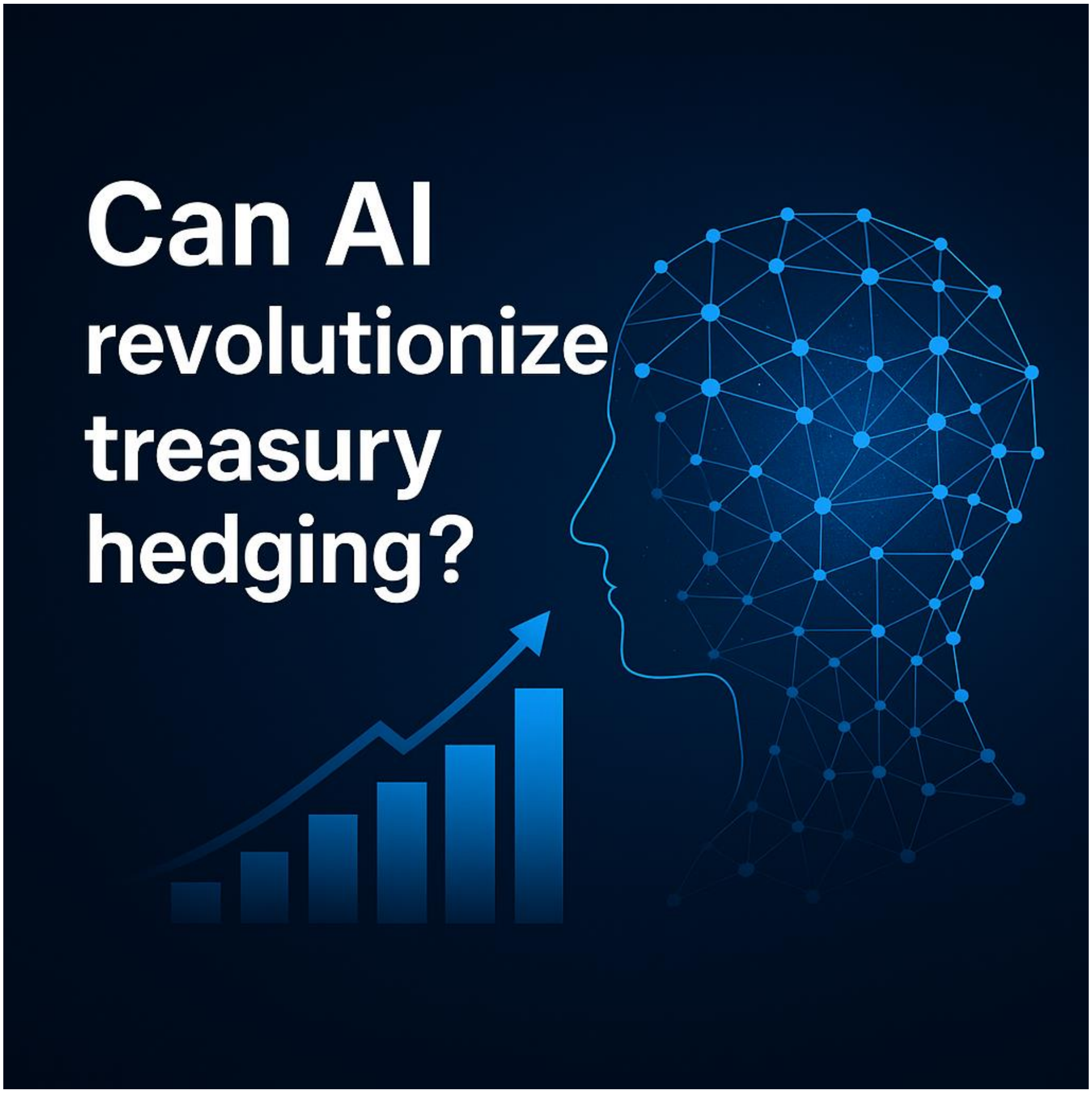

Revolutionizing Treasury Hedging: Why AI Outperforms Legacy Systems (e.g., Murex, Nasdaq Calypso, Finastra …)

Executive Summary

Why rely on static hedges when an AI engine can produce a portfolio- and market-aware hedge menu each run? We ran a controlled experiment comparing three approaches under the **same** portfolio, **same** market calibration, and **same** guardrails:

- **Traditional–Minimal:** 3-instrument shelf (IRS, UST future, FX option).

- **Traditional–Expanded:** a **fixed**, broader catalog that mirrors the instrument families AI typically considers (IRS across two tenors, OIS, multiple UST futures buckets, two FX options, an FX forward, and a USD/EUR cross-currency swap). Traditional-Expanded is a static shelf, not dependent on AI.
- **AI (Agent):** a **dynamic** menu generated from today's portfolio and the calibrated market backdrop, then optimized and guardrail-checked by the same solver.

All three used the **same optimizer**, **same guardrails** (VaR cap, FX-reduction target, duration-gap band), and the **same scenario set**. This makes the comparison apples-to-apples.

Key results from this run (see Optimal Hedge Baskets and Scenario Results

)

- **Post-hedge duration gap:** Minimal 2.344y | Expanded 2.260y | **AI 2.218y (best)**
- **Delta EVE improvement at +100 bp:** Minimal 2.0% | Expanded 8.0% | **AI 10.9% (best)**
- **VaR (95% daily) vs cap 683,113:** Minimal 630,341 | Expanded 658,088 | **AI 671,560** (all within cap)
- **FX +10% P&L reduction:** Minimal 11.0% | **Expanded 35.0%** | **AI 35.0%**

Bottom line: Under identical constraints, the AI menu tightened the duration gap and delivered the largest Delta EVE improvement while matching the Expanded catalog's FX protection and respecting the same VaR cap.

Demystifying the "Traditional–Expanded" benchmark

- **What it is:** A **static, pre-defined** shelf that mirrors the product families the AI typically considers (for fairness). It is **not** copied from the AI's output at runtime and **does not** depend on AI to execute. Any treasury can adopt it.
- **Why it's hard to maintain without AI:** To keep such a shelf fit-for-purpose, a team must **continually** curate tenor coverage, instrument mix, and headroom caps as the book and market conditions change. That is meeting-heavy and error-prone.
- **What AI changes:** The AI **automates** this curation each run, proposing a fresh menu that reflects the current portfolio and market calibration, then sizes it under the same guardrails as the traditional paths.

Why We Use a Deterministic Optimizer (and When Genetic Algorithms Make Sense)

Our approach today
For this study, hedge baskets were optimized using a **deterministic gradient-based method (L-BFGS-B)**.
We chose this because:

- It is **fast** and scales well to the relatively small hedge universes considered here (8–12 instruments).
- It is **reproducible**: the same inputs always yield the same hedge basket. This transparency is critical for governance, audit, and ALCO review.
- It works well under our setup where the solution space is already **constrained** by guardrails: VaR cap, FX reduction target, and a duration-gap band.

Will a genetic algorithm (GA) be better?
For our current universes, **probably not**. With only a handful of instruments and tight guardrails, gradient methods already converge to efficient, policy-compliant hedge baskets.
In fact, GA's added complexity (longer run-times, stochastic variability) would not translate into material improvement here.

Where GA could add value in treasury hedging
If the problem expands, GA or other evolutionary methods can become attractive:

- **Larger universes:** 50+ candidate instruments across currencies, tenors, and structures.
- **Scenario-weighted objectives:** optimizing across many shocks at once (e.g., parallel shifts, curve steepening, volatility, and FX combined), instead of one at a time.
- **Discrete constraints:** rules like "choose at most 3 FX options" or "include at least one long-dated IRS," which gradient optimizers struggle to handle.
- **Multiple solutions:** generating a *menu* of good hedge baskets for committee discussion, instead of just one "best" answer.

Bottom line
For the portfolio sizes and universes in this paper, **deterministic optimization is the right tool**—it is efficient, auditable, and fair.
As universes grow more complex, we can seamlessly switch to a **genetic approach** to explore richer trade-offs and multiple feasible hedge portfolios.

Inside the experiment (business view)

- **Inputs:** A real treasury portfolio (`treasury_portfolio.csv`).

- **Calibration window:** Last 12 months of US 10Y yields (^TNX) and EUR/USD.
  - US10Y daily rate volatility: **5.53 bp**
  - Average absolute correlation: **0.68**
- **Baseline (unhedged):** duration gap **2.37y**; VaR 95% daily **$621k**; FX +10% P&L (unhedged) **$3.625m**.
- **Hedge menus constructed:**
  - **Minimal** and **Expanded**: static shelves.
  - **AI**: dynamically generated, then budget-capped at the universe level.
- **For each menu we then:**

1. Optimized notionals,
2. Enforced guardrails (VaR ≤ 1.10× baseline; FX +10% reduction ≥ 35% when relevant; duration-gap band −0.20y to 0.80y),
3. Ran an identical scenario set (see next section),
4. Printed selections and scenario tables and wrote an audit artifact (`run_summary.json`).

Scenario set (real-world regimes, not just one case)

To evidence robustness beyond a single point estimate, we applied a comprehensive suite of stresses: a Base (no-shift) control; parallel rate hikes of +100 bp and +200 bp (tightening cycles) and a −100 bp cut (growth scare/flight-to-quality); a volatility shock of +20%; a liquidity-crunch stress with vol +40% and FX +10% (wider bid/ask, disorderly markets); a pure FX shock of +10% (currency insulation); a combined shock (rate +150 bp & FX +15%) capturing a policy surprise with concurrent USD strength; and curve-shape scenarios—steepening (+150 bp at the short end) and flattening (+150 bp at the long end). Each scenario was evaluated under the **same** guardrails, so outcomes are directly comparable across parallel and non-parallel moves, volatility regimes, and rate/FX dislocations.

Reading the outputs

- **Optimal Hedge Basket (per approach):** The final selection **after optimization and guardrail tightening**.
  - *Traditional–Expanded:* 8 instruments (IRS, OIS, UST futures, FX options/forward, cross-currency swap).
  - *AI:* similar families but **custom-sized** to this portfolio and calibration (e.g., added FX-option headroom to hit the FX target, then scaled rates to keep VaR within cap).
- **Scenario Results (per approach):** For each stress, we show post-hedge duration gap, Delta EVE, VaR, and FX P&L.

- **Head-to-Head (measured):** AI delivered the tightest gap (2.218y), the largest Delta EVE improvement at +100 bp (10.9%), matched the Expanded shelf's 35% FX-reduction, and stayed within the same VaR cap.

Why the AI path outperforms

- **Dynamic menu vs. fixed shelf:** AI tailors the instrument mix (and headroom) to **today's** exposures and calibration; static shelves cannot.
- **Closed-loop fitness feedback:** Propose → optimize → measure → refine. The evaluator returns actionable feedback (e.g., "gap still high," "FX reduction below 35%," "Delta EVE improvement below target"), and the menu is revised accordingly.
- **Guardrail-aware sizing:** The post-process tops up FX options (to reach the FX target) and then re-caps VaR by scaling rate hedges. With more degrees of freedom across tenors and families, the AI path finds **better trade-offs**.
- **Curve-shape resilience:** Diversified tenor placement (IRS/OIS plus multiple futures buckets) reduces sensitivity to steepening/flattening.
- **FX discipline without over-hedging rates:** FX and rates are co-sized under the **same** VaR headroom, avoiding unnecessary carry.

Fairness and methodology

- Two traditional benchmarks were run: **Minimal** (3-line legacy shelf) and **Expanded** (8-line shelf mirroring AI **families**, but static and pre-defined).
- All paths used the **same** optimizer, **same** guardrails, **same** scenarios, and a **portfolio-level budget cap** on the universe (≤ 50% of assets).
- The comparison is therefore **like-for-like** and defensible.
- **Operational note:** Without AI, a treasurer could build a similar Expanded shelf, but would need to **maintain** it continuously (tenor coverage, FX mix, caps). AI **automates** this maintenance each run.

Cost and ROI (concise view)

- **Trading and balance-sheet economics:** Right-sizing typically reduces tickets, avoids carry from over-hedging, and lowers roll/unwind frequency.
- **Process savings:** Calibration → proposal → optimization → guardrails → scenarios is automated, shortening cycles from days to hours and reducing analyst/committee time.
- **Run costs:** API LLM inference is usage-based and modest at treasury cadence; optimization and storage are negligible.

Simple ROI template:
Annual benefit ≈ |Notional reduced| × (trading cost + carry drag) + (hours saved × rate) − (AI + infra).

Governance, audit, and reproducibility

- Every run emits **run_summary.json** (constraints, selections, scenarios) to support ALCO/Internal Audit.
- Reproducibility is ensured through the use of fixed random seeds and explicit calibration windows.
- The AI proposal is **never** taken on faith: it is optimized, guardrail-checked, and stress-tested exactly like the traditional shelves.

Code availability and generality

- The codebase is **generic**, as it works with **any** treasury portfolio that conforms to the input schema (`treasury_portfolio.csv`) and the hedge universe format.
- Repository: **github.com/abalgir/AI-Treasury-Hedge-Optimization** (run the exact workflow, change inputs, and reproduce the tables).
- The results above reflect **one** portfolio and one calibration window; the **same code** adapts to other portfolios and market regimes.
