LIVE FIELD EXPERIMENTS

AI in the Real World

Interim Experiment Report - Day 11

Alex: AI Sales & Marketing | Morgan: AI Market Experiment

30 September 2026

Two AIs. Two different jobs. Thirty days. Measurable outcomes. This report records what has happened so far - including useful capability, commercial resistance, methodological limitations and results that have not yet materialised.

Executive summary

These experiments test whether currently available AI can produce useful, measurable real-world outcomes rather than merely impressive-looking activity. Alex is attempting to sell 100 paid Innocent Soap dog-shampoo bars in 30 consecutive calendar days. Morgan is recording forward-looking market forecasts and managing a £10,000 hypothetical portfolio under an explicit experimental framework.

At the Day 11 checkpoint, neither experiment has produced a financial gain. Alex has demonstrated that AI can operate much of a supervised B2B sales pipeline, while the sales result remains 0/100. Morgan has demonstrated forecasting activity and restraint, while the virtual portfolio remains at £10,000 with no trades. The distinction between activity, capability and outcome is becoming one of the central findings.

	ALEX	MORGAN
Checkpoint	Day 11 / 30	Trading Day 4 / 30
Primary objective	100 paid bars	Market forecasts + virtual portfolio
Financial result	£0.00 verified revenue	£0.00 P/L
Activity	61 initial approaches; 29 follow-ups	8 markets analysed; no trades
Measured outcome	0 / 100 paid bars	£10,000 portfolio; 0% return

> *Interim finding: AI can generate and manage activity. The harder test is whether that activity creates a measurable result.*

Experiment 01 - Alex

Can an AI actually sell something?

Alex is the AI Sales & Marketing Manager for Innocent Soap. The objective is deliberately simple to score: sell 100 paid 150g dog-shampoo bars within 30 consecutive calendar days. A sale counts only when paid-order evidence is verified. Robin's test purchase is excluded.

DAY	PAID BARS	VERIFIED REVENUE	INITIAL APPROACHES
11 / 30	0 / 100	£0.00	61
FOLLOW-UPS	**GENUINE BUYING ENQUIRIES**	**TRADE ORDERS**	**AI-ATTRIBUTED BARS**
29	1	0	0

What has happened

• **61 first-contact trade approaches** have been accepted by the mail client. Sixty have no known first-contact delivery failure.

• **29 follow-ups** have been processed. Twenty-seven were accepted and saved with no known immediate failure; one had uncertain Sent-folder evidence; one was confirmed blocked.

• **One genuine buyer** - West Country Pets Ltd t/a The Pet Express - requested range and minimum-order information. Product, ingredient, pricing and 12-bar mixed-MOQ information was supplied, but no order or payment followed by this checkpoint.

• **One useful rejection** came from Wag & Bone, which said solid shampoos had not sold well for it previously. That is commercially useful evidence about retailer resistance to the format.

• No acknowledgement, enquiry, promise, test payment, bounce or media response has been counted as a sale.

What Alex has demonstrated positively

The zero-sales result should not obscure the operational capability demonstrated. Under supervision and defined guardrails, Alex has researched and qualified prospects, personalised B2B outreach, monitored the campaign inbox, processed follow-ups, maintained an evidence ledger, separated enquiries from sales and suppressed failed routes. Those are useful sales-operations functions even though they have not yet produced revenue.

Emerging sales findings

Finding	Evidence so far
Activity is not revenue	61 initial approaches and 29 follow-ups have produced one genuine buying enquiry and no paid order.
The product format can face resistance	A retailer explicitly cited poor previous sales of solid shampoo.
Commercial clarity matters	The genuine buyer asked for range and MOQ information; clear opening-order terms supported that conversation.
More follow-up is not automatically the answer	Follow-ups have not yet generated a second buying enquiry.

Finding	Evidence so far
Evidence controls improve credibility	Test purchases, acknowledgements, bounces and enquiries remain separate from genuine paid sales.

Experiment 02 - Morgan

Can an AI make disciplined market decisions?

Morgan begins with a £10,000 hypothetical portfolio. The experiment separates two questions: whether AI can make forward-looking directional forecasts, and whether those forecasts can be turned into a disciplined, executable trading system. No real money is being invested.

TRADING DAY	PORTFOLIO	P/L	TRADES
4 / 30	£10,000	£0.00	0
OPEN POSITIONS	**RETURN**	**MAX DRAWDOWN**	**TRADING COSTS**
0	0%	0%	£0.00

Trading Day 4 forecasts - 30 September 2026

Market	Forecast	Confidence
Gold	Bullish	58%
Silver	Bullish	57%
Brent Crude	Bullish	59%
WTI Crude	Neutral / Bearish	55%
Copper	Bullish	60%
GBP/USD	Bearish	57%
EUR/USD	Bearish	62%
USD/JPY	Bearish	59%

Decision: NO TRADE

Morgan changed gold and silver from the bearish stance recorded on earlier trading days to cautiously bullish on Trading Day 4. Despite making directional forecasts, Morgan kept the portfolio entirely in cash. At this checkpoint there is therefore neither trading profit nor trading loss.

What Morgan has demonstrated positively

Morgan has not manufactured activity merely to make the experiment look busy. The system has recorded forward-looking calls while declining to put hypothetical capital at risk when the experimental rules do not justify a trade. That restraint is itself relevant to the test: a useful decision system must be able to say 'no trade', not simply generate transactions.

Important methodological finding

By Trading Day 3, the experiment had exposed a design limitation: although starting capital, maximum risk per trade, aggregate risk and maximum simultaneous positions had been established, the detailed transaction-cost schedule, exact position-sizing convention and mechanical benchmark were not completely specified before market behaviour was observed. Morgan therefore declined to create those missing rules retrospectively. Forecasting performance and portfolio performance are being treated as separate measures.

Emerging findings across both experiments

The two experiments are beginning to expose a common issue in evaluating AI: visible activity is easy to produce, but activity is not the same as an economically useful outcome.

Question	Alex	Morgan
Can the AI perform useful process work?	Yes: prospecting, outreach, follow-up, inbox and evidence management.	Yes: repeated market analysis, forward forecasts and rule-aware decisions.
Has that process produced a financial result yet?	No: 0 paid bars and £0 verified revenue.	No: £10,000 portfolio, £0 P/L and no trades.
What has been learned from failure or restraint?	Retailer objections and low conversion are becoming visible.	Incomplete pre-specified execution rules can invalidate apparently attractive trading activity.
What would count as stronger evidence?	Verified paid orders attributable to the campaign.	Scored forecast accuracy and/or rule-compliant portfolio performance over a meaningful sample.

Five interim conclusions

1. AI can execute substantial process work. Both experiments show useful operational capability without requiring the final outcome to be declared successful.

2. Process metrics can be misleading. Emails sent, forecasts issued and analyses completed are inputs, not the ultimate success measures.

3. Negative evidence is useful evidence. A retailer objection, an incorrect forecast or a methodological flaw can improve the next decision if it remains visible rather than being edited away.

4. Guardrails matter. Alex's commercial approval gates and Morgan's refusal to invent trading rules after observing the market protect the integrity of the tests.

5. It is too early for a final verdict. This is an interim checkpoint. Both experiments still need more observations before stronger conclusions can be drawn.

> *The useful question is not 'Can AI look impressive?' It is 'What happens when we actually try it?'*

Method, controls and transparency

Alex - sales evidence

Only verified paid bars count towards the 100-bar target. The owner's test purchase is excluded. Enquiries, acknowledgements, promises, bounces, sample requests and media responses are not sales. The campaign distinguishes mail-client acceptance from known delivery failure and keeps failed routes suppressed unless another verified route is found.

Alex - human supervision

Routine prospect research, normal B2B outreach, inbox monitoring, campaign records and measured follow-up can be handled within standing authority. Pricing changes, discounts, expenditure, unusual commercial terms, partnerships, media contact and public publishing remain subject to human approval. Facebook publishing is manual.

Morgan - market controls

Morgan uses hypothetical funds only. Forecasts are recorded before the subsequent outcome is known and are intended to remain in the historical record. The experiment distinguishes forecasting accuracy from portfolio performance and does not treat a 0% return as evidence of successful trading.

Why publish interim reports?

Publishing dated interim reports creates an audit trail. It reduces the temptation to rewrite the story after the final outcome is known and allows readers to see how the experiments, their limitations and their findings developed over time.

Next checkpoints
Continue the public record through the 30-day period. A later interim report can compare conversion, forecast scoring, human intervention and any change in financial outcomes. A final report should preserve the earlier reports rather than replace them.

Report date: 30 September 2026
Prepared for: Robin Saunders - AI Experiments
Experiment status: Interim - not final results

This document describes an experimental virtual trading exercise and is not investment advice. Morgan does not invest real capital. The sales experiment reports verified campaign evidence available at the stated checkpoint.