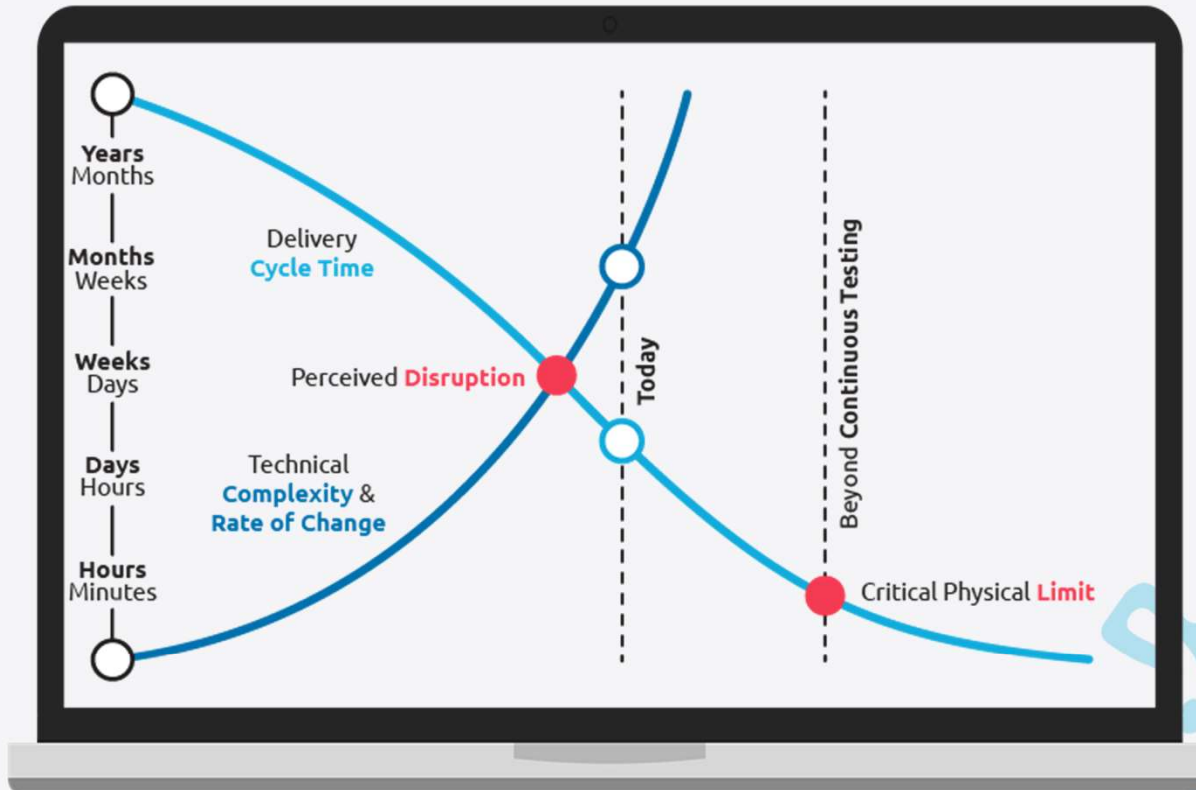


Generative AI applied to Quality Engineering and Testing

By Antoine Aymer, Global CTO

Oct 2024





1

AI-powered platforms complement and extend beyond human-designed testing

2

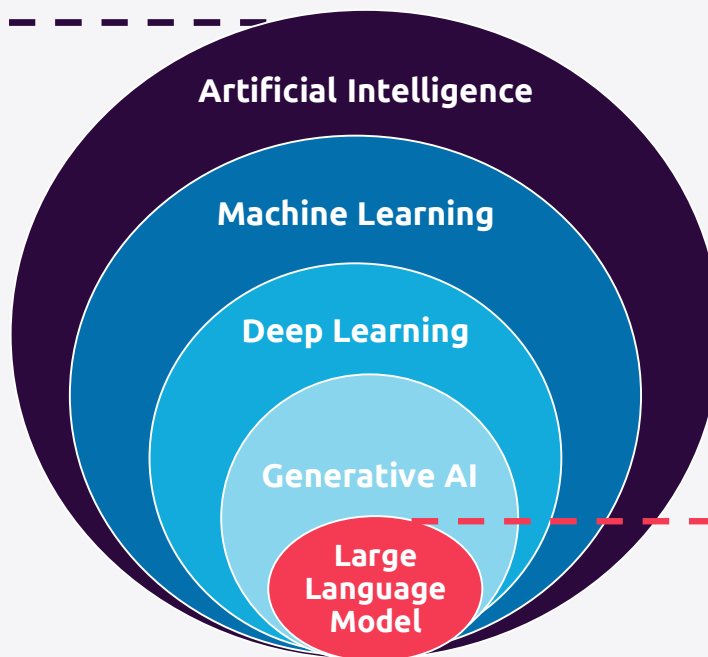
Generative AI amplifies human-driven quality engineering

1

Technical and analytical tasks

- Complex object detection
- Visual recognition
- self-healing automation tests
- Root cause analysis
- user flow validation
- Predictive metrics
- Synthetic data generation
- ...

<https://www.sogeti.com/ai-for-qe/>



2

Natural language-driven tasks

70% of software lifecycle activities unrelated to code

Last year, at the CxO summit, in Cannes

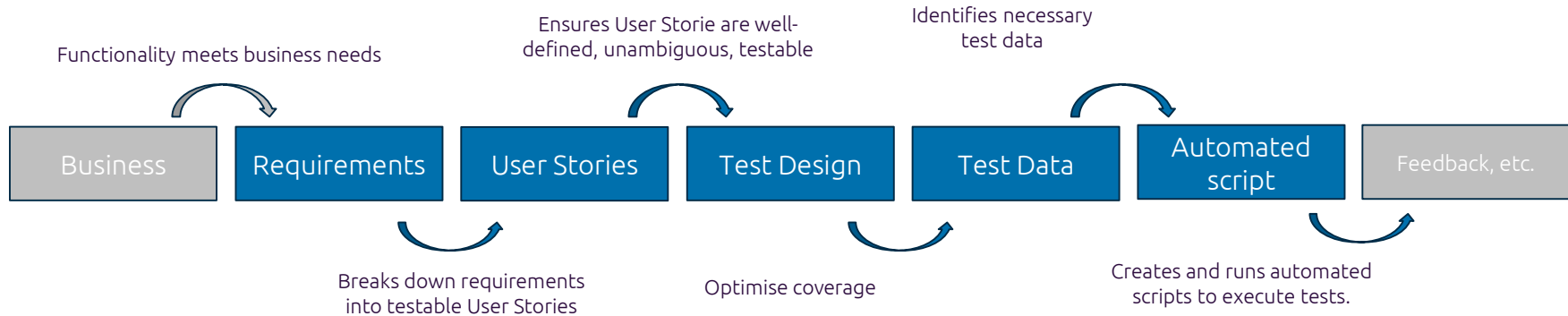



Importance in QE&T

The importance is based on a combination of factors like time spent on the activity, criticality of the activity, and the potential impact of the use case on overall testing outcomes.



What if testing could be completed before writing a single line of code?

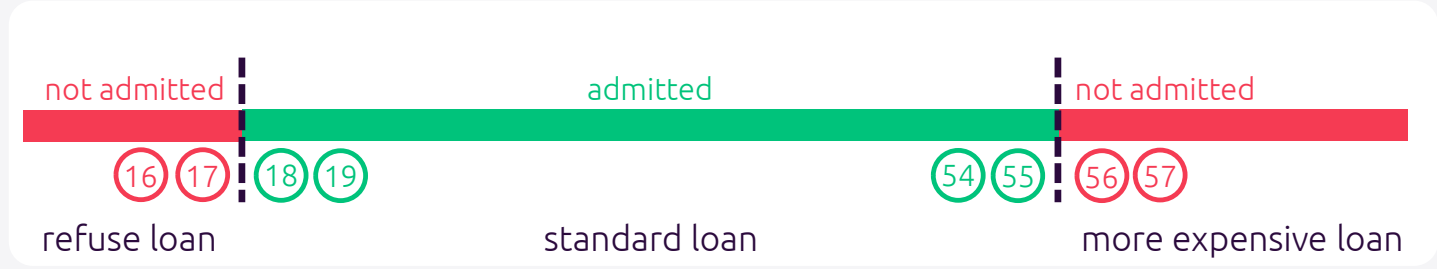
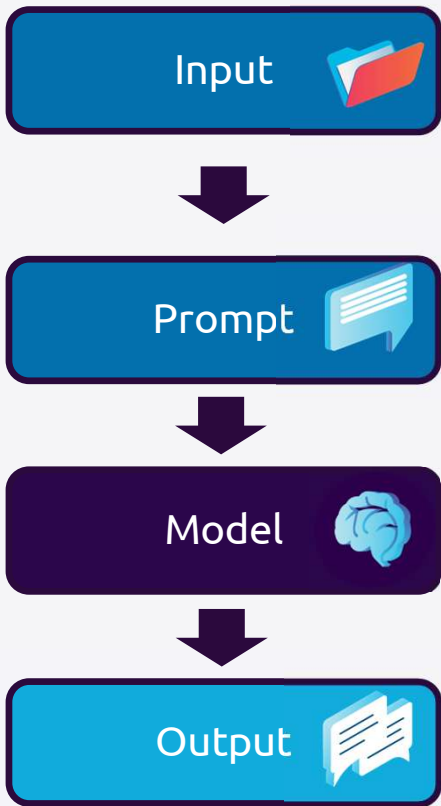




Section 2

How to generate
value?

Example of “out-of-the-box” test case generation



“Generate test cases applying **4-value** Boundary Value Analysis.”

Model 1

Here is a table applying the 4-value boundary value analysis for the scenario depicted in the image:

| Test Case # | Input Value | Expected Outcome |
|-------------|-------------|------------------------------------|
| 1 | 17 | Not Admitted (Refuse Loan) |
| 2 | 18 | Standard Loan Admitted |
| 3 | 55 | Standard Loan Admitted |
| 4 | 56 | Not Admitted (More Expensive Loan) |

These test cases target the boundaries between the refusal, standard, and more expensive loan categories.

2-value

Model 2

| Test Case ID | Credit Score | Expected Result |
|--------------|--------------|---------------------------|
| TC01 | 16 | Loan Refused |
| TC02 | 17 | Loan Refused |
| TC03 | 18 | Standard Loan Offer |
| TC04 | 55 | Standard Loan Offer |
| TC05 | 56 | More Expensive Loan Offer |
| TC06 | 57 | More Expensive Loan Offer |

3-value

Model 3

| Test Case | Input Value | Expected Output |
|-----------|-------------|---------------------|
| 1 | 16 | not admitted |
| 2 | 17 | not admitted |
| 3 | 18 | not admitted |
| 4 | 54 | standard loan |
| 5 | 55 | standard loan |
| 6 | 56 | standard loan |
| 7 | 57 | more expensive loan |
| 8 | 100 | more expensive loan |

Incorrect 4-value

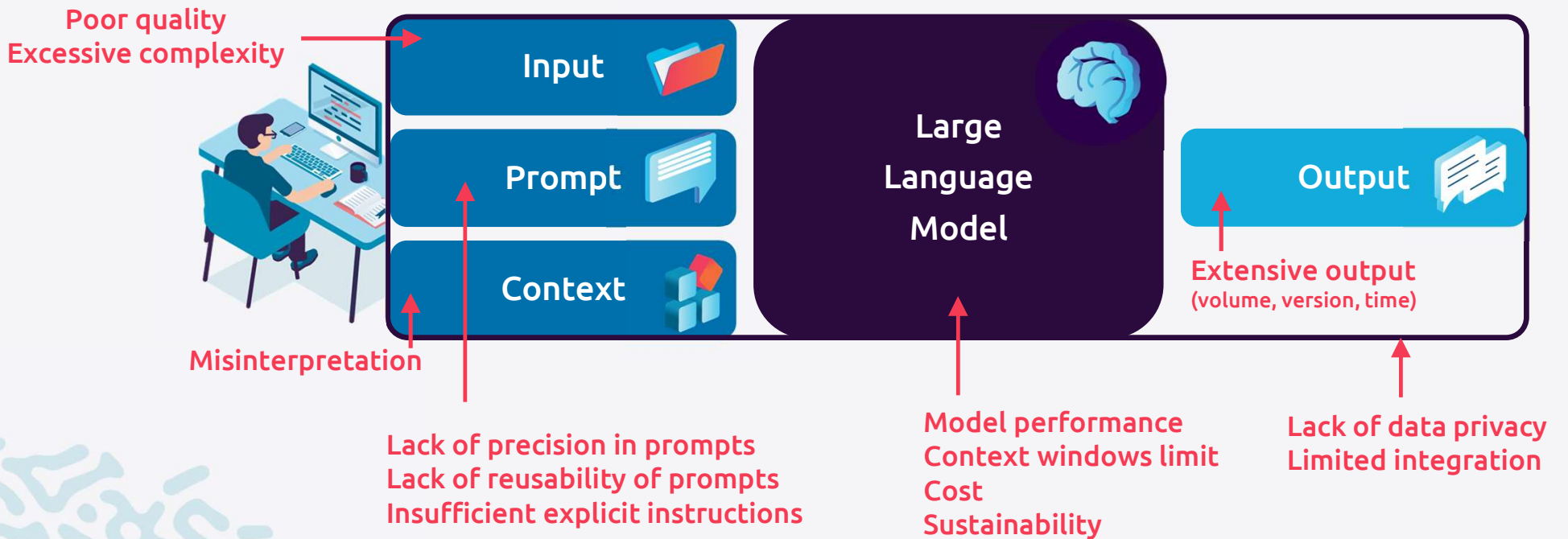
Model 4

Here are the test cases applying 4-value Boundary Value Analysis, presented as a table:

| Test Case | Input Value | Expected Output |
|-----------|-------------|---------------------|
| 1 | 16 | Refuse loan |
| 2 | 17 | Refuse loan |
| 3 | 18 | Standard loan |
| 4 | 19 | Standard loan |
| 5 | 54 | Standard loan |
| 6 | 55 | Standard loan |
| 7 | 56 | More expensive loan |
| 8 | 57 | More expensive loan |

4-value boundary

From using Large Language Models



| words | quality | time | cost |
|-------|---------|------|--------|
| 10 | ★★★★★ | | |
| 100 | ★★★★★ | 0 | \$ |
| 500 | ★★★★★ | 000 | \$\$ |
| >1K | ★★★★★ | 0000 | \$\$\$ |



Section 3

Introduction to Gen AI Amplifier

Gen AI Amplifier for Software & Quality Engineering

Homegrown Accelerator of Services

Ready-to-use

Fortified

60+ use cases

Bring-your-own-use-case

The screenshot displays the Gen AI Amplifier interface. On the left is a 'Menu' with options: Plan, Design, Build, Integration, Test, Deploy, Refactor, Generic, and BYOU Usecases. The main area is split into 'Requirement Creation' and 'Output'. 'Requirement Creation' includes fields for 'Choose Model', 'Upload File', 'Select Target Audience', and 'Choose Requirement Type'. The 'Output' section shows a flowchart and a 'Requirements Table'.

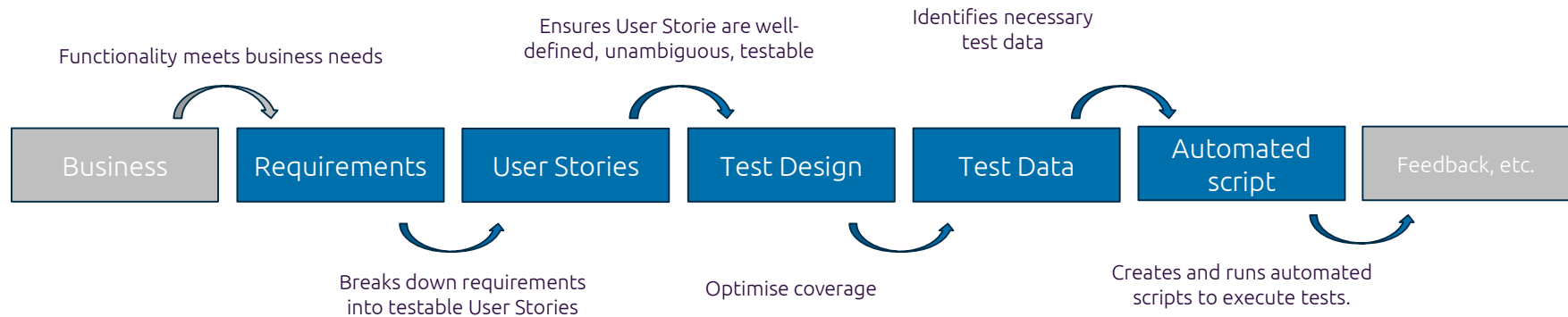
| Requirement ID | Requirement Type | Priority | Business Processes | Detailed Description | Acceptance Criteria | Total Function Points |
|----------------|------------------|----------|------------------------|--|--|-----------------------|
| 1 | Functional | High | Sales Order Management | The system must allow the Sales Order Manager to create sales orders. | The Sales Order Manager can create a sales order and the system records the order details. | 2 |
| 2 | Functional | High | Shipping Management | The system must display the status of partially shipped orders and indicate remaining items as 'Not Picked'. | The system shows the status 'Partially Shipped' for orders with unpicked items and marks remaining items as 'Not Picked'. | 2 |
| 3 | Functional | High | Invoicing | The system must prevent invoicing for items that are not picked and shipped. | The system blocks the user from generating invoices for items that have not been picked and shipped. | 2 |
| 4 | Functional | High | Invoicing | The system must allow the Invoice Processing Manager to generate customer invoices for shipped items. | The Invoice Processing Manager can generate an invoice (e.g., Customer Invoice 80011027) for shipped items, which is recorded in the system. | 2 |
| 5 | Functional | Medium | Shipping Management | The system must display planned shipping dates and statuses for remaining items. | The system shows planned shipping dates (e.g., February 23, 2020) and marks remaining items as 'Not Picked'. | 2 |
| 6 | Functional | Medium | Shipping Management | The system must track and display actual shipment dates for | The system records and displays actual shipment dates (e.g., February 20, 2020) for shipped items. | 1 |

Cloud native
LLM agnostic
Authentication
Data privacy
Guardrail
Cost control
User management
Integration

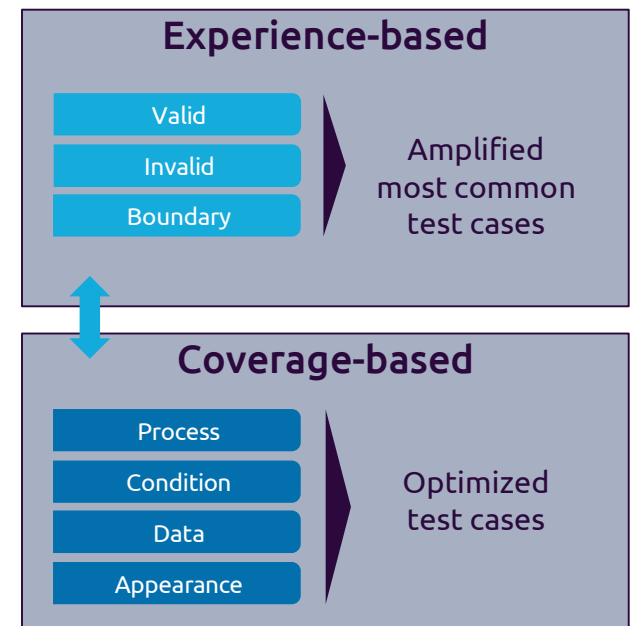
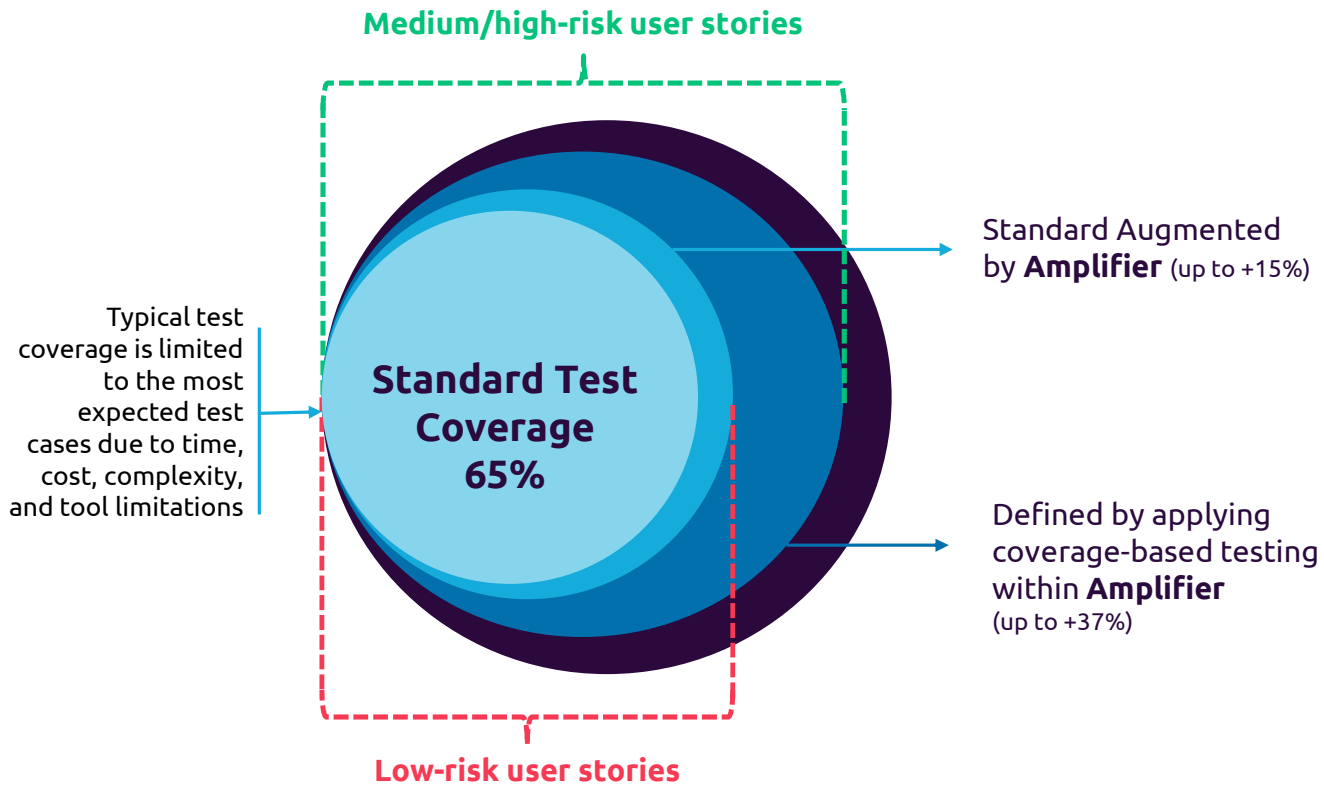
Knowledge System (RAG)

Output versioning
Low prompt no prompt

What if we could finish testing **before** a single line of code?



The amplified functional tester*



<https://www.tmap.net/building-blocks/test-design-techniques>

Test Design Technique Advisor



Analysis of the acceptance criteria and scenarios outlined in the risk-sensitive user story

Gen AI Amplifier
for Software & Quality Engineering

Test Advisor

This feature evaluates and recommends the best test design techniques and coverage levels for user stories, ensuring comprehensive testing tailored to the project's specific needs.

Choose Model:
OpenAI GPT-o1 Preview

Upload File | Enter Text

Upload file here:

ADV - Password user story.txt

Enter the Test Basis: information further guiding test creation context, skills, risks:

- High risk: If the password validation fails, it can lead to unauthorised access and security breaches.
- Medium risk: Poor error handling or unclear messages could lead to a poor user experience and non-compliance with user-friendly design standards.
- The password policy must comply with OWASP (Open Web Application Security Project) guidelines for password strength

Clear

Reset Generate

Output

US 1201: Password Security Enhancement

This user story primarily focuses on Conditions/Decision Points, Data, and Process. Recommendations are based on the available test basis. The risk level is High due to security implications.

| Coverage Group | Test Technique | Test Coverage Level | Precedence |
|----------------|-----------------------------------|----------------------------|------------|
| Conditions | Semantic Testing | MDCD (●●●) | 1 |
| Data | Equivalence Partitioning with BVA | EP combined with BVA (●●●) | 2 |
| Process | Process Cycle Test | TDS-2 (●●●) | 3 |

Explanation:

- **Conditions:** Semantic Testing with Modified Condition/Decision Coverage (MDCD) is recommended to thoroughly validate the password criteria, ensuring that each condition (minimum length, uppercase letter, lowercase letter, digit, special character) independently affects the outcome. This technique helps detect faults in the password validation logic by testing all possible combinations of conditions. This approach may require a high number of test cases due to the multiple conditions and their combinations.
- **Data:** Equivalence Partitioning combined with Boundary Value Analysis (BVA) is recommended to test valid and invalid password inputs, focusing on boundary conditions such as the minimum length of 12 characters. This technique efficiently identifies defects related to data input at the boundaries and within partitions. It is expected to generate a medium number of test cases, concentrating on data variations and boundary conditions.
- **Process:** Process Cycle Test with coverage level TDS-2 is recommended to ensure all password-related processes (account creation, password update, password reset) are thoroughly tested, including both successful and unsuccessful scenarios based on password compliance. This technique is expected to require a low number of test cases, as it focuses on covering the main process workflows.
- **Appearance:** Appearance testing is not relevant for this user story, as it focuses on backend functionality and validation logic rather than user interface elements or non-functional aspects.

Time taken for execution: 22.25 seconds | Total tokens: 6545

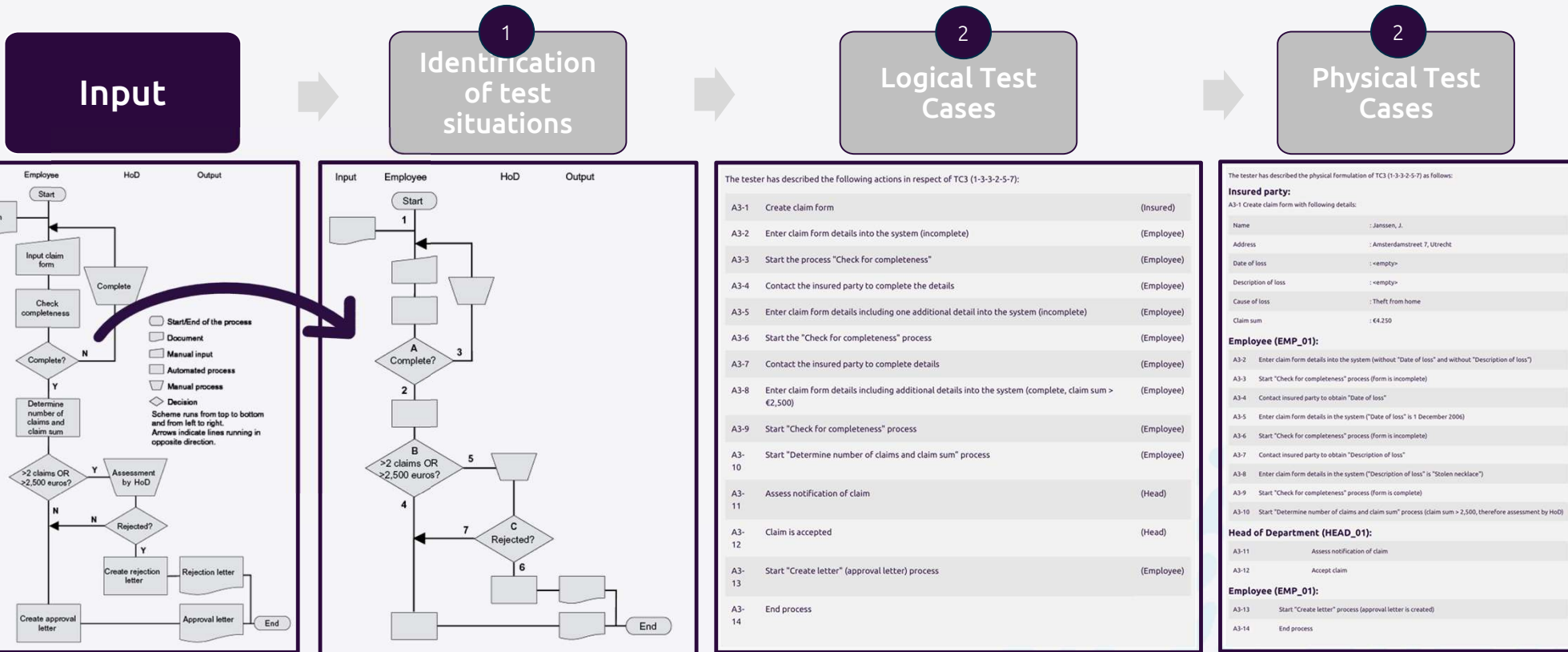
Price: € 0.22723

Reiterate on your output...

Tailored, precise, context-based recommendations for coverage levels and test design techniques

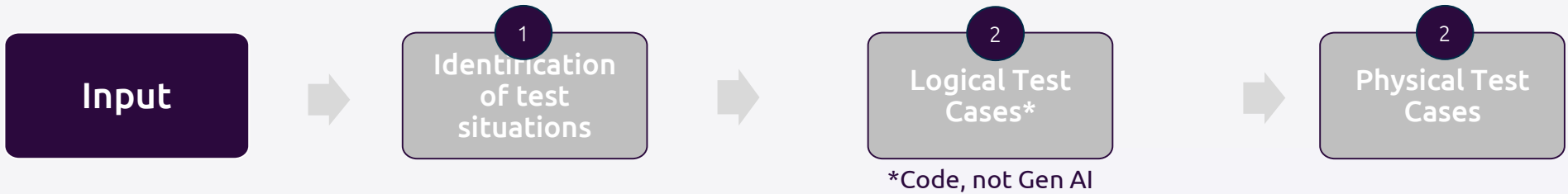
Process-driven testing: Process Cycle Test

Verify that individual actions within a procedural flow can be executed correctly by covering **all variations** in processing paths and decision points.

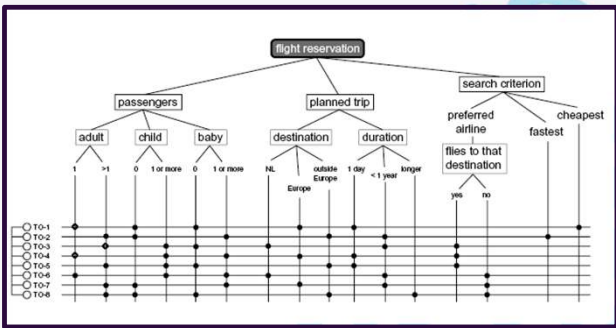
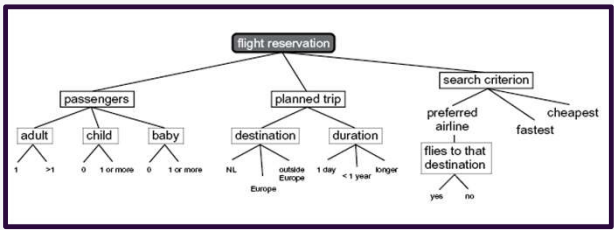


Data-driven testing: Pairwise testing

Verify that every pair of input parameters is tested at least once



2^8 Password Criteria
 2^3 Input Fields
2 Button State
 2^8 Error Message
 2^8 Visual Indicator
=
1,048,576 combinations



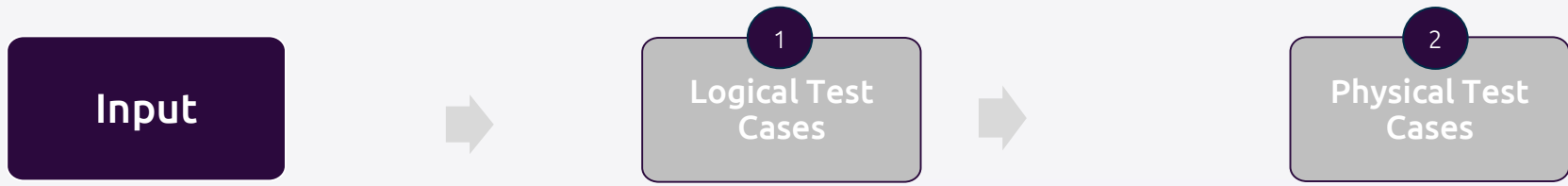
Example

In order to illustrate the principle, 4 test cases are made physical in the figure below. With every test case, the physical values are defined for all the necessary input data and the predicted result is described in concrete terms.

| | TG-1 | TG-2 | TG-3 | TG-4 |
|-------------------|------------|------------|-----------------------------------|---|
| Customer name | Jansen | Bireugel | Voort | Hansma |
| #Adults | 1 | 2 | 3 | 1 |
| #Children | 0 | 0 | 1 | 4 |
| #Babies | 0 | 2 | 0 | 1 |
| Destination | France-CDG | Singapore | The Netherlands Eindhoven Airport | The Netherlands Eindhoven Airport |
| Departure date | 12-02-2006 | 14-02-2006 | 15-02-2006 | 16-02-2006 |
| Return date | 12-02-2006 | 15-02-2006 | 15-04-2006 | 23-02-2006 |
| Search criterion | Cheapest | Fastest | KLM | Senegal Airlines |
| Predicted result: | | | | Message: "Airline does not fly to chosen destination" |
| Airline | Korean Air | Canada Air | KLM | |
| Flight number | KAD455 | CAN833 | KL1288 | |
| Price | € 44 | € 85 | € 83 | |

Condition-driven testing: Modified Condition / Decision Coverage

Verify that every possible outcome of a condition is the determinant of the outcome of the decision at least once.



```

IF password meets all criteria THEN
create account successfully ELSE
display appropriate error message

IF password_length >= 12
AND password_contains_uppercase = TRUE
AND password_contains_lowercase = TRUE
AND password_contains_digit = TRUE
AND password_contains_special_character = TRUE
AND password_contains_username = FALSE
AND password_contains_email = FALSE
AND password_is_common = FALSE

THEN password_validation_status := 'Valid'

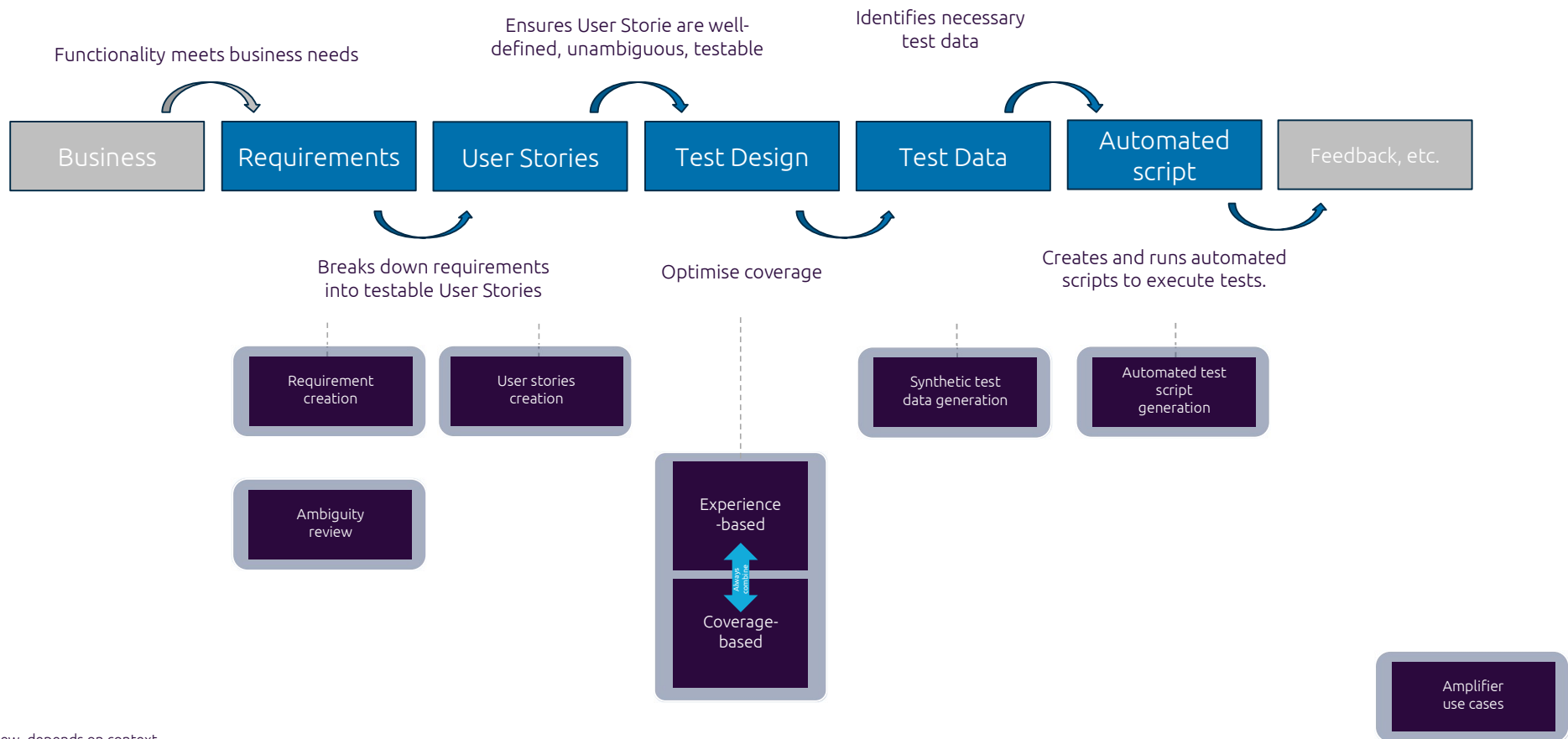
ELSE
IF password_length < 12 THEN
display_error_message := 'Your password must be at least 12 ch
ENDIF

IF password_contains_uppercase = FALSE THEN
  
```

| Test Case ID | Password_Length >=12 | Includes_Uppercase | Includes_Lowercase | Includes_Digit | Includes_Special_Ch |
|--------------|----------------------|--------------------|--------------------|----------------|---------------------|
| TC1 | True | True | True | True | True |
| TC2 | False | True | True | True | True |
| TC3 | True | False | True | True | True |
| TC4 | True | True | False | True | True |
| TC5 | True | True | True | False | True |
| TC6 | True | True | True | True | False |
| TC7 | True | True | True | True | True |
| TC8 | True | True | True | True | True |
| TC9 | True | True | True | True | True |

| Test Case ID | Password_Length >=12 | Includes_Uppercase | Includes_Lowercase | Includes_Digit | Includes_Special_Character | NOT Contains_Username | NOT Contains_Email_Address | NOT Is_Common_Password | Expected Result | Actual Results | Password |
|--------------|----------------------|--------------------|--------------------|----------------|----------------------------|-----------------------|----------------------------|------------------------|-----------------|----------------|--------------------------------|
| TC1 | True | True | True | True | True | True | True | True | Accepted | | P@ssw0rd1234! |
| TC2 | False | True | True | True | True | True | True | True | Rejected | | P@ssw0rd1! |
| TC3 | True | False | True | True | True | True | True | True | Rejected | | p@ssw0rd1234! |
| TC4 | True | True | False | True | True | True | True | True | Rejected | | P4ssw0rd1234! |
| TC5 | True | True | True | False | True | True | True | True | Rejected | | P@ssw0rd!!!! |
| TC6 | True | True | True | True | False | True | True | True | Rejected | | P4ssw0rd1234567 |
| TC7 | True | True | True | True | True | False | True | True | Rejected | | P@ssw0rd1234!user |
| TC8 | True | True | True | True | True | True | False | True | Rejected | | P@ssw0rd1234!email@example.com |
| TC9 | True | True | True | True | True | True | True | False | Rejected | | Password1234! |

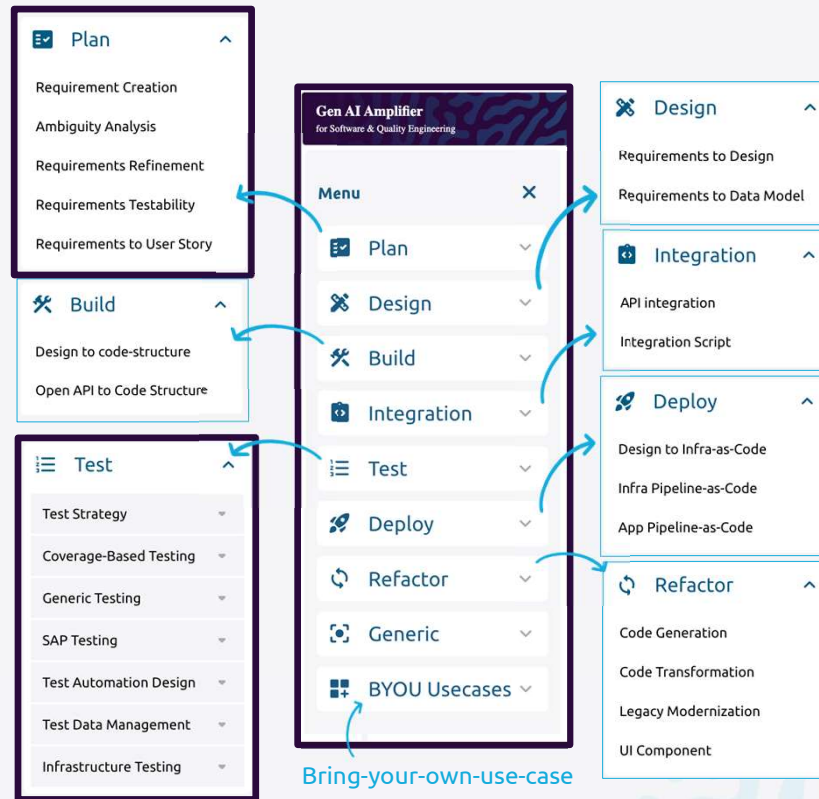
A journey inside the Amplified process*



*Simplified view, depends on context

Gen AI Amplifier for Software & Quality Engineering

Homegrown Accelerator of Services



Demo





Section 4

Productivity gains

Productivity Impact is Highest when Amplifier is Used End-to-End

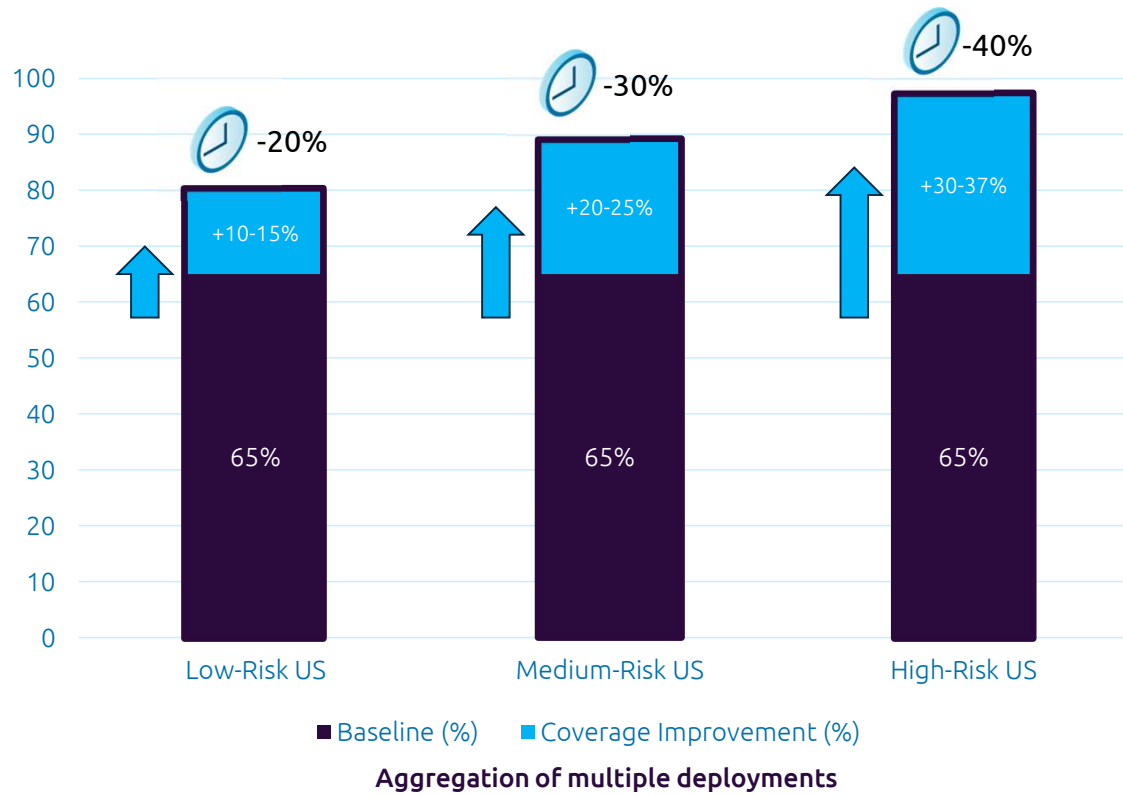


| POV Example | Input | Output | Manual Effort | Gen AI Amplifier | | | Savings |
|-------------|--------------|----------------------|---------------|------------------------|-------------------------|----------------------|---------|
| | | | | Pre-Process Effort [A] | Post-Process Effort [B] | Total Effort [A + B] | |
| 1. | User Stories | Test Cases | 11h27 | 5h27 | 3h43 | 9h10 | 20% |
| 2. | User Stories | Test Cases | 41h40 | 21h | 12h20 | 33h20 | 20% |
| 3. | User Stories | Test Cases | 12h30 | 1h47 | 6h05 | 7h52 | 37% |
| 4. | Transcript | Requirements | 16h00 | 0h30 | 6h00 | 6h30 | 59% |
| | Requirements | User Stories | 8h00 | 0h18 | 4h00 | 4h18 | 46% |
| | User Stories | Test Cases | 10h13 | 1h23 | 4h45 | 6h08 | 40% |
| | Test Cases | Katalon Test Scripts | 24h00 | 3h00 | 12h00 | 15h00 | 38% |

Productivity gain on Test Design

1

Productivity impact is maximized when applied across the entire process



2

The higher the risk, the greater the improvement in effectiveness and efficiency

Observations

1


The mandate for 'Expert in the loop'

2

Focus on the end-to-end process (e.g. high-quality input)

3

Productivity gains depend heavily on the context



Section 5

Deployment models & cost

Deployment model

Amplifier-as-a-Service

Internal-only shared environment operated by Sogeti

Sogeti hosted

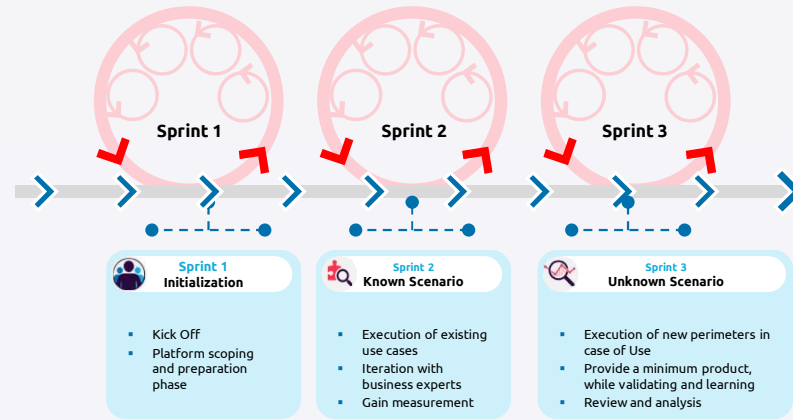
Sogeti deploys, hosts and manages for one single Client

Client hosted

Client deploys, hosts and manages
Sogeti provides updates for deployment

| | | | |
|---------------------------------------|--|---|---|
| Available immediately | Yes | No | No |
| Dedicated instance | No | Yes | Yes |
| Full feature access | Yes | Yes | Yes |
| Maintenance, Support, Upgrades | Yes | Yes | Yes Deployment by Client |
| Environment | Azure | Azure | Azure, AWS, GCP |
| Large Language Models | OpenAI | OpenAI | Client's decision |
| User Management Governance | Sogeti SPOC | Capgemini SuperAdmin | Capgemini SuperAdmin |
| Users | <ul style="list-style-type: none"> • Capgemini personnel • Capgemini authorized sub-contractors • No Client or no third-party personnel | <ul style="list-style-type: none"> • Capgemini personnel • Capgemini authorized sub-contractors • Client and third-party personnel <p>Note: Up to 100 users, maximum 1:1 Client/Capgemini user ratio</p> | <ul style="list-style-type: none"> • Capgemini personnel • Capgemini authorized sub-contractors • Client and third-party personnel <p>Note: Up to 100 users, maximum 1:1 Client/Capgemini user ratio</p> |

How to get started?



Discovery

Advanced demo

Contextualised Pilot

Production

at scale

Proposed Test Coverage Calculation

