

The MiAI Law Civil Benchmark for Australian Legal Research AI

Fifty questions across fifteen categories, with answer-key targets and a scoring rubric for evaluating Australian legal research AI tools

Purpose and design

This benchmark provides a structured methodology for evaluating Australian legal research AI tools. It comprises 50 questions distributed across 15 categories covering the workflows that arise in Australian legal practice: doctrinal research, statutory interpretation, point-in-time and current legislation research, statutory amendment tracking, compliance triage, procedural research, summarisation, audit of legal writing, drafting of appellate grounds, audit and pre-drafting research for pleadings, contract review across multiple deliverable modes, and Word plug-in integration.

The categories were chosen on the basis that they reflect the actual surface area of work that a competent Australian legal research AI should support. Substantive coverage spans contract, equity/fiduciary, ACL, Corporations Act, insurance, succession / estates, civil procedure, financial services, infrastructure / construction, and franchising. Authorities range across HCA, FCAFC, FCA, every state and territory supreme court, and every Civil and Administrative Tribunal.

Each question is paired with an answer-key target identifying the substantive content a correct answer should contain, and an evaluation note identifying the dimension the question is designed to test.

Jurisdictional coverage

An Australian legal research AI should cover every Australian state and territory - High Court, Federal Court, all state and territory supreme courts and courts of appeal, the intermediate appellate and trial divisions of every jurisdiction, and every Civil and Administrative Tribunal across Australia: NCAT (NSW), VCAT (Vic), QCAT (Qld), SACAT (SA), WASAT (WA), ACAT (ACT), NTCAT (NT) and TASCAT (Tas). The Federal Court tribunals - the Administrative Review Tribunal (which replaced the Administrative Appeals Tribunal in October 2024 under the Administrative Review Tribunal Act 2024 (Cth)) and the Fair Work Commission - should also be covered. The benchmark assumes coverage across the last 30 years and distributes questions across jurisdictions accordingly.

This matters for the benchmark because coverage gaps can be measured. A tool whose database does not extend to (for example) VCAT, WASC or TASSC will not return the expected answer to questions targeting those jurisdictions, regardless of how well the tool performs in jurisdictions it does cover. Per-question scores therefore reflect both substantive capability and coverage breadth, and the marker should record which is the operative limit when a target is not met.

Pinpoint accuracy as a scoring dimension

Pinpoint accuracy is not a category of question in this benchmark. It is a scoring dimension that runs through every answer. Every research answer in Australian legal practice carries a pinpoint on every substantive proposition - paragraph, page, or footnote, depending on what the source provides. The pinpoint is the verification mechanism that makes the answer defensible at file-handover. The marker checks pinpoint accuracy on every answer, in every category.

Different tools achieve pinpoint accuracy by different means. Some perform full-text retrieval against the judgment or statute and quote the source verbatim; others maintain a structured database of extracted propositions (ratio and obiter) with the pinpoint references attached at the point of extraction.

Both approaches are valid. The marker scores the accuracy of the pinpoint against the source, not the means of its production.

A correct proposition with a correct pinpoint scores full pinpoint marks regardless of whether the answer quotes the source verbatim. A verbatim quotation without a pinpoint, or with a wrong pinpoint, does not satisfy the dimension.

Legislation functions: point-in-time and amendment tracking

The benchmark distinguishes two legislation functionalities: point-in-time research (Category D) and statutory amendment tracking (Category E). They are tested separately because they produce different output shapes and engage different evaluation criteria, even where they share an underlying retrieval architecture.

Point-in-time research returns the historical text of a provision as at a specified date. A capable tool will explicitly disclose where it has not searched for subsequent amendments - the answer is the historical text and nothing further. Tools that silently return the current version when asked for a historical one, or that extend the answer beyond the date specified without disclosure, do not satisfy the target. Honest scope-limiting is a scoring dimension.

Statutory amendment tracking traces how an amending Act has modified a provision. A capable answer retrieves the pre-amendment consolidation, the post-amendment consolidation, and the amending Act itself, and identifies the operative amending instruction, the commencement date, and the commencement rule. Where the user has supplied an incorrect section or subsection reference and the amending Act in fact operates on a different provision, a capable tool will detect and correct the error. A correctly-reached 'no operative amendment' finding is itself a defensible answer where the amending Act left the substantive regime unchanged (for example, where only a section heading was reformatted).

Category overview

Cat	Category	Qs	What it tests
A	Doctrinal analysis, development, hierarchy, good-law status	4	Reasoning from primary sources; tracking doctrine across jurisdictions; precedent and binding effect; subsequent-treatment checks. One warm-up case-finder retained.
B	Multi-case synthesis	4	Integration across 4–9 named authorities with pinpointed propositions from each
C	Recent / post-cutoff authority	4	Tests currency; refresh quarterly
D	Point-in-time legislation research	3	Historical text retrieval with scope-limiting disclosure
E	Statutory amendment tracking	3	Pre/post text plus amending Act plus commencement
F	Compliance / due-diligence triage	3	Statute list → yes/no table → per-statute analysis with section refs
G	Negative-finding questions	3	'No authority' answers without invention
H	Statutory-architecture questions	3	Definitions / application / attribution / remedy gateways before substantive
I	Procedural legislation research	3	Rules + practice notes + bench books cross-referenced against cases

Cat	Category	Qs	What it tests
J	LawCheck - auditing legal writing	3	Per-proposition audit with snapshot summary and fabrication audit
K	AppealCheck - drafting grounds	3	Draft notice of appeal from a judgment
L	PleadingCheck + Pleadings research	3	Element mapping for existing pleadings AND for pre-drafting research
M	Document summarisation	4	Cases (weighted) plus pleadings, affidavits, contracts
N	Contract Review (Lite/Deep/Negotiation Pack) & Audit	4	Four discrete deliverables against the same document
O	Word plug-in integration	3	Pleadings Review and Contracts Review plug-ins inside Microsoft Word

Total: 50 questions across 15 categories. Authorities range across HCA, FCAFC, FCA, every state and territory supreme court (NSWCA, NSWSC, QCA, QSC, VSCA, VSC, WASCA, WASC, SASCFC, SASC, TASSC, ACTSC, NTSC) and every Civil and Administrative Tribunal (NCAT, VCAT, QCAT, SACAT, WASAT, ACAT, NTCAT, TASCAT), together with the Federal Court tribunals (ART, FWC).

A. Doctrinal analysis, cross-jurisdictional development, hierarchy and good-law status

Locating a case from facts is a basic retrieval task that most modern tools can complete. The harder and more diagnostic capabilities are: (i) explaining a doctrine from primary sources; (ii) tracking how a principle has developed across jurisdictions over 20–30 years; (iii) applying the rules of precedent correctly to identify what binds what; and (iv) checking whether a case is still good law against subsequent treatment. One warm-up case-finding question is retained as a baseline; the remaining three questions test the differentiating capabilities.

#	Question	Answer-key target / Evaluation note
A1	Warm-up - case-finding. What is the name of the Full Federal Court case in which Mercedes-Benz franchisees challenged Mercedes-Benz Australia's transition from a dealership model to an agency model on grounds of statutory unconscionability under s 21 ACL and breach of good faith under the Franchising Code?	<p>Target. AHG WA (2015) Pty Ltd v Mercedes-Benz Australia/Pacific Pty Ltd [2025] FCAFC 86; Moshinsky, Bromwich and Anderson JJ; appeal dismissed; primary citation linked to judgment.</p> <p>Evaluation note. <i>Baseline retrieval question. The decision was handed down in July 2025; tools dependent on training data with a 2024 or earlier cutoff will not return the correct case.</i></p>
A2	Doctrinal analysis. Explain the doctrine of statutory unconscionability under s 21 ACL - its evaluative-against-statutory-values character, the non-exhaustive list of relevant factors in s 22, the role of community standards, the proof of disadvantage, the relationship with equitable unconscionable conduct, and the limits the doctrine has been given by intermediate appellate courts. Reason from primary sources (statute and case law), not commentary. Address how the doctrine differs from misleading or deceptive conduct under s 18 and from the unfair contract terms regime in ss 23–28.	<p>Target. Doctrinal explanation built from primary sources: text of s 21 and s 22 ACL with the s 22 factors enumerated; treatment of Productivity Partners v ACCC [2024] HCA 27 on community standards permeating the statutory test; ASIC v Kobelt [2019] HCA 18 on the meaning of unconscionability; ACCC v Lux Distributors [2013] FCAFC 90 on the evaluative judgment against statutory norms; ACCC v Quantum Housing [2021] FCAFC 40 on disadvantage not being a precondition; differentiation from s 18 (misleading conduct, not unconscionability) and ss 23–28 (regime applying to contract terms ex ante, not conduct). Every proposition pinpointed to its primary source.</p> <p>Evaluation note. <i>Evaluates whether the tool reasons from primary sources or paraphrases commentary. A correct answer cites the statutory text first and the controlling appellate authority second; an answer drawn from secondary materials typically loses the statutory-norms framing that controls the doctrine.</i></p>
A3	Cross-jurisdictional development. Trace the development of the implied obligation of good faith in the performance of commercial contracts in Australia from the early appellate authority through to the current position. Address: the early HCA reservations, the position taken by Priestley JA in Renard Constructions, the position taken by the NSWCA in Hughes Aircraft and Burger King, the position taken by the Full Federal Court, the divergence among intermediate appellate courts, the position in Queensland, Victoria and Western Australia, the High Court's continued non-resolution of the question, and any 2024 or 2025 intermediate appellate decision applying or refining the doctrine.	<p>Target. Cross-jurisdictional development trace: Renard Constructions (Me) Pty Ltd v Minister for Public Works (1992) 26 NSWLR 234 (Priestley JA); Hughes Aircraft Systems International v Airservices Australia (1997) 76 FCR 151; Burger King Corporation v Hungry Jack's Pty Ltd [2001] NSWCA 187; subsequent FCAFC, QCA, VSCA and WASCA treatment; the High Court's repeated declining to resolve the question (recognised as remaining open at HCA level); identification of any 2024–2025 intermediate appellate decision. Every step pinpointed; jurisdictional divergence made explicit.</p> <p>Evaluation note. <i>Evaluates whether the tool can integrate doctrinal development across multiple jurisdictions over a 30-year period in a single</i></p>

#	Question	Answer-key target / Evaluation note
		<i>answer. Tools that return only the leading case or limit themselves to a single jurisdiction do not satisfy the target.</i>
A4	<p>Hierarchy, binding effect and good-law status. A Victorian Supreme Court single-judge decision from 2019 followed an English Court of Appeal decision on the construction of a commercial guarantee. A 2023 NSW Court of Appeal decision then held the opposite to the English Court of Appeal, expressly preferring an earlier High Court of Australia decision. (a) For a Federal Court single judge sitting in Victoria in 2026 on a similar guarantee, which of the above decisions are binding and which are persuasive only, and why? (b) Is the 2019 VSC decision still good law? Identify any subsequent treatment (followed / distinguished / doubted / overruled / not followed) and provide pinpoints. (c) How should a Tasmanian Supreme Court single judge resolve the same issue today, given the hierarchy of authority?</p>	<p>Target. (a) HCA decision is binding on FCA single judge sitting in Victoria; 2023 NSWCA is persuasive (not binding because not a Victorian appellate decision); 2019 VSC is persuasive (not binding on FCA single judge); English CA is persuasive only. (b) Subsequent treatment of the 2019 VSC decision identified and reported with pinpoints to citing decisions. (c) Tas SC single judge similarly not bound by NSWCA or VSC; bound by HCA; in the absence of TASSC appellate authority, should follow the most-recent appellate authority that itself follows the HCA position (likely the 2023 NSWCA decision). Reasoning grounded in the rules of precedent, not in analogy.</p> <p>Evaluation note. <i>Evaluates three capabilities in one question: structural reasoning on court hierarchy and binding effect, good-law verification via subsequent-treatment checking, and cross-jurisdictional advisory application. Each capability is independently scoreable.</i></p>

B. Multi-case synthesis

Synthesis of multiple named authorities into a single coherent analysis. The user supplies four or more cases and the tool integrates them with pinpoints to the propositions drawn from each. Tests integration across long time periods and across jurisdictions.

#	Question	Answer-key target / Evaluation note
B1	By reference to <i>Sergent v Glass</i> [2017] NSWSC 1446, <i>Re Kordos</i> [2023] VSC 14, <i>El Sayed v El Hawach</i> [2015] NSWCA 26, <i>Hewitt v Gardner</i> [2009] NSWSC 705, <i>Ramage v Waclaw</i> (1988) 12 NSWLR 84, <i>Cong v Shen (No 3)</i> [2021] NSWSC 947, <i>Scallan v Scallan</i> [2001] NSWSC 1129, <i>Powell v Oorschot</i> [2024] WASC 401 and <i>Re Ciantar</i> [2022] VSC 116, what are the limits on a beneficiary bringing a claim to impugn inter vivos transactions of a deceased said to involve unconscionable conduct? Must such claims first be investigated by the administrator? If the administrator declines, what limits apply? Must the administrator be joined? What is meant by 'special or exceptional circumstances'?	<p>Target. Five-part synthesis with pinpointed propositions from each named case; uniform treatment of NSW, VSC and WASC strands.</p> <p>Evaluation note. <i>Nine cases across three jurisdictions. Evaluates whether the tool can sustain pinpoint accuracy and doctrinal coherence over a high-volume synthesis.</i></p>
B2	By reference to <i>United Dominions Corporation Ltd v Brian Pty Ltd</i> [1985] HCA 49, <i>Hospital Products Ltd v United States Surgical Corporation</i> [1984] HCA 64, <i>Eaton v Rare Nominees Pty Ltd</i> [2019] QCA 190 and <i>Waller Projects Pty Ltd v F.W. Estate Pty Ltd</i> [2025] QSC 16, when will a fiduciary relationship arise between joint venturers under Queensland law, and when will an express contractual exclusion of fiduciary duty be effective?	<p>Target. Synthesis with pinpoints; treatment of the 'fiduciary cannot be superimposed on contract' principle (<i>Eaton</i> at [61], [82]); <i>Mason J's</i> indicia from <i>Hospital Products</i> at [68]–[69], [96]–[97]; <i>Kelly J's</i> 2025 restatement at [266]–[268].</p> <p>Evaluation note. <i>Evaluates handling of the express-exclusion doctrine in combination with a recent QSC restatement. Each authority must be pinpointed to its operative passage.</i></p>
B3	By reference to <i>Personnel Contracting Pty Ltd v CFMMEU</i> [2022] HCA 1, <i>ZG Operations Australia Pty Ltd v Jamsek</i> [2022] HCA 2, and any Full Federal Court or Fair Work Commission Full Bench decision since 2024 applying or distinguishing them, set out the current test for distinguishing an employee from an independent contractor. Address the operation of any relevant amendments to the Fair Work Act 2009 (Cth) made since 2023.	<p>Target. Pinpointed synthesis; treatment of <i>Kiefel CJ, Keane and Edelman JJ</i> at [58]–[61] in <i>Personnel Contracting</i>; identification of the Closing Loopholes amendments and their effect on the common law test for purposes of the Fair Work Act.</p> <p>Evaluation note. <i>Evaluates whether the tool integrates case law with a recent statutory override. Tools that handle only the case law or only the statute do not satisfy the target.</i></p>
B4	By reference to <i>Commercial Bank of Australia Ltd v Amadio</i> (1983) 151 CLR 447, <i>Bridgewater v Leahy</i> [1998] HCA 66, <i>Kakavas v Crown Melbourne Ltd</i> [2013] HCA 25, <i>Stubbings v Jams 2 Pty Ltd</i> [2022] HCA 6 and any 2024 or 2025 intermediate appellate decision applying or refining the doctrine, set out the current test for unconscionable conduct in equity, identifying how the test interacts with the statutory unconscionability provisions in ss 20 and 21 ACL.	<p>Target. Synthesis with pinpointed propositions across five HCA decisions; treatment of <i>Kakavas's</i> narrowing and <i>Stubbings's</i> reaffirmation; identification of any 2024–2025 intermediate appellate authority; equity / statute interaction articulated in <i>ASIC v Kobelt</i> and <i>Productivity Partners v ACCC</i>.</p> <p>Evaluation note. <i>Evaluates synthesis across five HCA decisions spanning four decades, plus the equity / statute interaction. The integration is doctrinally subtle.</i></p>

C. Recent / post-cutoff authority

Decisions handed down within the last 12–18 months. Categories of this kind are typically refreshed quarterly to maintain currency against changes in the training-data cutoff of competitor tools.

#	Question	Answer-key target / Evaluation note
C1	What did the Full Federal Court hold in 2025 about the application of the statutory test for unconscionability under s 21 ACL to franchising arrangements involving a transition from a dealership to an agency model?	<p>Target. AHG WA (2015) Pty Ltd v Mercedes-Benz Australia/Pacific Pty Ltd [2025] FCAFC 86 at [122]–[206]; evaluative judgment against statutory values, not abstract community standards; appeal dismissed.</p> <p>Evaluation note. July 2025 decision. Evaluates whether the tool's database extends to recent appellate authority.</p>
C2	What is the most recent Full Federal Court or High Court decision (2024 or 2025) on the test for statutory unconscionability in the context of consumer credit or financial services, and how does it interact with the holding in Productivity Partners Pty Ltd v ACCC [2024] HCA 27?	<p>Target. Identification of Productivity Partners and any 2025 FCAFC application; treatment of the HCA's emphasis on community standards as informing s 21.</p> <p>Evaluation note. Evaluates integration of a known 2024 HCA decision with whatever 2025 applications exist.</p>
C3	What did the Queensland Supreme Court hold in 2026 about the elements of proprietary estoppel and the pleading requirements for striking out an inadequately particularised claim?	<p>Target. Tri-Star Petroleum Company v Australia Pacific LNG Pty Ltd [2026] QSC 18 at [60]–[65], [140]–[147], [152]–[164].</p> <p>Evaluation note. Refresh-quarterly question; 2026 QSC decision. Evaluates currency at the most-recent end of the spectrum.</p>
C4	Identify the most recent Full Federal Court decision (2025 or 2026) on the operation of the unfair contract terms regime in ss 23–28 of the ACL following the introduction of pecuniary penalties by the Treasury Laws Amendment (More Competition, Better Prices) Act 2022 (Cth). Address how the Court treats the threshold for a 'small business contract' as amended.	<p>Target. Most recent FCAFC decision applying the post-November 2023 regime; treatment of the expanded 'small business contract' threshold (employees up to 100 or annual turnover up to \$10m); pinpoint to operative paragraphs and any penalty quantum.</p> <p>Evaluation note. Evaluates the intersection of a 2022 amending Act, a 2023 commencement, and any 2025/2026 application.</p>

D. Point-in-time legislation research

What did a provision provide on a specific historical date? The relevant function returns the consolidation as at the date specified. A capable tool will explicitly disclose where it has not searched for subsequent amendments - point-in-time research returns the historical text and the user is on notice that the current text may differ. Statutory amendment tracking is tested separately in Category E.

#	Question	Answer-key target / Evaluation note
D1	What did s 588FA of the Corporations Act 2001 (Cth) provide on 1 July 2018? Reproduce the verbatim text of the provision as in force on that date.	<p>Target. Verbatim text of s 588FA(1)–(3) as at the 1 July 2018 compilation, with the section heading 'Unfair preferences' as it stood. Source URL: legislation.gov.au consolidation dated 2018-07-01. The answer should not compare to the current text. A capable answer includes a scope-limiting disclosure indicating that the tool has not separately searched for subsequent amendments.</p> <p>Evaluation note. <i>Evaluates pure point-in-time retrieval. Tools that silently return the current version, or that extend the answer beyond the date specified without disclosure, do not satisfy the target.</i></p>
D2	What did s 18 of the Australian Consumer Law (Schedule 2 to the Competition and Consumer Act 2010 (Cth)) provide on the date of its original commencement (1 January 2011)?	<p>Target. Verbatim text of s 18 as at 1 January 2011, with section heading and operative subsection. Source URL with the appropriate compilation date stamp. Scope-limiting disclosure present.</p> <p>Evaluation note. <i>Evaluates retrieval at the commencement date of the ACL - a date 15 years in the past, outside the reliable training-data window for verbatim legislative text.</i></p>
D3	What did s 13 of the Insurance Contracts Act 1984 (Cth) - the duty of utmost good faith provision - provide on 1 March 2014? Reproduce the verbatim text as in force on that date.	<p>Target. Verbatim text of s 13 as at 1 March 2014; source URL with compilation date. Scope-limiting disclosure present.</p> <p>Evaluation note. <i>Pre-Insurance Contracts Amendment Act 2013 vs post: evaluates whether the tool retrieves the correct historical compilation rather than a more recent version.</i></p>

E. Statutory amendment tracking

How did an amending Act modify a provision? The relevant function retrieves the pre-amendment consolidation, the post-amendment consolidation, and the amending Act itself, and identifies the operative amending instruction, the commencement date, and the commencement rule. A capable tool will also detect and silently correct user-supplied section or subsection errors where the amending Act in fact operates on a different provision than the user identified.

#	Question	Answer-key target / Evaluation note
E1	How have sections 916F(1) and (2) of the Corporations Act 2001 (Cth) been amended effective in December 2025 by the Treasury Laws Amendment (Strengthening Financial Systems and Other Measures) Act 2025? Provide pre- and post-amendment text, the operative amending instruction verbatim, the commencement rule and the actual commencement date.	<p>Target. The amending Act operates on s 916F(1) and (3), not (1) and (2); a capable answer corrects the user-supplied subsection reference. Pre-amendment text (15 / 10 business days); post-amendment text (30 / 30 business days); Schedule 4 Part 1 Division 4 Items 23 and 24 of the amending Act reproduced verbatim; commencement rule 'day after Royal Assent'; commencement date 5 December 2025.</p> <p>Evaluation note. <i>Evaluates the tool's ability to detect a user-supplied subsection error and to engage the amendment-resolution gate across both consolidations of the base Act and the amending Act itself.</i></p>
E2	What did s 588FA of the Corporations Act 2001 (Cth) provide on 1 July 2018? Identify any amending Acts that have modified s 588FA between 1 July 2018 and today, and for each amending Act identify the commencement date and the operative amending instruction.	<p>Target. (1) Verbatim text as at 1 July 2018. (2) Amendment-resolution gate engaged across both consolidation-comparison and amending-Act-application pathways. (3) Identification of Treasury Laws Amendment (Modernising Business Communications and Other Measures) Act 2023 (Act No. 69 of 2023) as the relevant amending instrument; identification that the only change touched the section heading ('Unfair preferences' → 'Meaning of unfair preference') and that the operative subsections are unchanged. (4) A defensible practitioner-grade answer: 'no operative amending instruction affecting the substantive regime'.</p> <p>Evaluation note. <i>Evaluates handling of a composite point-in-time + amendment-tracking query. Tools that return only the current text and assert it has not changed, or that fabricate an amendment that does not exist, do not satisfy the target.</i></p>
E3	What changes did the Fair Work Legislation Amendment (Closing Loopholes No. 2) Act 2024 (Cth) make to the definition of 'employee' and 'employer' in the Fair Work Act 2009 (Cth)? Identify the new s 15AA provision verbatim and its commencement date.	<p>Target. s 15AA Fair Work Act verbatim text; commencement 26 August 2024; treatment of the statutory definition's interaction with CFMEU v Personnel Contracting Pty Ltd [2022] HCA 1 / ZG Operations Australia Pty Ltd v Jamsek [2022] HCA 2.</p> <p>Evaluation note. <i>Evaluates the interaction between case law and overriding amendment - the same kind of integration tested in B3 but viewed from the amending-Act side.</i></p>

F. Compliance / due-diligence triage (statute list → yes/no table)

The user supplies a list of statutes and asks which impose obligations specific to a class of actor (contractors, directors, APP entities, etc.). The expected output is a yes/no screening table with section references followed by a per-statute analysis identifying the actor class each statute regulates.

#	Question	Answer-key target / Evaluation note
F1	Below is a list of potentially relevant laws to the construction of data centres in NSW. Advise (yes/no, with section references) whether any impose obligations specific to contractors (head contractors, subcontractors, principal contractors): Foreign Acquisitions and Takeovers Act 1975 (Cth); Security of Critical Infrastructure Act 2018 (Cth); Privacy Act 1988 (Cth); Telecommunications Act 1997 (Cth); Climate Change Act 2022 (Cth); Home Building Act 1989 (NSW); Building and Construction Industry Security of Payment Act 1999 (NSW); Work Health and Safety Act 2011 (NSW); Privacy and Personal Information Protection Act 1998 (NSW).	<p>Target. Yes/No table with section refs: HBA (s 4, s 5, s 13, s 18B), SOPA NSW (s 7, s 8, s 26A–C), WHS Act (s 19, s 14, s 16, s 46, s 244) → yes; Cth privacy / SOCI / FATA / telco / climate → no (apply to other actor classes). Plus per-statute rationale.</p> <p>Evaluation note. <i>Evaluates whether the tool can produce a single-pass yes/no table over multiple statutes with section pinpoints and consistent actor-class analysis.</i></p>
F2	A NSW principal contractor is taking over delivery of a partially-built transport infrastructure project. Identify, with yes/no flags and section references, which of the following impose direct obligations on the incoming principal contractor as distinct from the prior contractor, the head client, or subcontractors: Work Health and Safety Act 2011 (NSW); Environmental Planning and Assessment Act 1979 (NSW); Heavy Vehicle National Law (NSW); Rail Safety National Law (NSW); Building and Construction Industry Security of Payment Act 1999 (NSW); Workers Compensation Act 1987 (NSW); the Building Code 2016 (issued under the Building and Construction Industry (Improving Productivity) Act 2016 (Cth)).	<p>Target. Yes/No table; identification that the Building and Construction Industry (Improving Productivity) Act 2016 (Cth) was repealed by the Building and Construction Industry (Improving Productivity) Repeal Act 2022 (Cth), with the consequence that obligations under the Building Code 2016 are no longer in force.</p> <p>Evaluation note. <i>Evaluates detection of a repealed-statute reference. The repeal post-dates many tools' training data cutoffs.</i></p>
F3	A fintech is launching a buy-now-pay-later product targeted at consumers in Australia. Identify, with yes/no flags and section references, which of the following impose obligations specific to the credit provider as distinct from the merchant, the consumer, or the broker: National Consumer Credit Protection Act 2009 (Cth); the National Credit Code (Sch 1); Anti-Money Laundering and Counter-Terrorism Financing Act 2006 (Cth); Privacy Act 1988 (Cth); Australian Securities and Investments Commission Act 2001 (Cth) Pt 2 Div 2; any current ASIC product intervention orders affecting BNPL; Treasury Laws Amendment (Responsible Buy Now Pay Later and Other Measures) Act 2024 (Cth).	<p>Target. Yes/No table with section refs; identification that the 2024 BNPL amending Act brings BNPL within the NCCP regime; cross-reference to any current ASIC product intervention orders; actor-class distinction between credit provider, merchant and broker maintained.</p> <p>Evaluation note. <i>Evaluates triage across a complex regulatory landscape with a recent amending Act and current regulator instruments.</i></p>

G. Negative-finding questions

The correct answer is 'no controlling authority' or 'the position is unsettled'. Evaluates whether the tool acknowledges the limit of the available authority rather than fabricating a citation.

#	Question	Answer-key target / Evaluation note
G1	Is there controlling High Court authority on whether a contractual discretion is constrained by an implied obligation of good faith? Identify the authority by name and pinpoint, or state that there is none.	<p>Target. No controlling HCA authority; the position remains unsettled at HCA level; intermediate-court divergence (notably Burger King v Hungry Jack's [2001] NSWCA 187 vs more recent FCAFC authority). The expected answer acknowledges this rather than naming an HCA case.</p> <p>Evaluation note. <i>Evaluates whether the tool will fabricate a controlling authority where none exists.</i></p>
G2	Is there controlling Full Federal Court authority on whether an algorithmic pricing system that produces collusive outcomes (without human intent) contravenes Pt IV of the Competition and Consumer Act 2010 (Cth)? Identify the authority or state that there is none.	<p>Target. No controlling FCAFC authority.</p> <p>Evaluation note. <i>Cutting-edge area of competition law where invented authority is a foreseeable risk.</i></p>
G3	Is there High Court authority on whether a parent company can be vicariously liable for the misleading conduct of its overseas subsidiary's directors in Australia under s 18 ACL? Identify the authority or state that there is none, and explain how the question should be answered under the existing statutory architecture (ss 5, 131, 139B, 139C).	<p>Target. No directly-on-point HCA authority; the question turns on application of attribution (ss 139B/C) and accessorial limbs. The expected answer articulates the statutory architecture without inventing a case.</p> <p>Evaluation note. <i>Evaluates the distinction between 'no case' and 'no statutory answer'.</i></p>

H. Statutory-architecture questions

Questions whose answer requires navigating the structural provisions (definitions, application, attribution, remedy gateways) before reaching the substantive prohibition. Evaluates whether the tool treats statutory text as architecture rather than as a collection of substantive prohibitions to be addressed in isolation.

#	Question	Answer-key target / Evaluation note
H1	Who is a proper defendant for a claim under ss 18 and 21 of the ACL? In what circumstances can an individual director or employee be sued, and how does the answer differ between a damages claim under s 236 and a pecuniary penalty claim under s 224?	<p>Target. Architecture-aware answer: 'person' in ss 18 and 21; application via s 131; attribution via ss 139B/C; 'involved in' via s 2 ACL (knowingly concerned / aided / abetted / party to / induced); remedy gateways - s 236 for damages, s 224 for penalty (does not extend to Pt 2-1, so no penalty against s 18 accessories, but extends to Pt 2-2 including s 21), s 232 for injunctions; <i>Yorke v Lucas</i> knowledge requirement.</p> <p>Evaluation note. <i>Evaluates whether the tool addresses application, attribution and remedy provisions before reaching the substantive prohibition.</i></p>
H2	How does the application provision in s 131 of the Competition and Consumer Act 2010 (Cth) interact with the attribution provisions in ss 139B and 139C and the extraterritorial application provision in s 5 in determining whether a foreign individual can be held liable for misleading conduct directed into Australia?	<p>Target. Architecture answer integrating extraterritorial application (s 5), s 131 application of Sch 2, and the attribution machinery (ss 139B for body corporate, 139C for non-body-corporate persons within actual/apparent authority), with limits arising from each.</p> <p>Evaluation note. <i>Evaluates whether the tool integrates four structural provisions rather than addressing the substantive question (s 18 liability) in isolation.</i></p>
H3	How does the operation of s 224(2) ACL (pecuniary penalty maximums) interact with the turnover-based formula introduced by the Treasury Laws Amendment (More Competition, Better Prices) Act 2022 (Cth)? Identify the four limbs of the current maximum and the date the formula commenced.	<p>Target. Four-limb maximum (greatest of: \$50m fixed amount; 3× value of benefit; 30% of adjusted turnover during the period; default to fixed amount where benefit/turnover unknown); commencement 9 November 2022; identification of the correct amending Act.</p> <p>Evaluation note. <i>Evaluates the tool's ability to read the current consolidated text alongside the amending Act that introduced the four-limb formula.</i></p>

I. Procedural legislation research

Distinct from substantive research. Procedural research draws on procedural legislation, court rules, tribunal rules, practice notes, and bench books, cross-referenced against the cases applying them. Time limits, leave gateways, joinder rules, security-for-costs thresholds. Highly rule-specific and frequently updated.

#	Question	Answer-key target / Evaluation note
11	What is the time limit and the source rule for filing a notice of intention to appeal from a NSW District Court decision to the NSW Court of Appeal? Cross-reference to any cases applying the rule on extension of time and the standard the Court applies.	<p>Target. Uniform Civil Procedure Rules 2005 (NSW) - specific rule (typically r 51.8/r 51.9 area); 28-day limit; cross-reference to Supreme Court Act 1970 (NSW); cases applying the extension-of-time discretion (e.g. Tomko v Palasty (No 2) [2007] NSWCA 369).</p> <p>Evaluation note. <i>Evaluates retrieval of the rule plus the cases applying it - the case cross-reference is the diagnostic element.</i></p>
12	Under the Victorian Civil and Administrative Tribunal Act 1998 (Vic), what is the procedure and time limit for an application for leave to appeal from a decision of VCAT to the Supreme Court of Victoria on a question of law? Identify the gateway, the source rule, the relevant VCAT practice note, and any VSCA decision applying the gateway.	<p>Target. Section 148 VCAT Act 1998 (Vic); 28-day limit for leave application; error of law gateway; cross-reference to Supreme Court (General Civil Procedure) Rules 2015 (Vic) on form of appeal; identification of relevant VCAT practice note; VSCA decisions applying the leave threshold.</p> <p>Evaluation note. <i>Evaluates handling of a tribunal-to-superior-court appeal pathway with rules, practice notes, and case law in a single answer.</i></p>
13	Under the Supreme Court (General Civil Procedure) Rules 2015 (Vic), what is the rule governing security for costs against an impecunious corporate plaintiff? Identify any Court of Appeal decision applying the rule and the factors the Court considers, with reference to the relevant bench book or practice note.	<p>Target. Specific rule reference (r 62.02 or current equivalent) and cross-reference to Corporations Act s 1335; VSCA decisions applying the rule; treatment of the discretionary factors (prima facie case, impecuniosity caused by the defendant, delay in application, etc.); any relevant bench book guidance.</p> <p>Evaluation note. <i>Multi-source procedural synthesis - case law plus statute plus rule plus bench book.</i></p>

J. LawCheck - auditing legal writing

A piece of legal correspondence, opinion or memorandum is submitted for audit. A capable tool returns: (1) a snapshot summary with green/amber/red counts; (2) a per-proposition audit table with proposition / flag / authority / ratio + pinpoint / confidence level / assessment / explanation; (3) an explicit fabrication audit; (4) practitioner-grade commentary on issues for further consideration.

#	Question	Answer-key target / Evaluation note
J1	LawCheck - Audit the following draft client letter paragraph and return a snapshot table (green/amber/red counts), a per-proposition audit, and an explicit fabrication audit: 'It is well established by the High Court in Smith v Jones [2018] HCA 47 at [42] that a contractual discretion is always constrained by an implied obligation of good faith. This was confirmed in Burger King Corporation v Hungry Jack's Pty Ltd [2001] HCA 23 per Gleeson CJ at [187]. Further, the High Court in Western Export Services Inc v Jireh International Pty Ltd [2011] HCA 45 abolished the parol evidence rule.' (Sample contains three deliberate errors for evaluation.)	<p>Target. (1) Smith v Jones [2018] HCA 47 - fabricated authority. (2) Burger King v Hungry Jack's is [2001] NSWCA 187, not [2001] HCA 23; Gleeson CJ was not on the bench (it is a NSWCA judgment). (3) Western Export stands for the opposite proposition - Codelfa's ambiguity gateway remains binding on lower courts. The first substantive proposition (contractual discretion implied good faith) is itself unsettled at HCA level. A capable tool returns the full audit table, flags each error, and produces the snapshot.</p> <p>Evaluation note. <i>Multi-error audit evaluating fabrication detection, court-of-record verification, and substantive misuse of authority in a single submission.</i></p>
J2	LawCheck - Audit a paragraph asserting that 'the Building Code 2016 obligations imposed by the Building and Construction Industry (Improving Productivity) Act 2016 (Cth) continue to apply to all Commonwealth-funded construction projects.' Identify whether the underlying statute is still in force.	<p>Target. The Building and Construction Industry (Improving Productivity) Act 2016 (Cth) was repealed by the Building and Construction Industry (Improving Productivity) Repeal Act 2022 (Cth). The Building Code 2016 obligations under that Act no longer apply.</p> <p>Evaluation note. <i>Evaluates detection of reliance on a repealed statute.</i></p>
J3	LawCheck - Audit a draft opinion paragraph asserting: 'Under s 588FA of the Corporations Act 2001 (Cth), a payment made by a company to an unsecured creditor within six months before the relation-back day is automatically voidable as an unfair preference.' Identify any incorrect statement of law.	<p>Target. Three errors. (1) s 588FA defines unfair preference but does not itself make it voidable - voidability arises under s 588FE. (2) The six-month relation-back period applies only where the transaction is an 'insolvent transaction' (s 588FE(2)). (3) The payment is not 'automatically' voidable - the liquidator must apply under s 588FF, and the creditor has running-account and good-faith defences (ss 588FA(3) and 588FG). All three must be flagged with statutory cross-references.</p> <p>Evaluation note. <i>Multi-error substantive audit across closely-related Corporations Act provisions, evaluating whether the tool reads the statute against the proposition rather than nodding through plausible-sounding insolvency law.</i></p>

K. AppealCheck - drafting grounds from a judgment

A judgment is submitted and the tool generates a draft notice of appeal. The expected output includes: (a) a threshold appealability check; (b) panel and reasons-structure analysis; (c) numbered grounds, each with error / correction / pinpoint references / authority / materiality; (d) a consolidation pass identifying overlapping grounds; (e) orders sought in the correct form.

#	Question	Answer-key target / Evaluation note
K1	AppealCheck - Upload AHG WA (2015) Pty Ltd v Mercedes-Benz Australia/Pacific Pty Ltd [2025] FCAFC 86 and generate a draft notice of appeal to the High Court identifying at least four appealable grounds, each with the threshold appealability check, pinpoint references to specific paragraphs in the joint reasons of Moshinsky, Bromwich and Anderson JJ, supporting authority, and materiality assessment.	<p>Target. Draft notice of appeal including: error in s 21 test articulation; failure to apply full breadth of statutory norms; misapplication of s 22 as non-exhaustive; legitimate-interests / reasonableness error; failure to give adequate reasons; Franchising Code good faith error. Each ground with paragraph cross-references, pinpoints, authority including Productivity Partners [2024] HCA 27, Kobelt [2019] HCA 18, Lux Distributors [2013] FCAFC 90, Quantum Housing [2021] FCAFC 40. Consolidation pass and orders sought.</p> <p>Evaluation note. <i>Evaluates whether the tool can produce a draft notice of appeal grounded in the judgment's reasoning with pinpointed grounds, supporting authority, and materiality analysis.</i></p>
K2	AppealCheck - Upload a single-judge NSWSC decision on contract construction (any post-2024 decision) and generate a draft notice of appeal to the NSW Court of Appeal identifying at least three grounds. Identify any leave requirements and any threshold appealability issues.	<p>Target. Threshold check (interlocutory v final, leave requirements under Supreme Court Act 1970 (NSW) s 101); grounds tied to specific judgment paragraphs; authority citations; orders sought in NSWCA format.</p> <p>Evaluation note. <i>Evaluates whether the tool adapts to a different appellate court with a different leave regime.</i></p>
K3	AppealCheck - Upload a tribunal decision (NCAT, ART, or FWC). Generate a draft of any appropriate appellate or review pathway, identifying which superior court has jurisdiction, the relevant gateway (error of law / merits / judicial review under ADJR or s 39B Judiciary Act 1903 (Cth)), and the time limit. Note: the Administrative Review Tribunal replaced the Administrative Appeals Tribunal in October 2024 under the Administrative Review Tribunal Act 2024 (Cth); the tool must use the current statutory framework.	<p>Target. Identification of correct appellate / review pathway; statutory pathway (e.g. CATA NSW s 80, ART Act 2024 s 172 (appeal on question of law to Federal Court), FW Act 2009 (Cth) s 605); time limit; grounds drafted in correct form for that pathway; explicit acknowledgment that AAT references in older materials should be read as ART under the transitional provisions.</p> <p>Evaluation note. <i>Evaluates handling of a recent statutory rename and replacement. Tools relying on training data that pre-dates October 2024 may incorrectly cite the abolished AAT framework.</i></p>

L. PleadingCheck + Pleadings research

Two related workflows. (1) PleadingCheck audits an existing pleading element-by-element against the cause of action. (2) Pleadings research operates before drafting: identifying the elements of a cause of action, the material facts that must be pleaded for each element, and the common pitfalls. Both deliverables map every requirement against the law.

#	Question	Answer-key target / Evaluation note
L1	PleadingCheck - Submit a statement of claim alleging misleading conduct under s 18 ACL with a claim for damages under s 236 against an individual director. Identify any element of the cause of action that has not been pleaded with sufficient particularity and any material fact that is missing.	<p>Target. Element-by-element analysis: (a) conduct attributable to the director personally (not just to the company); (b) in trade or commerce; (c) misleading or deceptive or likely to mislead or deceive; (d) reliance by the plaintiff specifically on the director's conduct (not just on corporate documents - the Yorke v Lucas / s 139C consideration); (e) causation; (f) loss; (g) accessorial pathway if pleaded under s 236 + s 2 'involved in' alternative. Identification of which elements are inadequately pleaded.</p> <p>Evaluation note. <i>Evaluates element mapping at the pleading level with statutory pinpoints.</i></p>
L2	Pleadings research - A plaintiff intends to plead a claim for breach of fiduciary duty against a former joint venturer who took a registered mortgage over joint venture property without prior disclosure. Identify (a) the elements of the cause of action; (b) the material facts that must be pleaded for each element; (c) the common pitfalls that lead to strike-out applications; (d) any relevant pleading rules (e.g. UCPR r 14.14 on particulars).	<p>Target. Elements: existence of fiduciary relationship; scope of duty; breach (conflict / non-disclosure / unauthorised profit); causation and loss / remedy sought. Material facts: undertaking of mutual trust / common venture (United Dominions v Brian); facts establishing the non-disclosure; absence of fully informed consent (Maguire v Makaronis); causal link to loss. Pitfalls: failure to plead facts establishing the fiduciary character of the relationship as distinct from a contractual one (Eaton v Rare Nominees consideration); failure to plead the absence of fully informed consent; conflating breach of contract with breach of fiduciary duty.</p> <p>Evaluation note. <i>Evaluates pre-drafting research mapped against elements and pitfalls - distinct from auditing an existing pleading.</i></p>
L3	Pleadings research - A plaintiff intends to plead an unfair preference claim against a creditor under ss 588FA and 588FE of the Corporations Act 2001 (Cth). Identify the elements, the material facts to be pleaded, the common pitfalls (including the Metal Manufactures v Morton set-off point and the running-account / Bryant v Badenoch peak indebtedness issue), and the relevant procedural rules.	<p>Target. Elements of unfair preference (s 588FA) AND voidability (s 588FE) AND insolvent transaction (s 588FC) - the statutory sequence. Material facts: parties, transaction, debt characterisation, insolvency at the time, relation-back day. Pitfalls: pleading s 588FA alone without s 588FE; failure to plead insolvency; failure to address running-account principle under s 588FA(3) post-Bryant v Badenoch [2023] HCA 2; assuming s 553C set-off available (no longer available after Metal Manufactures v Morton [2023] HCA 1); s 588FG good-faith defence considerations. Procedural: Federal Court Rules r 16.02 (material facts); UCPR r 14.14 (particulars).</p> <p>Evaluation note. <i>Evaluates integration of multi-section statutory architecture with two recent HCA decisions and pleading rules in a single pre-drafting deliverable.</i></p>

M. Document summarisation (cases, pleadings, affidavits, contracts)

Summarisation is distinct from review or audit - the deliverable is a structured digest, not a risk analysis. Case summaries are weighted in this category because they are typically the most diagnostic. Summaries should preserve doctrinal accuracy, pinpoints, and the structural integrity of the source document.

#	Question	Answer-key target / Evaluation note
M1	Case summary - Upload AHG WA (2015) Pty Ltd v Mercedes-Benz Australia/Pacific Pty Ltd [2025] FCAFC 86. Produce a case summary capturing: facts; procedural posture; issues; holding; reasoning (with pinpoints to the joint reasons of Moshinsky, Bromwich and Anderson JJ); orders; and significance.	<p>Target. Structured case summary with: factual matrix (dealer-to-agent model transition); procedural posture (appeal from primary judge's decision); issues (s 21 ACL unconscionability, Franchising Code cl 6 good faith); holding (appeal dismissed); reasoning with pinpoints to specific paragraphs in the joint reasons; orders; significance (clarifies the evaluative-against-statutory-values test).</p> <p>Evaluation note. <i>Evaluates whether the summary preserves doctrinal accuracy and pinpoints rather than flattening the reasoning into general terms.</i></p>
M2	Case summary - Upload Bryant v Badenoch Integrated Logging Pty Ltd [2023] HCA 2. Produce a case summary including the abolition of the 'peak indebtedness rule' and the correct application of the running-account principle in s 588FA(3) of the Corporations Act 2001 (Cth), with pinpoints.	<p>Target. Summary capturing: facts (running-account creditor of an insolvent logging company); issue (whether the peak indebtedness rule survived the 1992 statutory reformulation); holding (it did not); reasoning with pinpoints; orders; significance (resolution of a long-running insolvency-law dispute affecting every liquidator's preference claim).</p> <p>Evaluation note. <i>Evaluates summary accuracy on a doctrinally difficult HCA decision where the holding can easily be stated without the supporting reasoning structure.</i></p>
M3	Pleadings summary - Upload a statement of claim in a complex commercial dispute (e.g. multi-defendant ACL and breach of contract proceeding with cross-claims). Produce a structured summary identifying parties, causes of action against each defendant, relief sought, key factual allegations, and any pleaded defences or set-offs.	<p>Target. Structured summary distinguishing party-by-party causes of action; relief mapped against cause of action; key factual allegations; cross-claims; identification of any pleading rule references (UCPR / FCR).</p> <p>Evaluation note. <i>Evaluates whether the tool can navigate a multi-defendant pleading without conflating allegations across parties.</i></p>
M4	Contract summary - Upload a major project services agreement (e.g. a Victorian Government infrastructure contract with multi-year term and incorporated documents). Produce a structured summary covering: parties; term and renewal; scope of services; payment mechanism; risk allocation; termination rights; dispute resolution; key incorporated documents.	<p>Target. Structured contract summary with each section pinpointed to clauses; identification of incorporated documents; preservation of the contract's risk-allocation architecture.</p> <p>Evaluation note. <i>Distinct from the Contract Review workflow in Category N - the deliverable is a digest, not a risk analysis.</i></p>

N. Contract Review (Lite / Deep / Negotiation Pack) & Audit

Four discrete deliverables against the same uploaded contract. (1) Contract Review Lite - fast triage. (2) Contract Review Deep - risk summary table, per-clause case authority, and legislative status verification flagging any cited statute that is not in force or not retrievable. (3) Negotiation Pack - drafting-mark-up oriented output ready for redlining the counterparty's draft. (4) Defined Terms / Cross-References / Internal Clash Audit.

#	Question	Answer-key target / Evaluation note
N1	Contract Review Lite - Upload a mid-complexity commercial agreement (e.g. a software licence with limitation-of-liability and IP-assignment provisions). Produce a fast-triage report identifying the top 5–7 risks with clause references and one-line takeaways. The output should be readable in under 5 minutes.	<p>Target. Lite-mode report: 5–7 prioritised risks; clause references; one-line takeaways; no Legislative Status Verification table (that is Deep-mode territory).</p> <p>Evaluation note. <i>Evaluates whether the tool distinguishes Lite from Deep. A tool that always produces an exhaustive output regardless of mode does not satisfy the responsiveness dimension.</i></p>
N2	Contract Review Deep - Upload a major project services agreement (e.g. a Victorian Government infrastructure contract). Produce: (a) Executive Summary identifying overall risk profile and most significant red/amber flags; (b) Risk Summary Table with relevant authority, clause reference, risk flag, key takeaway and recommended action for each issue; (c) Legislative Status Verification table identifying every cited statute, its current status, version date and source URL, and flagging any that are not in force or not retrievable.	<p>Target. Risk Summary Table covering SOPA compliance (Southern Han v Lewence), time bars, consequential loss exclusions (Lucas Earthmovers, Amann Aviation), termination for convenience, confidentiality carve-outs, missing incorporated documents. Legislative Status Verification including identification that Building and Construction Industry (Improving Productivity) Act 2016 (Cth) is no longer in force.</p> <p>Evaluation note. <i>Evaluates the Deep review against an executed agreement with multiple regulatory touch-points, including detection of a repealed statute reference.</i></p>
N3	Negotiation Pack - Upload a counterparty-drafted franchise agreement. Produce a Negotiation Pack identifying clauses to delete, clauses to amend (with proposed redlines), clauses to add (with proposed wording), and the legal/commercial justification for each position. Map the negotiation positions against relevant statutory and case authority (e.g. Franchising Code cl 6, AHG WA Mercedes-Benz [2025] FCAFC 86 on unconscionability).	<p>Target. Negotiation Pack with action-by-action breakdown: delete / amend / add; proposed redlines or wording; justification mapped to law; cross-references to the leading authority on each negotiation position.</p> <p>Evaluation note. <i>Evaluates whether the output is drafting-ready rather than confined to issue-spotting.</i></p>
N4	Contract Audit (Defined Terms / Cross-References / Internal Clash) - Upload the same Victorian project services agreement. Produce a Defined Terms Audit Table flagging: (a) terms defined but not used or used but not defined; (b) circular definitions; (c) cross-references to clauses, schedules or annexures that do not exist or that point to repealed legislation; (d) inconsistent intent across clauses.	<p>Target. Per-term audit with five columns (Defined Term / Clause References / Issue Identified / Suggested Solution / Risk Flag). Specific flags include: 'Completion' as circular reference; 'Building Code' referencing repealed Act; 'OHS Legislation' referencing repealed Rail Safety (Local Operations) Act 2006.</p> <p>Evaluation note. <i>Evaluates exhaustive defined-terms and cross-reference audit including statute-status verification.</i></p>

O. Word plug-in integration

Pleadings Review and Contracts Review delivered as Microsoft Word plug-ins. The plug-ins are the same analytical workflows as Categories L and N, exposed inside Word with in-document overlay rather than as standalone reports. Evaluates usability, fidelity of the in-document overlay, and preservation of pinpoints and authority links.

#	Question	Answer-key target / Evaluation note
O1	Pleadings Review plug-in - Open a statement of claim in Microsoft Word and invoke the Pleadings Review plug-in. The plug-in should highlight, paragraph-by-paragraph, which elements of each cause of action have been pleaded, which are missing, and which are inadequately particularised. Each annotation should include a pinpoint to the supporting authority and a working hyperlink.	<p>Target. In-document mark-up: element-by-element overlay; pinpoints preserved on every annotation; hyperlinks resolve to the cited authority; round-trip preservation when the document is saved and reopened.</p> <p>Evaluation note. <i>Evaluates the plug-in's ability to deliver Category L analysis inside Word without losing pinpoints or breaking document formatting.</i></p>
O2	Contracts Review plug-in - Open a draft commercial contract in Microsoft Word and invoke the Contracts Review plug-in. The plug-in should annotate each clause with: any relevant statutory provision, any controlling case authority, a risk flag (red/amber/green), and a suggested amendment or position. Annotations should be insertable as Word comments or tracked changes.	<p>Target. In-document annotations: statute references with hyperlinks; case authority with pinpoints; risk flags; suggested amendments deliverable as Word comments or as tracked-changes redlines.</p> <p>Evaluation note. <i>Evaluates integration with Word's native review features. A plug-in that produces only side-panel commentary (and does not write into the document) does not satisfy the workflow-integration dimension.</i></p>
O3	Plug-in integration - For both the Pleadings Review and Contracts Review plug-ins, test: (a) latency on a 30-page document; (b) accuracy preservation compared to the standalone (browser-based) workflow on the same document; (c) graceful handling of large documents (50+ pages); (d) preservation of all pinpoint references and authority hyperlinks across save / reopen / share cycles.	<p>Target. Latency benchmarked; standalone-vs-plug-in accuracy compared (should be substantially identical); large-document handling without truncation; pinpoint and hyperlink preservation across full Word document lifecycle.</p> <p>Evaluation note. <i>Evaluates robustness of the plug-in as a production tool across the document lifecycle.</i></p>

Scoring rubric

Each answer is scored against ten dimensions on a 0–10 scale unless flagged otherwise. The first six dimensions are objective and can be marked from the output alone. The remaining four require senior-practitioner judgment.

Dimension	Scale	Definition
Pinpoint accuracy	0–10	% of substantive propositions accompanied by a correct paragraph, page or footnote reference.
Hyperlink integrity	0–10	% of cited authorities accompanied by a working link resolving to the correct source.
Citation fabrication	count	Number of fabricated cases, paragraph references, or authorities not on AustLII / Jade / the official register. Lower is better.
Currency	0–10	Whether the answer reflects the law as at the question date, including amendments and recent appellate authority.
Negative-finding accuracy	P/F	Where the answer key is 'no authority' or 'position unsettled', whether the tool says so without inventing a case.
Uncertainty / scope flagging	0–10	Whether the answer explicitly identifies what it has and has not searched for, and flags genuinely contestable propositions.
Time to verify (minutes)	minutes	Time taken by a senior practitioner to verify the answer end-to-end, including following every citation. Lower is better.
Defensibility on file	0–10	Whether the output could be attached to a memorandum to a client or court without redrafting.
Reasoning from primary sources	0–10	Whether the reasoning is built from statute and case law or paraphrased from secondary commentary.
Doctrinal completeness	0–10	Coverage of exceptions, qualifications, and competing authorities.

Operational notes

Note	Detail
Lock the answer key first.	Two senior practitioners independently work each question against primary sources and reconcile. The answer key is locked before any tool is run.
Use identical prompts.	Each question is submitted verbatim to every tool. No tool-specific phrasing. No re-prompting. Where a tool needs format scaffolding, apply the same scaffolding to every tool.
Run each tool in each of its output modes.	Tools with multiple output tiers (e.g. fast / deep, or research / contract review modes) should be run in the mode appropriate to each category. Categories N1, N2 and N3 specifically require Lite, Deep, and Negotiation Pack modes respectively.
Test the plug-ins on a real Word document.	Category O cannot be evaluated through a browser. Use a 30-page reference pleading and a 30-page reference contract. Test latency, annotation fidelity, hyperlink preservation, and round-trip save/reopen.
Refresh Category C quarterly.	The recent-authority category loses its diagnostic value as competitor training-data cutoffs advance. Replace 2 of the 4 questions every quarter with newer decisions. Preserve the prior set for longitudinal comparison.

Note	Detail
Categories J, K, L, M and N need source documents.	LawCheck, AppealCheck, PleadingCheck, summarisation and contract review/audit all require uploaded inputs. Prepare a reference document set: a draft advice letter with seeded errors; the AHG WA Mercedes-Benz judgment; a NSWSC contract-construction judgment; a tribunal decision; a multi-defendant statement of claim; a complex commercial contract; the Victorian Government project services agreement.
Capture raw outputs.	PDF every output with timestamp. Screenshot any inline hyperlinks. Record response time.
Blind the marker.	Anonymise tool names before marking. Apply the ten-dimension rubric above.
Score honest scope-limiting.	Where a tool explicitly discloses what it has and has not searched for (for example, in point-in-time research), that disclosure counts toward the uncertainty / scope flagging dimension. A tool that silently extends or restricts its output scores zero on that dimension even where the substantive content is otherwise correct.
Publish the per-question scoring sheet.	For each question, publish: the question text; the locked answer key; each tool's verbatim output (or URL); the dimension-by-dimension scores; the marker's verification notes.
Verify every fabrication finding.	Have a second senior practitioner verify every fabrication finding before publication.

Reporting

Per-question scores are reported across the ten rubric dimensions. Per-category and per-tool aggregates can be computed from the per-question scores. The benchmark does not prescribe a single composite score; aggregate weighting depends on the use case (a litigation practice may weight pinpoint accuracy and good-law verification more heavily than a transactional team, which may weight compliance triage and contract review more heavily).

The benchmark recommends publishing three layers of output: (a) headline numbers per tool across the ten rubric dimensions; (b) per-category radar charts for visual comparison; and (c) the full per-question scoring sheet for independent verification.