GenAI in application design: does it make a difference?

A field experiment using Pega's Blueprint

Atos Amplify

Executive Summary

Generative Artificial Intelligence (GenAI) is commonly seen as one of the most promising applications of AI. Global organizations and governments are quick to hop onto the AI bandwagon in an attempt to maximize this growing trend, but little is known about how it can benefit professionals in their daily operations.

Pegasystems (also known as Pega) is a leading vendor of a low-code platform aimed at dynamic case management. The company created Blueprint as a tool to make a first design of a new case management application — an activity that is usually carried out in the early stages of a project by a design team consisting of business experts, business architects and system architects.

In our experiment, we aimed to answer the following questions

1. Does Pega Blueprint replace the design team?
2. Can it simply play a supportive role in assisting the design team?
3. Does it add value in any way? If so, what is the most effective use?

In the spring of 2025, Atos' team of AI experts conducted an experiment that involved human experts as well as Pegasystems' GenAI tool, Blueprint. Based on a four-page (fictious) case description of the European Food Quality Testing Facility (EFQTF), five teams were tasked with creating a new high-level application to fulfil the requirements that were outlined in the case description.

This design consists of a description of several types of elements: applicable case types, the processes underlying these case types, the interfaces to other systems, the data model as well as the personas that play a role in the processes.

One team consisted of just the Blueprint application, where the only human task was to upload the case description in Blueprint and upload the application created. Two teams consisted of the traditional human set-up with a business architect and a system architect. And the last two teams had a traditional set-up but also had Blueprint at their disposal.

This paper showcases the experiment hypotheses and setup, the team findings and observations, and our learnings and final conclusion. Dive in to read the entire paper and chart our journey and findings.

>> Learn more about Atos' AI offerings and how we are making a difference in the real world.

>> Connect with our team of experts to optimize GenAI in your business today.

Aiming. Assessing. Hypothesizing.

The experiment aimed to deliver five Pega applications as a response to the business case description.

- **GENAI:** One GenAI-only application
- **HUMAN1 and HUMAN2:** Two applications developed by human experts only
- **MIXED1 and MIXED2:** Two applications developed by human experts with GenAI at their disposal

Team name	Characteristics
GENAI	Only Pega Blueprint with a case description fed into Pega Blueprint
HUMAN1	Team consists of a business architect and a system architect that do **not** have Pega Blueprint at their disposal
HUMAN2	See HUMAN1
MIXED1	Team consists of a business architect and a system architect that **do** have Pega Blueprint at their disposal.
MIXED2	See MIXED1

Table 1: Team composition

GenAI's massive **data processing capabilities** are the USP of this technology. Hence, it was expected that the design generated by Blueprint will contain elements that cannot be found in the applications that are built by the human-only teams. On the other hand, humans – given their non-artificial intelligence – would provide certain elements that Blueprint could not come up with.

Hence the hypotheses were as follows:

Hypotheses H1a, H1b, H1c

- *GENAI will contain elements that cannot be found in HUMAN1 and HUMAN2.*
- *HUMAN1 and HUMAN2 will contain elements that cannot be found in GENAI.*
- *MIXED1 and MIXED2 will contain elements that cannot be found in GENAI, HUMAN1 and HUMAN2.*

When it comes to **speed**, it was expected that the speed of Blueprint could not be matched by humans. Although we did not instruct the human participants in the experiment — except for the deadline — we expected that they would use the allotted time till the end.

In other words:

Hypothesis H2

GENAI will be delivered within 10 minutes of the final delivery of answers to questions posed to the business. HUMAN1, HUMAN2, MIXED1 and MIXED2 will be delivered just before the closing deadline of the experiment.

At the end of it all, **quality of the design** matters. For this purpose, an expert assessment of all the designs delivered would be carried out.

Every design will be scored against three axes:

1. **What is the depth of the design?** Is the level of detail sufficient to be used as a starting point for DevOps teams to start developing?
2. **What is the consistency of the design?** Do all the elements in the design properly interact with each other? Are there no duplicates or even contractionary elements in the design?
3. **What is the coverage of the design?** Does it cover the full business scope of the case provided?

Scoring would be done using a blind assessment, investigating the Pega application delivered. All three axes would be scored on a five-point **Likert scale** (1=low, 5=high). We expected that GENAI would produce many elements that strongly contribute to the coverage and depth, but it would have difficulties in preventing inconsistencies. We expected that the teams where GENAI and humans work together would be able to avoid these inconsistencies.

Hypothesis H3a and H3b

The GENAI application will have the best Coverage and the best Depth, whilst it will score low on Consistency.

The mixed teams will have a better Coverage and Depth than the human-only teams, whilst their consistency will be better that the GENAI "team".

Establishing the experimental set-up

The experiment was carried out using 9 participants: 4 business architects, 4 system architects and a Blueprint operator.

A case description "EFQTF, the European Food Quality Testing Facility" was written by a business owner.

None of the participants had seen this case description in advance.

In addition, the case dealt with a business area that none of the participants had any prior experience in. The case description was four pages long and described in business language the requirements of the new application.

The assignment was stated as follows:

"The European Commission has selected your company to establish a core support system to operate the EFQTF facility. Your company has proposed to develop the system using the Pega platform. You are being asked to come up with a first design of the system before 18:00 pm today. This design will serve as a starting point that the 4 DevOps Pega teams can use to start building the application. It is the EU's wish to have the EFQTF operational on January 1st, 2026."


The European Food Quality Testing Facility Case Description

EFQTF

The European Food Quality Testing Facility

Case Description

Introduction

Citizens of the European Union have recently been alarmed by several news items on the amount of PFAS in eggs produced by chickens that are kept as a hobby.

The abbreviation PFAS stands for per- and polyfluoroalkyl substances. These are man-made substances that do not naturally occur in the environment. PFAS are widely used, long lasting chemicals, components of which break down very slowly over time. Because of their widespread use and their persistence in the environment, many PFAS are found in the blood of people and animals all over the world.

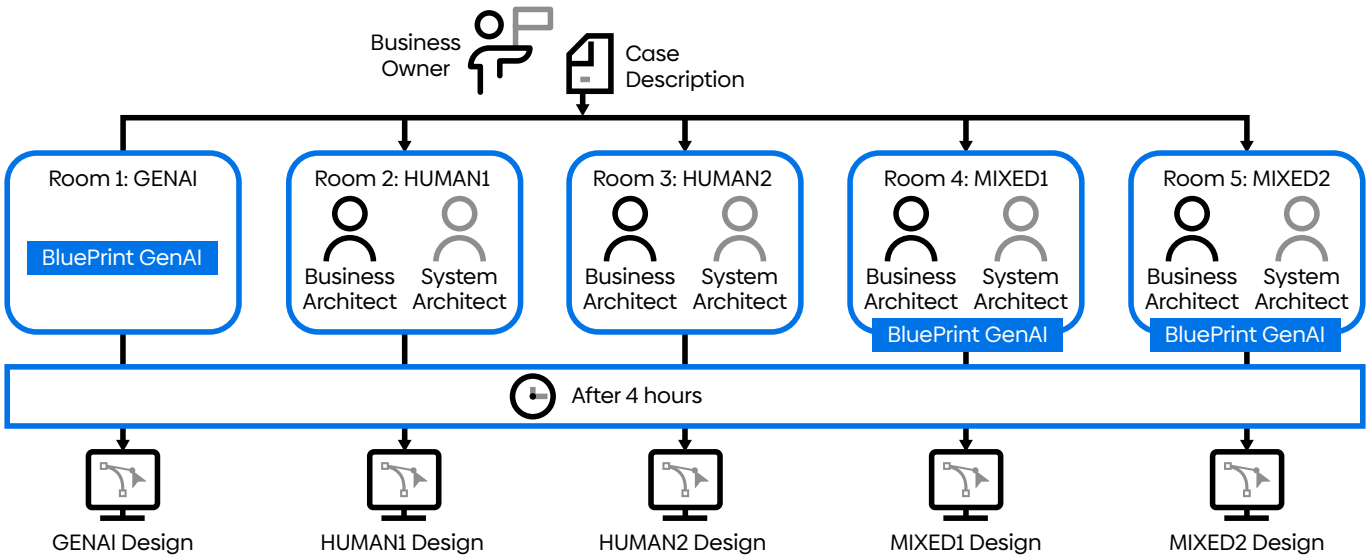
PFAS are found in water, air, fish, and soil at locations across Europe. Scientific studies have shown that exposure to some PFAS in the environment may be linked to harmful health effects in humans and animals. There are thousands of PFAS chemicals, and they are found in many different consumer, commercial, and industrial products. This makes it challenging to study and assess the potential human health and environmental risks.

European consumers are worried about PFAS concentrations in the products they eat and drink. The European Commission has decided to set up a large scale laboratory where every European citizen or organization can submit food samples to that are then tested on PFAS concentration. This facility will be established in the Netherlands. As it is the intention to expand the tests offered to other chemicals than PFAS, the facility has been given the more generic name EFQTF, the European Food Quality Testing Facility.

All participants had been trained in Blueprint as part of their continuous professional development in Pega, integrated in the educational framework of Atos. The participants had been randomly assigned to the teams. None of the participants had access to the public internet, the only sites that could be accessed were the Pega application site and – for GENAI, MIXED1 and MIXED2 – the Pega Blueprint site.

Each of the five teams was assigned a room for the course of the experiment, which was held on a Wednesday in May 2025. The teams were not allowed to communicate with each other during the four-hour long experiment. During the first two hours, the participants were allowed to ask questions in writing to the business expert, who then distributed the questions and his answers to all teams.

After four hours, all teams were required to hand over their design in the form of a first Pega application. The five applications were then assessed by a team consisting of the business owner and a principal Pega architect.



All participants were asked to take notes during the process and filled out a questionnaire after submission of their design work.

Tabulating outcomes and analysis

Team GENAI submitted their application design 10 minutes after the answer to the last question was provided. The other four teams delivered their design work at the end of the allotted time.

The table below provides a high-level overview of the results:

Team name	Result
GENAI	Initial Pega application
HUMAN1	Initial Pega application
HUMAN2	Paper-based design of a first Pega-application
MIXED1	Initial Pega application
MIXED2	Initial Pega application

Table 2: High-level overview of the results

As illustrated by the table above, one of the human-only teams did not have the time to convert their paper-based design into a Pega application. The other four teams delivered a Pega application.

The elements in the design

The business owner and the principal Pega architect have together compared the five designs that have been delivered. They first assessed the elements that could be found in the designs: case types, processes, data objects, integration points and personas.

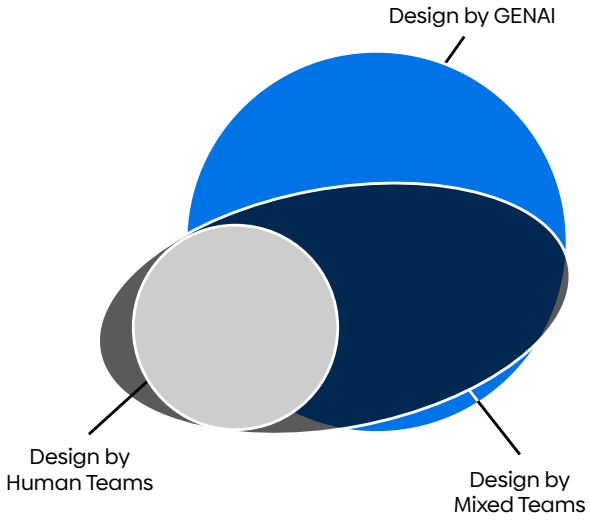
Team name	#Case types	#Processes	#Data Objects	#Integration Points	#Personas	Total
GENAI	10	59	6	0	10	85
HUMAN1	6	5	3	4	4	22
HUMAN2	1	6	2	0	4	13
MIXED1	5	28	5	0	5	43
MIXED2	6	34	9	0	5	54

Table 3: Breakdown of the results per team

Looking at the numbers alone, the GENAI application contains the most elements. The HUMAN1 and HUMAN2 applications contain the least. An in-dept analysis of the contents of the elements leads to the following:

- All five applications have **elements in common**, in something that may be called the kernel of the application. This kernel focuses on the main process of the case description: ordering and fulfilling Food Quality tests. The largest part of the HUMAN1 and HUMAN2 application design deals with this process.
- The GENAI application design contains many **duplicate elements**. The testing process of samples is defined within three different case types, for example. The MIXED1 and MIXED2 teams stated that they removed these duplicates manually, allowing them to end up with an application design that not only consists of the kernel but also defines a set of additional processes: the handling of failed tests and the maintenance of equipment.

This is reflected in the picture below:



If we refer to Hypothesis H1a, H1b and H1c, we have some supportive evidence.

The GENAI model contains elements that cannot be found in the human-based designs, especially in the area of secondary processes. In this experiment, the tendency of the two human-only teams was to focus on the core process, and leave aside the less important processes, even though these were explicitly mentioned in the case description.

In the discussion afterwards it was agreed that design teams often probably focus too much on the most frequent process — the "happy flow" as it is often called. The GENAI tool helps the teams here to also give focus to processes that are less "happy" but that are nonetheless important for customers, for example a good working luggage handling process is important for every airliner, but the design team should also spend much effort in the process of properly handling luggage that is lost somewhere.

Therefore, GENAI produced design elements that the human participants did not produce within the timeframe given. The reverse is also true: the design of the human teams contained several elements that were not produced by GENAI. Here is where the **human touch** appears, based on creativity, experience and interpretation of the case description in a different way. One could argue that these elements make the difference, as the rest is just a commodity. One human team clearly made a distinction between the delivery phases of the product or, in Pega terms, the minimal lovable product (MLP). The other team had detailed out the various channels that could be used to communicate with the customer. Both design elements were mentioned in the case description, but a bit implicit, and were overlooked by GENAI.

On the time spent

Although it was not stated that the participants had to use the full timeslot allotted, they all did. The four teams that had human participants in it all worked on the design to the very last minute.

There was a remarkable difference between the teams though.

Teams HUMAN1 and HUMAN2 – the teams without Blueprint – had the feeling that they had not finished the assignment yet, and they would need several extra hours to properly finish their work. The teams MIXED1 and MIXED2 – which could use Blueprint as a support tool – had completed the assignment by the end of the time allotted to them.

We observed a remarkable difference in the way of working. Teams HUMAN1 and HUMAN2 spent the first 75% of their time designing a brown paper process map, understanding the requirements, detailing the process, and translating it into the design elements requested. As stated before, they were only able to do this for the main purpose of the case. Team HUMAN2 ran out of time. In fact, they could not convert their work into an initial Pega application. The overall feedback of these two teams was that "it is a lot of work; we need more time".

The teams MIXED1 and MIXED2 took a different approach. They both decided to use Blueprint as a supportive tool and – after an initial analysis of the case – started to feed the GENAI tool with prompts based on the case. They checked the outcome, decided what was good and what wasn't, and then recreated the prompt and fed it into Blueprint again. This process was iterated a number of times, until some stability was reached in the outcome of Blueprint, after which this outcome was altered manually.

As both teams stated, "when the time was up, we felt that we had finished."

On the quality of the design

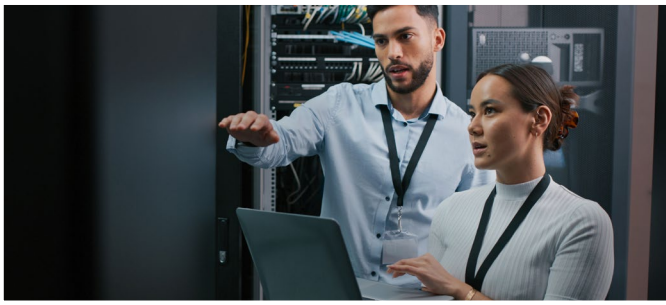
The four designs that have led to an initial Pega application have been rated by a principal Pega architect and the business owner.

To allow for a blind assessment, the only design delivered on paper by team HUMAN2 was not included in the overall assessment.

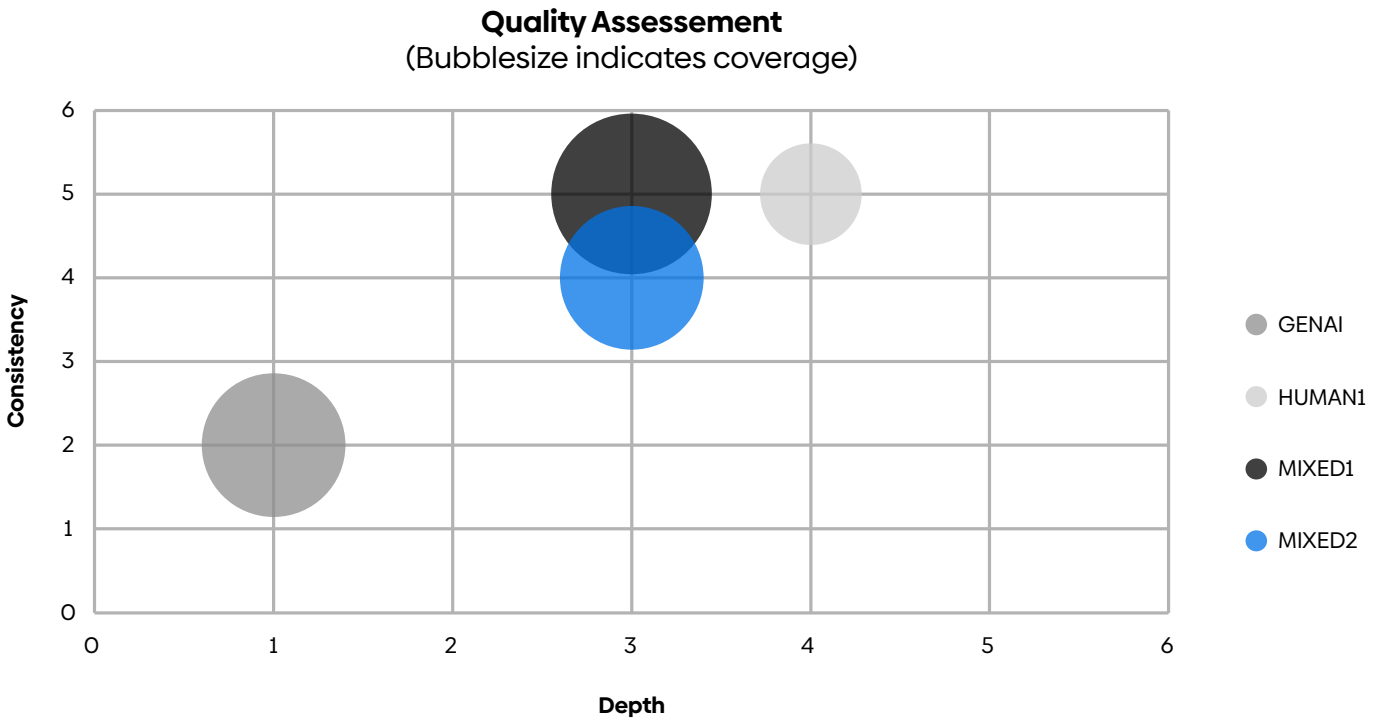
The scores for each, based on the three axes described above, are provided in the table below:

Team name	Coverage	Consistency	Depth
GENAI	10	59	6
HUMAN1	6	5	3
MIXED1	5	28	5
MIXED2	6	34	9

Table 4: Breakdown of the assessment scores per team



A graphical representation of this outcome is depicted below. The size of the bubbles is an indication of the coverage.



During the assessment, the assessors made several observations. One team had created a very detailed design for a part of the business case. That detailed design was very consistent.

In the opinion of the assessors, there was too much detail in the design. In many aspects the design already went from the "what to do" to the "how to do it" stage, typically an action that is with the DevOps teams. The coverage of this design was relatively low. It turned out that this was the HUMAN1 team, and it was indicated by the team that they had truly focused on what they (rightfully) believe to be the most important part of the application.

One team had a broad coverage, but on a shallow level. This was insufficient for a DevOps team to start working on it, according to the assessors. The internal consistency was also low as there were many duplicate elements. This turned out to be the work created by the GENAI team.

The other two teams both had high coverage, sufficient depth and a good consistency mark. These turned out to be two MIXED teams. In the judgement of the assessors, one of these two teams created the **best design**, with a **coverage that was broader than any of the other teams, a depth that was good enough** and a **good consistency** in the design.

Analysis, review and discussion

This experiment does not have a pretention to directly lead to results that are statistically significant, amongst others due to the small sample size. However, if we carefully review our hypotheses, then we can make the following observations:

Hypothesis	Expectations	Observations
H1a	GENAI will contain elements that cannot be found in HUMAN1 and HUMAN2.	We identified several elements that were found by GENAI, but (within the given time conditions) not by the human-only teams.
H1b	HUMAN1 and HUMAN2 will contain elements that cannot be found in GENAI.	We found several elements that were found by the human-only teams and not by the GENAI team.
H1c	MIXED1 and MIXED2 will contain elements that cannot be found in GENAI, HUMAN1 and HUMAN2.	The Mixed teams had the highest number of elements.
H2	GENAI will be delivered within 10 minutes of the final delivery of answers to questions posed to the business. HUMAN1, HUMAN2, MIXED1 and MIXED2 will be delivered just before the closing deadline of the experiment.	This is exactly what happened.
H3a	The GENAI application will have the best coverage and the best depth, whilst it will score low on consistency.	The experiment supports the hypothesis that GENAI provides a broad coverage, but we did not find support for the level of depth. The consistency level of the GENAI design was relatively low.
H3b	The mixed teams will have better coverage and depth than the human-only teams, while the consistency will be better than the GENAI team.	The experiment provides support for this hypothesis.

The initial high-level design of a dynamic case management application is a very important part of the implementation process. It has a large impact on all the steps taken thereafter. It is therefore usually placed in the hands of an experienced team of experts.

After design submission and while the assessment was carried out, the experts filled out a questionnaire on how they would evaluate the way of working. There was a remarkable difference between the two human-only teams and the two mixed teams:

The human-only teams had felt quite a time pressure, and they felt that the time was up before they were finished, insufficient time was allotted. The mixed teams had the feeling that they were just about done when their time was up.

The human-only teams had spent about 75% of the allocated time in understanding the case, drawing elements such as processes on the wall, and discussing various elements.

The mixed teams had done the same for about 25% of their time and then started to work with the Blueprint GenAI tool.

Team MIXED1 indicated that they started to work on improving the Blueprint prompt in an iterative cycle: “Enter-the-Prompt, Generate the Design, Review the Results, Improve-the-prompt etcetera”. They did so until they saw no improvements in the design generated by Blueprint, and at that point started to improve the design manually, for example by establishing internal consistency.

Team MIXED2 first created a design without Blueprint and then ran Blueprint to produce a design that was used to validate their initial work. They were unhappy with this result and switched to using Blueprint as a generative tool.

Overall, the mixed teams were positive about using Blueprint:

- “It really has an added value.”
- “We can work really faster.”
- “Blueprint has pointed out business areas that we would not have given much attention too.”
- Or, as one of the participants in one of the human-only teams stated it: “Can you place me in a mixed team next time, please?”

The overall opinion was that all architects will use Blueprint as a standard tool in next design activities.

The way forward: Collaboration and optimization

This experiment helped supportive evidence that – both from an objective and from a subjective perspective – GenAI adds significant value in the initial design of dynamic case management systems.

In our experiment, the designs delivered by the teams that were able to use GenAI are better than the designs delivered by the human-only teams. The coverage of the design is better, the time pressure is lower, and the general feeling is that it is a tool that every architect wants to apply in future ventures. It helps in identifying blind spots and even the iterations that were done to create a better prompt for Blueprint have helped the human architects to create better structure and express the initial design.

It is also demonstrated the GenAI cannot do the work by itself. The winning team is a combination of GenAI and human competencies.

With technology determining the future of the world as we know it, it only stands to reason that we need to balance human skillsets and reasoning with technological advancements to optimize opportunities and collaborate for a sustainable tomorrow.

Acknowledgements

We would like to thank our colleagues who volunteered to participate in this experiment and have fueled us with ideas and useful comments:

- Alexander Berkhout
- Remi Bos
- Niek Draaijer
- Wessel Ferwerda
- Alex Hofstra
- Naveen Maddula
- Bob van der Wende
- Martin van Zoelen
- Tijmen Witte

Your assistance, valuable time and input were much appreciated.

Authors

Colette van Boxtel
Lead Strategy Consultant
Colette.vanboxtel@atos.net

Colette van Boxtel is a Lead Strategy Consultant at Atos Amplify, responsible for business development and driving strategic growth initiatives. Colette focuses on large and complex organizations, including government and defense, where alignment between strategy, culture, and technology is critical.

Colette has led large-scale Pega implementations across multiple accounts, bringing together advisory and delivery to ensure successful outcomes. With a strong advisory background, Colette supports organizations in translating strategy into effective and sustainable execution.

Robert-Jan Streng
Partner, Business Process Improvement,
PEGA Lead System Architect
robert-jan.streng@atos.net

Robert-Jan Streng is a Partner at Atos Amplify, specializing in bridging the gap between business and IT. Robert-Jan holds a PhD in Business Process Management from Delft University of Technology and is a Lean Six Sigma Black Belt. As a certified Pega Lead System Architect (LSA), Robert-Jan brings years of experience in guiding organizations to optimize the use of modern technologies, including AI.

Thijs van Werven
Senior Consultant Operational Excellence,
PEGA Business Architect
Thijs.vanwerven@atos.net

Thijs van Werven is a Senior Consultant at Atos Amplify, driven by improving how people work with technology. As a certified Pega Business Architect (BA), Thijs brings years of experience in business and process analysis, supporting organizations in transforming processes through automation and AI. Thijs has a strong focus on user adoption and sustainable change and believes that technology should always be designed around people.

Rik van der Zwan
Senior Consultant, Operational Excellence
PEGA Business Architect
Rik.vanderzwan@atos.net

Rik van der Zwan is a Senior Consultant at Atos Amplify, focused on delivering business value through technology. As a certified Pega Business Architect (BA), Rik brings years of experience in aligning business and IT. With a background in innovation management, business analysis, and Lean thinking, Rik supports organizations in driving effective and lasting change. and Lean thinking to drive organizational change.

About Atos Group

Atos Group is a global leader in digital transformation with c. 56,000 employees and annual revenue of c. €7.2 billion (at the go-forward perimeter), operating in 54 countries under two brands - Atos for services and Eviden for products and systems. European number one in cybersecurity and a leader in cloud, Atos Group is committed to a secure and decarbonized future and provides tailored AI-powered, end-to-end solutions for all industries. Atos Group is the brand under which Atos SE (Societas Europaea) operates. Atos SE listed on Euronext Paris.

Find out more about us
atos.net
atos.net/career

Let's start a discussion together

107548 - EM + MB - Whitepaper GenAI - Amplify

Atos Amplify