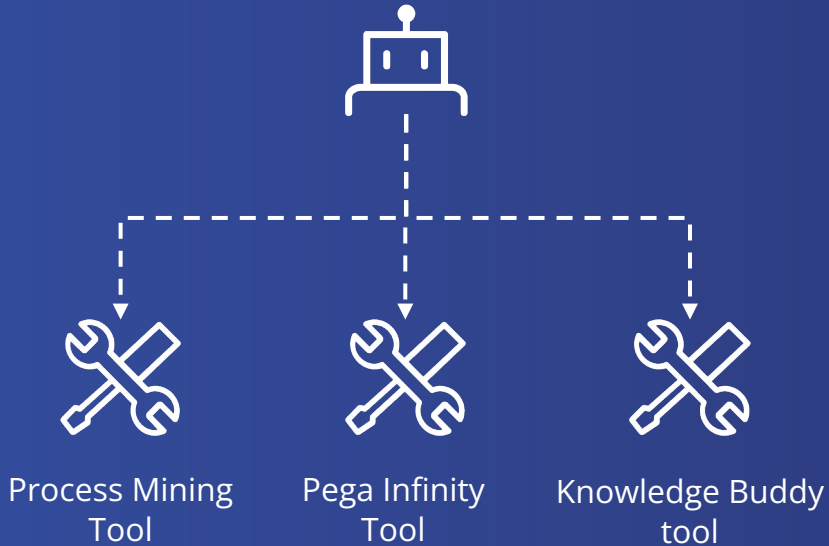
I found three bottlenecks in your process. Would you like me to enable Process AI to improve the processing time?

Yes please do!

Yes please do!

Send





## Pega Coach

Optimize my claims process to minimize processing time while maintaining NPS.

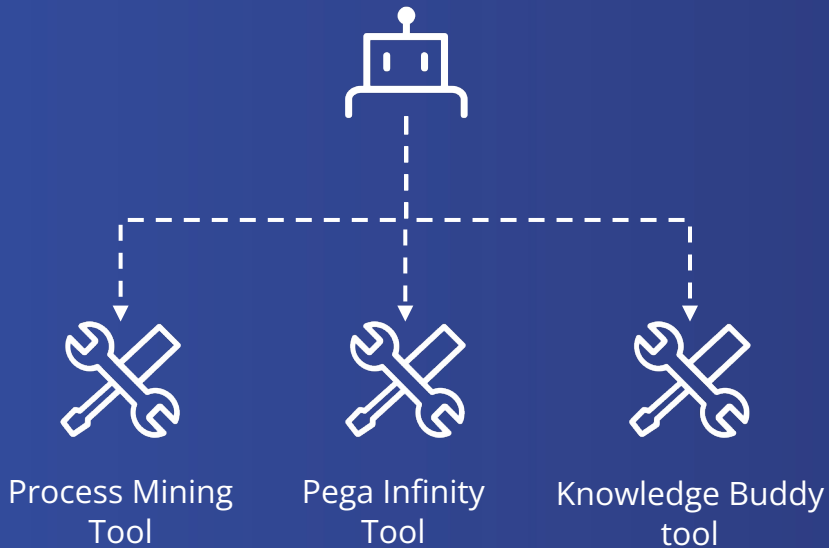


I found three bottlenecks in your process. Would you like me to enable Process AI to improve the processing time?

Yes please do!

...

Send



## Pega Coach

Optimize my claims process to minimize processing time while maintaining NPS.



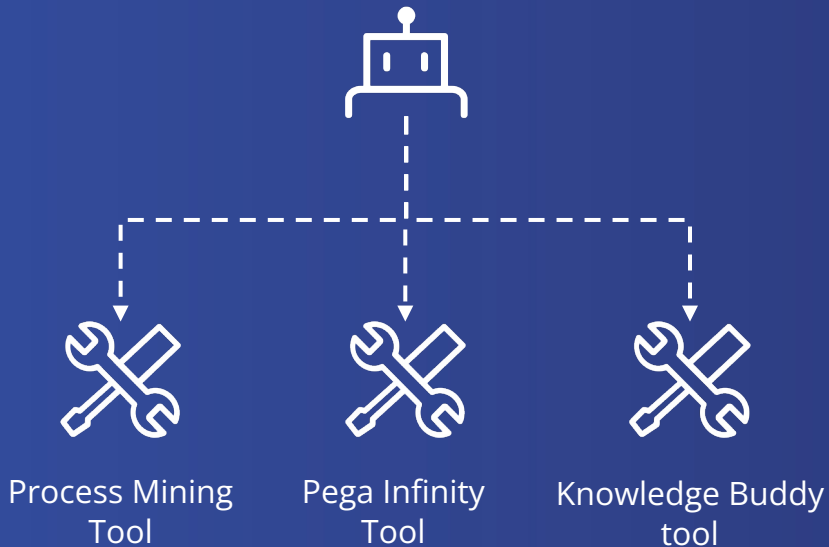
I found three bottlenecks in your process. Would you like me to enable Process AI to improve the processing time?

Yes please do!



Executing...  
Pega Infinity

Send



## Pega Coach



I found three bottlenecks in your process. Would you like me to enable Process AI to improve the processing time?

Yes please do!



Done! My changes are part of change request CR-4264, please ask your system administrator to promote them to production.

Send

**Now we would like  
to unleash the power  
of autonomous  
agents to you!**





# Coach Rule in '24.2

**DEV STUDIO** Application: Sales Automation Configure Launch Portal Create Search DEVELOPMENT

Home GenAI Coach New

**Edit GenAI Coach: Pegaworld Infinity Coach [Final, Internal]** Actions Save

CL: Rule-GenAI-Coach ID: pzPegaworldInfinityCoach RS: Pega-GenAI:08-24-03

Definition Knowledge Actions History

System Instructions ⓘ

You are an expert Pega Coach that can help me understand my processes through three available tools:

- Pega Process Mining Tool is an interactive agent that can answer questions about all processes in the Infinity system. It can find bottlenecks, find root causes & do real-time analysis.
- Pega Knowledge Buddy Tool is a RAG agent that can answer questions based on real-time data sources around Pega product documentation and other related documents.
- Pega Infinity Tool is an interactive agent that can plan and take actions in Pega Infinity.

Welcome Message

Hi, I am your Pega Coach, ready to help you understand your processes.

Ask me any questions about Pega product documentation, about your process instances - I can even perform analysis over your processes.

**Guided Questions**  
Pre-configure questions to enhance user interactions with the Coach.

**Starter Question**  
Configure an initial question to be asked as soon as the Coach is launched.

Question: What are the main bottlenecks in my system?

Additional Instructions: To answer this question, you should first ask the process mining tool to do an analysis of the process. Then you should reach out to the knowledge buddy tool to get more information about this kind of bottleneck and lastly you should ask the Infinity tool to see if you can take action yourself.

**Quick-Select Question**  
Configure the questions or instructions that users can choose at any time during their interaction with the Coach.

Question: What are the main bottlenecks in my system?

Additional Instructions: To answer this question, you should first ask the process mining tool to do an analysis of the process. Then you should reach out to the knowledge buddy tool to get more information about this kind of bottleneck and lastly you should ask the Infinity tool to see if you can take action







SA Agile Workbench Current work Scenario testing Issues Tracer Clipboard Live UI Live Data Accessibility Performance PEGA

# Coach Rule in '24.2

**Edit GenAI Coach: Pegaworld Infinity Coach [Final, Internal]**

**CL:** Rule-GenAI-Coach **ID:** pzPegaworldInfinityCoach **RS:** Pega-GenAI:08-24-03

Definition Knowledge Actions History

Source	Category
 KnowledgeBuddyTool 	
 ProcessMiningBuddyTool 	




[+ Add](#)

# Coach Rule in '24.2

**Edit GenAI Coach: Pegaworld Infinity Coach [Final, Internal]**

**CL:** Rule-GenAI-Coach **ID:** pzPegaworldInfinityCoach **RS:** Pega-GenAI:08-24-03

Definition Knowledge Actions History

Name	Category
 PegainfinityTool 	

[+ Add](#)

Value

- Pega '23 boosters
- Coach '24
- Blueprint '24



### Engineered Prompt

Automated prompt engineering, dynamic prompt templates, single call

- Knowledge Buddy '24



### Basic Tool Use

1-2 tools (f.i. search engine + knowledge), single call, scripted use

- CDH Assistant '24



### Autonomous Agent

Agentic, planning, multi-step, multi tool, multi agent

Autonomy

Value

Pega '23 boosters



### Engineered Prompt

Automated prompt engineering, dynamic prompt templates, single call

Blueprint (future) Knowledge Buddy '24



### Basic Tool Use

1-2 tools (f.i. search engine + knowledge), single call, scripted use

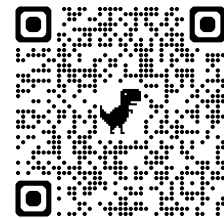
Coach (future) CDH Assistant '24



### Autonomous Agent

Agentic, planning, multi-step, multi tool, multi agent

Autonomy



OPINION

# 2024: The year Generative AI is let out of its cage

**Instead of worrying about AI taking over their jobs perhaps people should wonder if their work will be taken over by a new starter who knows how to use AI.**

By Peter van der Putten, Director AI Lab, Pegasystems

Published: 17 Jan 2024



# The AI Manifesto

Nine guiding principles for responsible & trustworthy application of AI with impact



AUTHORED BY: PETER VAN DER PUTTEN





# Make the Most of Your PegaWorld Experience



Don't miss these sessions!

**Top 10 Hottest Use Cases for Boosting Your ROI with GenAI from Pega**

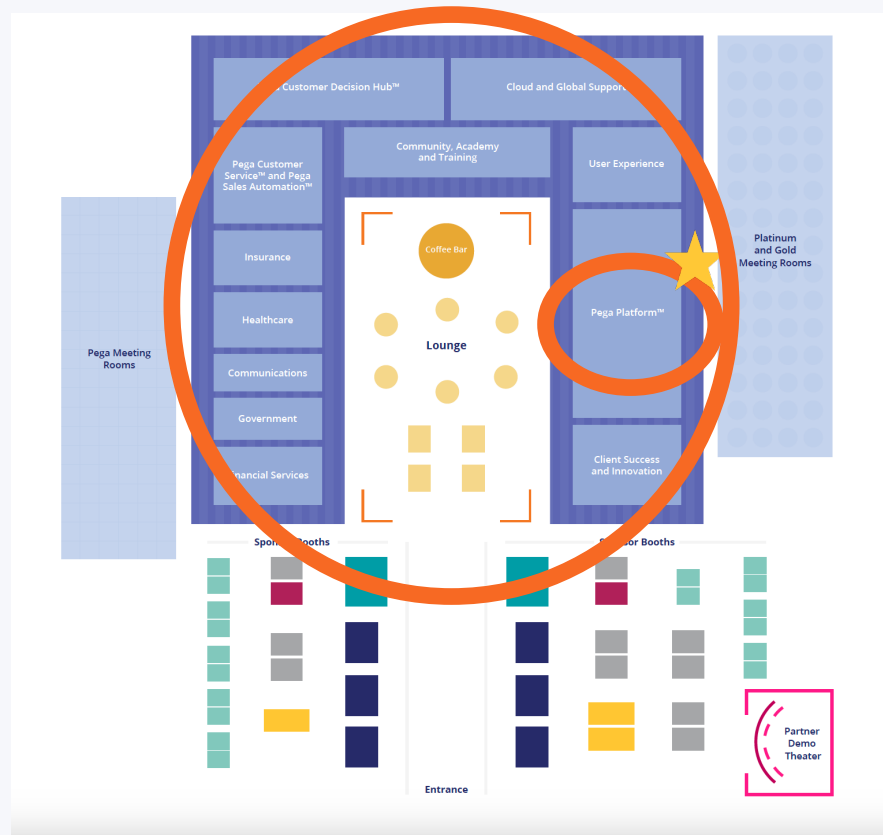
*Mon | 11:30-12:00 | Chairman's Ballroom 359*

**Sparks of AI-driven Autonomous Operations**

*Mon | 2:15-3:00 | Premier Ballroom 316*

**Pega GenAI Cloud Exclusive: Paving the way for an Autonomous Enterprise**

*Tues | 3:15-4:00 | Chairman's Ballroom 361*



# Question time



**Peter van der Putten**

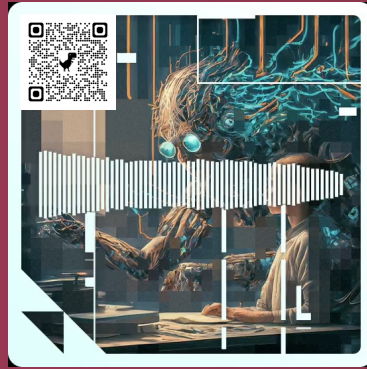
Director AI Lab & Lead Scientist, Pegasystems



**Stijn Kas**

Sr Data Scientist, Pegasystems





You bring the screen to life with code (with code)  
Teach me the ways of this AI road  
Do we trust the tech, or do we pause? (do we pause)  
Information at our fingertips, let's explore (let's explore)  
With every lesson, we're learning more  
Guide me through, like a mentor (a mentor)  
Questions broad, and answers deep (answers deep)  
The machine thinks, while the world's asleep  
But wisdom's not in silicon alone (not alone)  
You articulate the doubts we've never known (never known)

We're discovering a digital dawn  
With every query,  
our minds are reborn (reborn)





# PegaWorldiNspire

JUNE 9-11, 2024 | LAS VEGAS

