

Operationalising AI in payments testing

Navigating deterministic gates
and human accountability



Phillip Stranger

AI Subject Matter Expert,
Unifits



Carl Slabicki

Head of Commercial,
Global Payments & Trade,
BNY



Michael Shklovskiy

ISO20022 and Payments
Transformation Lead Quality
Manager, Deutsche Bank

Contents

- 00 | **Market Context**..... 03
- 01 | **Introduction**..... 04
- 02 | **The story so far: Where AI is already providing value in payments testing**..... 05
- 03 | **Ensuring auditable and explainable AI model orchestration**..... 07
- 04 | **ISO 20022 as a case study for AI in payments testing** 09
- 05 | **Conclusion**..... 11
- 06 | **About**..... 12

00 | Market Context

In the world of payments — with 24/7 rails, instant schemes, ISO 20022 migration, richer data payloads, and tighter fraud and sanctions controls — complexity is rising while delivery windows shrink. In this context, the question for senior leaders is not whether to use AI, but how it fits in a controlled testing architecture.

When it comes to payments testing, AI can deliver immediate value by accelerating code creation, test-case design, data translation and synthetic data creation for ISO 20022 environments, and 24/7 operations support. Yet in order to achieve these benefits, financial institutions need to create tight guardrails to ensure hallucination risk is mitigated, and auditability remains throughout.

This report summarises the discussion from a Finextra webinar, hosted in association with Unifits, featuring a panel of industry experts, including:



Phillip Stranger

AI Subject Matter Expert, Unifits



Carl Slabicki

Head of Commercial,
Global Payments & Trade, BNY



Michael Shklovskiy

ISO20022 and Payments Transformation
Lead Quality Manager, Deutsche Bank

01 | Introduction

Payments testing has historically been a complex challenge for financial institutions. Strict validation, reliable regression testing, full traceability, and intricate regulatory requirements leave little room for error, especially as payment methods continue to diversify.

AI can be a key enabler in streamlining testing, improving automation, usability, and contextualisation — but only if set up appropriately. Particularly in ISO 20022-based environments, the probabilistic nature of AI systems can clash with deterministic nature of regulated testing where validation rules need to be continuously applied.

So, how can financial institutions create payments testing setups that leverage the best that AI has to offer without adding risk or friction? Creating hybrid testing models enables organisations to use AI to complement analysis, insight generation, and efficiency gains, while deterministic systems remain responsible for execution and validation under human oversight. Especially in the ISO 20022 era, AI-augmented payments testing can be a key differentiator for financial institutions.

This webinar report seeks to answer:

- ① What are the key considerations for AI in payments testing? Where does it excel, and what are its limitations?
- ② How can financial institutions use AI to create efficient, usable, and scalable testing setups without compromising the reliability and accountability required in regulated environments?
- ③ How can financial institutions leverage AI to test in ISO 20022-based environments? What are the opportunities for AI testing as we move towards the next ISO 20022 milestones?
- ④ What governance, controls, and auditability frameworks are required when integrating AI into payments testing within regulated financial environments?
- ⑤ How can financial institutions balance AI-driven test generation and optimisation with the need for explainability, reproducibility, and regulatory transparency?

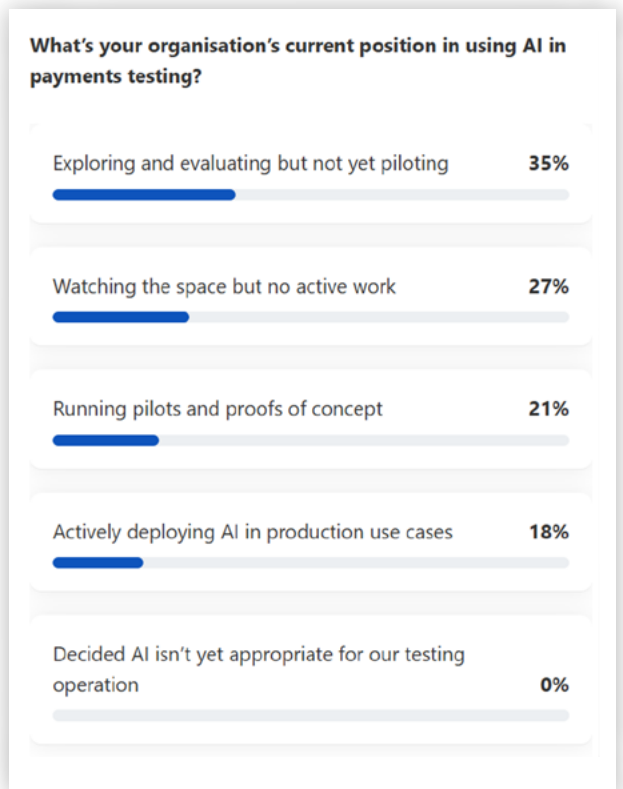
02 | The story so far: Where AI is already providing value in payments testing

The world of payments has evolved rapidly over the last decade. From the proliferation of real-time payments, the expansion of payment rails, and the transition to ISO 20022 — testing has been crucial in maintaining high straight-through processing (STP) rates as complexity increases.

To set the scene for the session, a first audience poll (pictured on the right) found that when it comes to payments testing, only 18% of participants were already actively deploying AI in production use cases. The majority of respondents (35%) were exploring and evaluating AI, but not yet piloting.

The results reflect a good pulse check on the industry which, in many aspects, is still in the infancy stages of AI rollout compared to other industries. Explaining the slower rollout in financial services, Shklovskiy commented:

“Financial institutions operate with specific guardrails for regulations. The probabilistic nature of AI is a direct clash with what regulation demand, which is absolutely certainty and being able to audit to 100%. While AI can be incredibly accurate and highly probable, it’s not exactly 100%. The question here is not whether AI is useful — it is. It’s whether the output is sufficiently reliable for specific control testing.”



While the AI rollout may be slower and more controlled in financial services, there are already successful use cases specifically in payments testing. Slabicki explained the approach at BNY:

“When it even comes to production use cases, what are the things that banks and global transaction entities worry about? It’s: how do we best route payments smartly? How do we optimise FX? How do we choose the best rails? How do we work with data translation and how do we optimise straight-through processing and handle exceptions? And AI is really helping all of that.”

AI can support both in the back and front office, in writing user stories, drafting code, writing test cases and helping the acceleration of engineering teams. Yet it can be hard for financial institutions to determine where to start.

Stranger highlighted the key is to start by determining where AI fits, and keeping a clear separation between AI and non-AI layers. For example, in validation, a payment message is either valid or not valid — not a probability in between — making validation an inappropriate fit for AI.

“The starting point is actually the architecture,” Stranger emphasised. “So, where do we need our deterministic test system, that the AI can then use as a tool? Then, you have your audit trail, you know exactly which message was validated with which rules.”

He added: “So, how do we construct that? The LLM can do its thing, but it’s boxed in, and the deterministic side of things stays deterministic. We already see this in coding agents, which have come to life especially this year with agentic tool use. And the same pattern can be applied to payments testing beautifully. In coding, for example, you have your test scripts, linters, and compilers. The AI can use all of these tools, create code, check the code, see if it has made an error, and then self-heal. And that can be translated beautifully to payments testing.”

Unlocking this potential, however, requires meticulous planning and strict operational control. So, what are the specific design, discipline, and governance challenges that need to be addressed in the process?



In coding, for example, you have your test scripts, linters, and compilers. The AI can use all of these tools, create code, check the code, see if it has made an error, and then self-heal. And that can be translated beautifully to payments testing.

Phillip Stranger, Unifits

03 | Ensuring auditable and explainable AI model orchestration

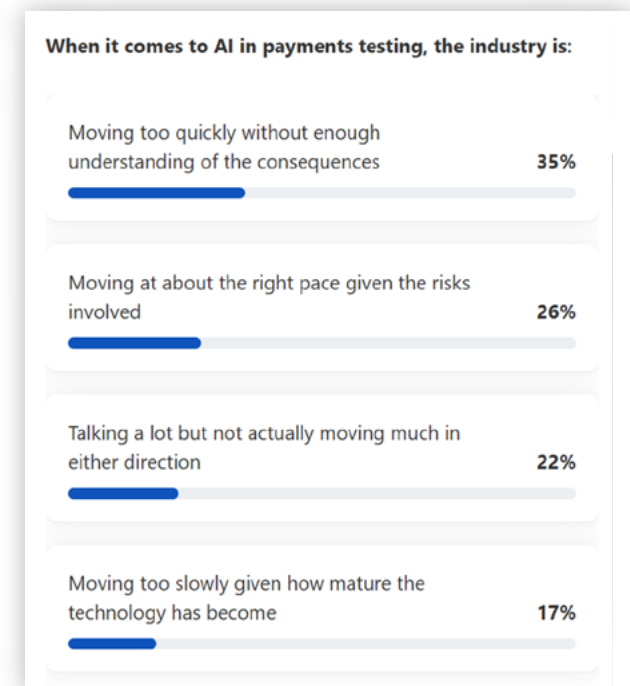
Careful orchestration becomes key in the process of integrating AI into the payments testing process. To set the scene, a second audience poll (pictured on the right) asked participants about the speed of implementation across the industry. Emphasising the risks that AI can pose in payments testing if not rolled out safely, 35% of participants stated the industry moves too quickly without enough understanding of the consequences.

26% of participants thought the industry was moving at the right pace given the risks involved, while another 22% believed there was a lot of talk but not enough movement in either direction.

Commenting on the poll, Slabicki explained: “I think it goes back to understanding the audit trail decision process, to understand the sources of the information and where the recommendations are coming from. That’s how the AI orchestration ultimately works.”

There are three key considerations the panellists highlighted in the context of AI in payments testing: architecture setup, the risk of hallucinations, and the human-in-the-loop.

When it comes to the architecture itself, organisations need to have data that’s appropriate for the LLM. “There are certain types of data types which work better for LLMs,” highlighted Stranger. “That’s stuff that you have to think about. And then the second part is the framework. Where in your setup does the AI fit and can interact with your other systems? [BNY’s Eliza] is probably very integrated, and so it’s more useful than an AI system that’s not fully integrated with your systems.”



How easily financial institutions can process this integration depends on the underlying infrastructure. The eternal question in finance applies here as well: how much legacy setups is a financial institution still running on? AIs can work very well with APIs, so if an organisation has already started making the transition to modern data types, it will ultimately also become easier to integrate AI.

Once the layout is set, how can financial institutions address the hallucination risk? Stranger explained: “If you split the deterministic test system and the AI as a separate layer, then the deterministic system can check the AI layer. And if the AI starts to hallucinate, the deterministic test system will catch it.”

This concept of ringfencing and repressibility is the foundation for successfully integrating AI into testing processes, yet it’s equally crucial to keep the human-in-the-loop.

Shklovskiy commented: “For AI to be trusted as a verified output, the results generated must be understandable by human experts. If an LLM generates a complex ISO message for testing, for example, it should be in the format that business analysts can reasonably inspect and validate against the standard and business rules. In automated testing, the process must be reproducible in exactly the same way. Any randomness introduced by AI must be managed to allow deterministic recreation of specific run test cases.

“Again, I’m going to go back to the fact that we are in regulated environments. ‘Almost correct’ or ‘highly probable’ validation is not acceptable. Human responsibility also is needed to emphasise that AI is a tool to empower humans, not to replace their ultimate responsibility for integrity and compliance of payment processes.”



If you split the deterministic test system and the AI as a separate layer, then the deterministic system can check the AI layer. And if the AI starts to hallucinate, the deterministic test system will catch it.

Phillip Stranger, Unifits

04 | ISO 20022 as a case study for AI in payments testing

As the industry enters the ISO 20022 era and continues to transition according to Swift's roadmap, AI becomes a crucial tool in the testing process. The transition to ISO 20022 has been one of the most fundamental changes in financial services recently. As the first milestones have passed with more on the horizon, financial institutions are working hard on their data estates to ensure compliance.

"The biggest challenge is the complexity of the standard," emphasised Stranger. "There is so much business knowledge involved and you need proper experts to understand everything. If one of your experts leaves the company, you might have a real hole in terms of business knowledge. I think this is where AI can definitely help. Once you have the knowledge in a system where you can effectively use it, that's where you get some real value."

ISO 20022 milestones in 2026 and 2027 centre around structured address migration and statement and reporting message types. These deep system changes mean that projects cannot be rushed, testing needs to be extensive, and teams that have not already started are setting themselves up for more pressure further down the line.

Shklovskiy emphasised: "I think people underestimate the length of work that is required to verify all the testing that has been done and all the testing that they have to redo. That's why starting now is important, especially in the context of AI-augmented testing. The ISO messages are incredibly rich and granular compared to legacy formats. AI is a great helper here. End-to-end means that ISO impacts multiple internal systems, so people should not underestimate that they need to change a lot of the systems, data, and touch points."



I think people underestimate the length of work that is required to verify all the testing that has been done and all the testing that they have to redo. That's why starting now is important, especially in the context of AI-augmented testing.

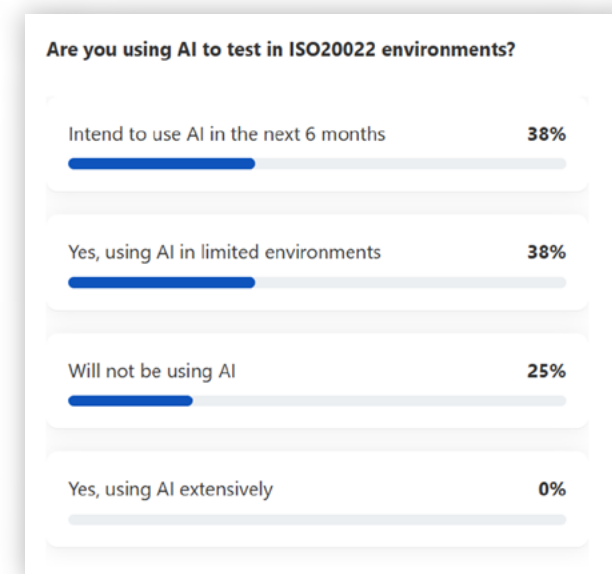
Michael Shklovskiy, Deutsche Bank

Within the context of the upcoming structured address migration, there are multiple ways in which AI can support teams. One is by using synthetic data for message testing (unless specific systems are tested that need real data for deterministic decisions). Another way AI can support is by helping to structure address data based on all available consumer information.

Slabicki highlighted AI can also help by learning from historical code deployments and tests. “Adding robust test cases to test plans based off of experience of other code deployments, applying it to the current situation and amplifying a super robust testing plan in the matter of seconds or minutes, and ultimately helping run scenario analysis and testing. Again, it really helps provide such a greater scale to whatever the problem is.”

This adaptability becomes key for AI in payments testing environments. AI is great at predicting where vulnerabilities are most likely to occur, meaning that frameworks have to be adaptive to account for changes overtime, which becomes even more crucial in standard-led environments such as ISO 20022.

Underscoring the role AI can play in the ISO 20022 transition, a final audience poll (pictured on the right) found that three out of four participants were either already using AI in limited production to test in ISO 20022 environments, or were planning to do so within the next six months.



05 | Conclusion

The financial services industry is only starting to reap the full potential of AI. Today, it is already delivering benefits when it comes to payments testing, and acting as a translation layer for complex transitions such as ISO 20022.

However, to full tap into the potential that AI can offer, financial institutions need to consider many factors: data readiness, machine readable imports, architecture, governance, audit trail, and rail clarity. Separating AI layers from deterministic systems becomes crucial to maintain auditability and drive down risk.

As with many technological innovations, AI success depends on strong foundations. Successful model orchestration will turn AI into a scale enabler — enhancing human employees, speeding up processes, and improving payments testing processes.

About

Finextra Research

This report is published by Finextra Research.

Finextra Research is the world's leading specialist financial technology news and information source. It offers more than 130,000 fintech news, features and TV content items to some 800,000 monthly visitors to www.finextra.com.

Finextra covers all aspects of financial technology innovation involving banks, institutions and vendor organisations within the wholesale and retail banking, payments and cards sectors worldwide. Finextra's unique member community consists of over 40,000 fintech professionals and 200,000 social followers working inside banks and financial institutions, specialist fintechs, consulting organisations and technology providers.

The Finextra community actively participates in contributing opinions, ideas and comments on the evolution of fintech.

For more information visit www.finextra.com and become a member, follow [@finextra](https://twitter.com/finextra) or email us via contact@finextra.com.

Unifits

Unifits: Your Global Partner for ISO 20022 Compliance

Unifits stands as the trusted global partner for ISO 20022 compliance in the financial industry. With the Unifits Test Engine, financial institutions can leverage out-of-the-box clearing simulations to automate transaction testing, significantly enhancing quality and efficiency. This innovative solution is kept evergreen by the dedicated Unifits team, comprising software specialists and financial industry experts, allowing financial institutions to prioritise customer satisfaction and business growth, free from the complexities of managing test cases, test data, and compliance concerns.

Learn more at: www.unifits.com

For more information

Finextra Research

77 Shaftesbury Avenue
London
W1D 5DU
United Kingdom

contact@finextra.com

[@finextra](#)

www.finextra.com

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording or any information storage and retrieval system, without prior permission in writing from the publisher.

© Finextra Research Ltd 2026