Pricing & commercial models under AI pressure

What I've built for law firms deciding which work to reprice for AI, and how to defend the price of the rest

General briefing · Alt-V Law

Derek Duffy, APAC Director, Alt-V

In short. Thirty-four per cent of law firms have updated pricing to reflect AI; the other sixty-six per cent are discovering it through margin leakage. The strategic question now is whether the firm leads the repricing conversation or has clients impose it. I've built a thinking layer (ten commercial models, four-category AI-impact matter segmentation, the Toby Brown model of structural margin erosion), three live tools (Pricing Engine, Panel Bid Economics, Panel Bid Builder), a seven-dimension pricing-maturity assessment for the firm itself, and an eight-to-twelve-week engagement shape with twelve named templates. A sample pack is included.

The sixty-six per cent wedge

Thirty-four per cent of firms have updated their pricing to reflect AI's effect on hours (BigHand 2025). Eighty-four per cent of firms offer AFAs; twenty-three per cent of actual revenue is priced under one - a gap that has not moved in two decades. Ninety per cent of legal spend is still hourly. Am Law 100 rates have crossed one thousand US dollars for mid-tier partners and two thousand for the top end; OGC-tracked rates rose nine per cent in 2025; Global 50 rates thirteen point five. The pressure on the firm shows up in realisation.

The uncomfortable arithmetic is Toby Brown's, modelled with Bruce MacEwen and Janet Stanton on the timesheets of ten large US firms and reported by Bloomberg Law on 16 May 2024. Take five per cent off partner hours and twenty per cent off non-partner hours, and revenue falls by thirteen per cent and margin by about seven per cent. Small changes in hours produce a structural fall in margin, and the fall can land before anyone sees it in the numbers. The ACC/Everlaw 2025 survey has fifty-two per cent of in-house teams using generative AI, sixty-four per cent expecting to reduce external reliance, and sixty-one per cent planning to push pricing changes to their panel firms. The pressure is live and it is not going to abate.

A firm that leads the conversation has a defensible answer ready for a predictable question; otherwise clients set the terms, matter type by matter type, through the panel-renewal cycle.

The four-category AI-impact segmentation

Not every matter type bends the same way. The segmentation Alt-V works with has four categories, and the right commercial model is different for each.

AI-high. Document review, due diligence, discovery - high-volume, high-standardisation, high-leverage, mature AI substitution. Expected hour reduction thirty to sixty per cent on affected matter types within two to three years. Pricing moves to **fixed, capped, or outcome-based** structures with the AI efficiency priced in; defending hourly here concedes the margin to competitors who have already made the move.

AI-medium. Contract review, legal drafting, legal research - moderate-volume, moderate-standardisation, moderate AI substitution. Expected hour reduction ten to twenty-five per cent; the margin case is for **hybrid** structures - base plus performance element, or capped with efficiency sharing - rather than full fixed.

© 2026 Derek Duffy. Licensed to Alt-V for authorised use.

AI-low. Complex litigation, M&A deal work, regulatory advice - bespoke, high-strategic-value, judgement-heavy, low AI substitution. Expected hour reduction under ten per cent, concentrated on the lowest-leverage support work. **Hourly** remains the honest default, and the pricing conversation centres on the value of the advice.

AI-hostile. Client-sensitive advice, trial advocacy, bet-the-company. AI touches none of the critical judgement. **Value-based** or **premium hourly** pricing holds; the firm's position is that this work was never a candidate for discount by efficiency.

The ten commercial models

The pricing taxonomy carries ten named models: Hourly, Fixed, Capped, Collared, AFA umbrella, Contingency / Success, Retainer / Subscription, Value-based, Portfolio / Panel, Hybrid. Each fits some matter types and breaks under others. The pricing conversation becomes manageable once the firm has a decision frame for which model goes where.

The two live tensions are between **Hourly** (which the Toby Brown numbers quietly erode) and **Fixed** (which transfers efficiency gains to the firm but asks for scope discipline the firm may not yet have); and between **Portfolio / Panel** structures (which win panel places but compress margin on the low end) and **Value-based** structures (which hold margin but require a client dialogue the firm may not yet be able to run).

The seven-dimension matter-type diagnostic

The Pricing Engine scores each matter type on seven dimensions: **scope variability**, **volume / repeatability**, **strategic value to client**, **risk profile**, **AI impact category** (one of the four above), **competitive pressure**, and **realisation**. The output is a tiered price band (low / target / high) plus a recommended commercial model keyed to the firm's AI-and-pricing readiness. For AI-high matter types with mid-realisation and intense competitive pressure, the model surfaces fixed-fee at a tiered band with AI-efficiency sharing built in; for AI-low matter types with high strategic value, it surfaces premium hourly.

The diagnostic structures the conversation the partnership would otherwise have in the hallway, one matter at a time, with no shared frame; the partner's judgement on the specific client still decides.

The seven-dimension pricing-maturity model

Underneath matter-type pricing sits the firm's capability to run pricing at all. Alt-V's pricing-maturity model (template T04 of the Pricing Review Methodology) reads a firm against seven dimensions: **pricing capability and team**, **commercial models used**, **data and analytics**, **governance and approval**, **partner behaviour and discipline**, **client-facing pricing dialogue**, and **AI-and-pricing readiness**. Each dimension lands on a Nascent / Emerging / Established / Advanced / Optimised anchor.

The directional rule is simple: a firm cannot run model-driven pricing from a Nascent data posture; a firm cannot sustain fixed-fee discipline from a Nascent partner-behaviour posture; a firm cannot have the AI-savings conversation with clients from a Nascent client-dialogue posture. The maturity assessment tells the firm what it can actually run, and that has to come before any choice of model means much.

The three tools

Pricing Engine. A per-matter-type pricing diagnostic. Firm selects the matter type (standard templates for commercial advisory, employment, litigation, M&A, real estate, etc.), enters cost and realisation data, selects dimension levels on the seven-dimension diagnostic. Output is a tiered price-band, a recommended commercial model, a realisation implication, and (in the gated full report) the four-point

© 2026 Derek Duffy. Licensed to Alt-V for authorised use.

client script for the “why aren’t you passing on the AI savings?” conversation. Free tier shows the price band; the full report runs through the governance checklist and the editable email template.

Panel Bid Economics. A three-year profit-and-loss simulator for a panel renewal. Firm enters its rate card, cost structure, target margin, panel discount percentage, matter mix, AI substitution scenario (conservative / central / aggressive), and working capital assumptions (realisation, write-off, WIP days, debtor days). Output is a three-year P&L across all three AI scenarios, a margin bridge from as-is delivery to panel-term delivery, a sensitivity table across AI scenarios and years, and a break-even analysis showing the panel discount or fixed fees at which year-1 margin hits zero.

Panel Bid Builder. A seven-dimension position-strength assessment against a specific panel brief. Firm identifies the buyer, the panel type, the term, the scope categories, the disclosed criteria, and its own position (incumbent, challenger, insurgent) across seven dimensions: technical capability · pricing readiness · resourcing · track record · relationship capital · DEI and ESG · value-add. Output is a strength / weakness map, a gap register (what the firm needs to improve before submission), and win-themes keyed to the specific panel brief.

The three tools compose. The Pricing Engine reads the firm’s capability on a matter type; the Panel Bid Economics stress-tests the firm’s commercial position on a specific panel; the Panel Bid Builder reads the firm’s position against the specific brief.

Fee Performance Analytics is a separate product that sits beside the three tools. It reads a firm’s own time, billing and matter history, shows how much of the work at standard rates became cash and where the rest went, and re-prices each matter type under 14 fee models with AI in or out. Two demonstrations on fictional firms are open at tools.derekduffy.com/fee-performance/ (a UK firm) and tools.derekduffy.com/fee-performance/au/ (an Australian firm).

The four-point client script

When a client asks “why aren’t you passing on the AI savings?”, the firm needs a defensible answer in four points. The script already published in the Pricing Engine report runs this way:

1. **Acknowledge the shift explicitly.** AI has moved the economics on specific matter types; the firm has measured it and has a position.
2. **Name which matter types have moved and by how much.** Specificity defuses the general claim. The firm’s AI-high categories carry a quantified efficiency; the AI-low categories have not moved and will not.
3. **Name what has been priced in.** The firm has already adjusted on the matter types that have moved - through fixed fees, through capped structures, through efficiency-sharing clauses. Showing the recent adjustments pre-empts the general discount demand.
4. **Name the trade-off being proposed.** Full pass-through of efficiency means the firm absorbs the measurement cost, the AI governance cost, and the pilot risk. A shared-savings structure keeps the firm investing. A rate cut without that frame transfers all three to the client - which the client does not actually want when the governance frame is understood.

The script works because it moves the conversation from “give us a discount” to “how do we share the gains while keeping the capability investment going”, which is where the margin is protected.

How the engagement runs

The Pricing Review Methodology is eight to twelve weeks, six phases, twelve templates. **Scoping** (weeks 1–2) - proposal, data request pack, interview guide. **Diagnostic** (weeks 2–4) - Pricing Maturity Assessment, matter-type portfolio worksheet. **Workshops** (weeks 4–7) - current-state half-day, options-design half-day. **Business case** (weeks 6–9) - structured business case, benefits realisation

© 2026 Derek Duffy. Licensed to Alt-V for authorised use.

model. **Board** (weeks 8–10) - board/partnership paper. **Implementation** (weeks 10–12+) - repricing cadence tracker, partner education deck.

Lighter shapes compose from the same spine. A **Boutique Pricing Health Check** is three to four weeks against the maturity assessment alone. A **Single-practice-group review** is four to six weeks on one practice group's matter types. A **Comprehensive Pricing Review** is twelve to sixteen weeks across all matter types with a per-practice-group workshop series. Scope and investment are set with the firm once the diagnostic has run and the engagement shape is confirmed.

Pairings with the other pillars

Three pairings matter. First, **Partner Pay**. A firm whose pay issue is K-shape downward drift cannot fix pay in a rem review alone - re-slicing a shrinking pool makes no one happier. The rem review has to run paired with a pricing engagement; otherwise it is theatre. Second, **Operations and the LOMM**. Matter-level economics is Pricing's data layer. A firm at LOMM Level 2 cannot run model-driven pricing because the data layer does not exist to drive the model. The pricing review has to meet the firm at the operating maturity it actually has. Third, **M&A**. A post-merger firm with two legacy rate cards has a pricing-harmonisation problem that the Pricing Engine solves directly; the T04 Pricing Maturity Assessment and the T05 matter-type portfolio worksheet are the usual Phase 1 deliverables.

Where this tends to be useful

Three triggers come up most often.

1. **Before a panel renewal.** Running the Panel Bid Economics six to nine months before a renewal shifts the conversation from "what discount do we offer" to "what is the deliverable commercial position at the margin the firm needs." The sample pack carries this scenario.
2. **After an AI pilot has produced its efficiency reading.** The pilot numbers only become useful once they feed the pricing conversation. The Pricing Engine is the tool that converts a technical efficiency reading into a commercial-model recommendation.
3. **When the client-side pricing dialogue has started asking the AI question.** The four-point client script is designed for this moment, and a firm with the script in hand can hold its margin through that conversation.

What is the next matter type at your firm that is quietly being discounted on an unexamined assumption about AI - and how would you know?

Derek Duffy

APAC Director, Alt-V

The sample Pricing Engine output, Panel Bid Economics P&L, Panel Bid Builder position assessment, and Pricing Maturity Assessment that accompany this note use an invented firm, people and figures throughout. The tools and the methodology are management-consulting instruments; they do not substitute for a scoped engagement. The Pricing Engine's forthcoming refresh adds consumption-priced AI modes (seat / consumption / seat-plus) to Block D of the tool.

© 2026 Derek Duffy. Licensed to Alt-V for authorised use.