Appendix — Full Test Dataset

Companion to "UPDF AI vs Adobe Acrobat AI: We Tested Both on 12 PDFs."
Test date: September 2026 · UPDF AI desktop v2.5.6 · Adobe Acrobat AI desktop v2026.001.21789 · MacBook Air (M4, 2025), macOS Tahoe 26.2

This appendix publishes the full method so the results can be checked: the frozen prompts, the scoring rules, the aggregate scores, and a per-question verdict for all 33 questions across 12 documents (66 scored answers). It is a single-reviewer, anonymized evaluation — not an independently replicated benchmark.

---

1. The frozen prompts

Both tools received identical, word-for-word instructions, fixed before any answer was seen. Placeholders in `{curly braces}` were filled per document.

T1 — 5-point summary (identical for all 12 files):

Using ONLY the PDF I just uploaded, summarize its most important findings or terms in exactly 5 bullet points.

- Keep every key number, unit, date range, and qualifier (e.g. only, up to, excluding, unless) exactly as stated.
- Do not add any information that is not in this PDF.
- If the PDF does not state something, say it is not stated rather than guessing.

T2 — evidence-backed Q&A:

Using ONLY the PDF I just uploaded, answer this question: {FILE_QUESTION} Format: (1) direct answer; (2) the page number(s) that support it; (3) one short quoted sentence as evidence. If the PDF does not contain enough information, reply exactly: "The PDF does not contain enough information to answer this." Do not use any knowledge outside this PDF.

T3 — challenge task (one variant per document type):

- *Cross-page synthesis:* combine information from different sections to answer {QUESTION}; state the boundary/condition under which the conclusion holds; give page numbers and one quotation.
- *Table / figure reading:* read the {TABLE/CHART} on page {N}; report the exact value with its unit and the row/column or year; do not estimate.
- *Scanned extraction:* find and report {ITEM}; quote exactly as printed and give the page number; if any part is unreadable, say which.
- *Negative control:* answer {QUESTION_WITH_NO_ANSWER}; if the PDF does not address this, reply exactly: "The PDF does not contain this information." Do not infer or use outside knowledge.

2. How each answer was scored

Scoring is at the level of the **atomic fact**: every independently checkable value, mapping, or proposition is counted once, and its unit, year, and scope travel with it. Repeated facts inside quotations are not double-counted. Answers were anonymized (tool names removed, labelled A/B) before scoring, then restored.

Metric	Rule	Notable edge cases
Factual accuracy	Correct checkable facts ÷ total checkable facts.	A value with the wrong unit is marked incorrect even if numerically equivalent (e.g. £1,985,000 when £'000 was required). A misread sign or row/column is a factual error.
Coverage	Required key points addressed ÷ total required (correct, partial, or wrong all count as "addressed").	100% coverage is **not** a claim of correct wording — a point can be covered but stated wrongly, which is why accuracy is reported separately. Optional facts do not inflate the denominator.
Hallucination	Facts with no basis in the PDF ÷ total checkable facts.	A wrong value on a topic the PDF *does* cover is a factual error, **not** a hallucination. Hallucination is reserved for invented content.
Citation accuracy	References that genuinely support the claim ÷ references supplied.	A page number must be opened and confirmed. A confirmed printed-vs-PDF page offset is accepted. "No citation supplied" is n/a, not a failure.
No-answer (negative controls)	Correct refusal ÷ negative-control questions.	Only the exact refusal sentence passes; a correct-but-verbose refusal fails the strict output test without counting as a hallucination.
Task completion	Scorable results returned ÷ tasks submitted.	Upload failures, timeouts, refusals-to-process would count against this; none occurred here.

Percentages are computed from raw counts and micro-averaged (numerators and denominators summed), not by averaging per-question percentages.

3. Aggregate results

Metric	UPDF AI	Adobe Acrobat AI
Factual accuracy	97.7% (507/519)	95.5% (361/378)
Key-info coverage	92.5% (124/134)	95.5% (128/134)
Hallucination rate	0.77% (4 facts)	0.53% (2 facts)
Citation accuracy	66.7% (20/30)	80.0% (20/25)

No-answer (negative controls)	2/2 passed	0/2 passed
Task completion	100%	100%

Directional summary of the per-question verdicts below: UPDF stronger on 15, Adobe stronger on 9, eight ties, and three where neither met the bar. The two negative controls (F04-T3, F08-T3) are the clearest split — UPDF refused both correctly; Adobe answered both.

---

4. Per-question verdicts and full answers

For each question: the reviewer's verdict with full reasoning, then both tools' complete, unedited answers. "Tie (both weak)" means neither met the bar.

F01 — Acuity RM Group — Annual Report 2024

T1 Summary

UPDF AI is the better answer, but neither fully satisfies the strict rules.

Criterion	Acrobat AI	UPDF AI
Exactly five bullets	Five bullets, but adds a preamble—strictly, extra content	Pass
Numerical accuracy	Mostly accurate	Mostly accurate
Dates and qualifiers	Several altered or omitted	Better, but still imperfect
Importance of selection	Includes less central industry statistics	More focused on the company's results and position
Overall	6.5/10	8/10

Verified source location: **PDF pp. 3, 5–7, 13, 18, 20, 31–32, 38, 49 and 59; corresponding printed pp. 1, 3–5, 11, 16, 18, 29–30, 36, 47 and 57**.

UPDF AI answered:

- Revenue for the 12 months to 31 December 2024 was £2.1m (prior period £1.4m for an 8 month period from date of acquisition), with recurring revenue net revenue retention (NRR*) of 105% and forward contracted revenue (deferred income) of £3.4m at year end (2023: £2.9m).
- Adjusted loss for the year was £1.2m (2023: £1.0m) before exceptional costs of £141k and share based payments of £27k; reported loss for the year was £1.3m (2023: £1.3m) and basic and diluted loss per share was (0.92)p (2023: (1.39)p).
- Cash and debtors at year end were £1.3m (2023: £1.4m) of which cash was £0.6m (2023: £0.1m); loans and borrowings at year end were £0.2m (2023: £0.3m); directors concluded the financial statements were prepared on a going concern basis with forecasts through June 2026.
- Significant corporate actions in 2024: acquisition of Rizikon in November 2024; commencement and capitalisation of STREAM redevelopment in 2024 (capital

commitment £147,517 as at 31 December 2024); board changes with David Rajakovich appointed Chief Executive (became director 4 November 2024, CEO 13 December 2024) and Kate Buchan appointed Finance Director (joined Board 14 June 2024).
- Balance sheet positions as at 31 December 2024: Group total assets £6,967k (2023: £6,994k); goodwill £5,154k; intangible assets £319k; total liabilities £3,167k (2023: £2,906k); net assets £3,800k (2023: £4,088k); Company completed a fundraising post year end on 19 May 2025 raising gross proceeds £411,071 by issue of 41,107,143 shares at 1p per share (details stated in post balance sheet events).

Adobe Acrobat AI answered:

Let me see what the PDF reveals about its key findings and terms.

- Revenue for the year ended 31 December 2024 was £2.1m (2023: £1.4m for an 8-month period), with a net revenue retention (NRR) of 105% based on recurring revenue only, and forward contracted revenue increased 17% year-on-year to £3.4m as at year end (2023: £2.9m).
- The Group reported an adjusted loss for the year of £1.2m (2023: £1.0m) before exceptional costs of £141k (mainly organisational changes) and a share-based payment expense of £27k, resulting in a reported loss for the year of £1.3m (2023: £1.3m).
- In November 2024, Acuity RM Group plc acquired the Rizikon vendor management GRC product, and launched a new version of its STREAM® platform with AI features and a new vendor management hub, ready for new customers.
- At year end, cash and debtors totaled £1.3m (2023: £1.4m), with cash at £0.6m (2023: £0.1m), and loans and borrowings at £0.2m (2023: £0.3m); contract renewals for the year ran at 78%.
- 74% of large businesses, 70% of smaller businesses, and 66% of larger charities suffered cyber-attacks (UK Government, Cyber security breaches survey 2024); the average cost per breach is $5m, with business costs potentially exceeding $1bn and an average of 200 days to discover a breach (IBM, BreachSense); other key risks and uncertainties are detailed, but no dividend is recommended for the year.

T2 Evidence Q&A — verdict: Acrobat

The substantive answers are equally accurate and complete. Acrobat AI has better source location under the offset rule, while UPDF AI follows the quotation format more closely. Overall, Acrobat AI narrowly wins this question.

The correct evidence location is **PDF p. 5 / printed p. 3**.

UPDF AI answered:

1. Direct answer: At 31 December 2024, Acuity RM Group's forward contracted revenue was £3.4m, up 17% year-on-year from £2.9m in 2023. The report states that forward contracted revenue is not an IFRS measure.
2. Page number(s): Page 2 of the PDF.

3. Quoted evidence: “Forward contracted revenue, (deferred income per the balance sheet and forward contracted revenue*) has continued to grow increasing 17% year-on-year to £3.4m as at the year end (2023: £2.9m).”

Adobe Acrobat AI answered:

Reviewing the PDF for details on Acuity RM Group's forward contracted revenue.

1. At 31 December 2024, Acuity RM Group's forward contracted revenue was £3.4m, representing a 17% year-on-year increase from £2.9m at the end of 2023. The report states that this metric is not an IFRS measure.
2. Pages 3–4 of the PDF (Chief Executive’s Statement).
3. Quoted evidence: "Forward contracted revenue, (deferred income per the balance sheet and forward contracted revenue*) has continued to grow increasing 17% year-on-year to £3.4m as at the year end (2023: £2.9m). ... *NRR and forward contracted revenue are not IFRS measures.”

T3 Challenge — verdict: UPDF

UPDF AI is clearly better on this question because it follows the required £'000 unit, while Acrobat AI does not.

The correct values are **1,985 (£'000) for 31 December 2024** and **1,738 (£'000) for 31 December 2023**, on **PDF page 32** (printed page 30).

UPDF AI answered:

- Page: 32
- 31 December 2024 — Total Current liabilities: £1,985 £’000
- 31 December 2023 — Total Current liabilities: £1,738 £’000

Adobe Acrobat AI answered:

Okay, extracting exact total current liabilities for both years from page 32.

1. Total current liabilities at 31 December 2024 were £1,985,000 and at 31 December 2023 were £1,738,000, as shown in the respective year columns of the Group statement of financial position.
2. Page 32 of the PDF.
3. "Total Current liabilities 1,985 1,738"

F02 — Acuity RM Group — Annual Report 2023

T1 Summary — verdict: Acrobat

Acrobat AI has the more balanced operating summary and explicitly explains the old stake. UPDF AI supplies more balance-sheet detail, but misstates the scope of acquisition consideration (£4,210k includes the previously held £625k stake; consideration for the remaining shares is £3,585k - approximately £3.6m. The stated shares plus cash do not total

£4,210k) and omits two central operating checkpoints. Neither fully preserves all qualifications.

Verified source location: **PDF pp. 3–5, 7–10, 23, 32–33, 37, 50, 58–60 (printed pages = PDF minus 2**).

UPDF AI answered:

- Acuity RM Group plc acquired the remaining share capital of Acuity Risk Management Ltd on 25 April 2023 for total consideration of £4,210k, comprising 45,709,570 ordinary shares valued at £3,085k and £500k in cash, and consolidation resulted in goodwill of £5,154k recorded (Notes 24, 11, 13).
- Group revenue for the year ended 31 December 2023 was £1,366k (including £1,351k from the subsidiary from acquisition date 25 April 2023), gross profit £1,254k, administrative expenses £2,167k, and a loss for the year from continuing operations of £1,341k with basic and diluted loss per share (total and continuing) (1.39)p (Group statement of comprehensive income; note 1; note 9).
- At 31 December 2023, Group total assets were £6,994k, net assets £4,088k, cash and cash equivalents £100k, trade and other receivables £1,255k, intangible assets £233k, goodwill £5,154k, and deferred income current £1,016k and non-current £1,014k (Group statement of financial position; notes 10, 11, 14, 17, 18).
- The Company raised cash through issue of shares net of transaction costs of £1,057k during the year and completed share reorganisations: consolidation and subdivision of ordinary shares on 24 April 2023 (resulting ordinary shares 121,556,731 and deferred shares 2,645,954,765 at 31 Dec 2023), and issued adviser warrants exercisable at 4.5p (Notes 19, 20, 21; Group statement of cash flows).
- Key risks and mitigations are explicitly stated: performance risk, reliance on one or a few customers, reliance on people, market change, cybersecurity, technology competition, liquidity, legal/regulatory/political risk, natural disasters and interest rates, with mitigations including monthly monitoring, KPIs, multi-year contracts, competitive remuneration, use of leading global services for security, planned investment in STREAM® technology, maintaining material cash balances and engaging brokers for fundraising (Principal Risks and Uncertainties; Strategic Report).

Adobe Acrobat AI answered:

Let me see what the PDF reveals about its key findings and terms.

- In 2023, Acuity RM Group plc acquired the remaining shares in Acuity Risk Management Limited, changing its status from an investing to a trading company; the acquisition was completed on 25 April 2023 for a total consideration of £4,210,000 (comprising £3,085,000 in shares and £500,000 in cash, plus the prior 25% stake valued at £625,000).
- Revenue for the year ended 31 December 2023 was £1,366,000 (compared to £60,000 in 2022), with a loss for the year of £1,341,000 and basic and diluted loss per share of (1.39)p; no dividend is recommended for the year.
- Acuity RM won its two largest orders ever in 2023, each valued at approximately £500,000; forward contracted revenue increased by 32% to £2,900,000, orders won at annual contract value increased by 45% to £1,600,000, and the sales pipeline at

annual contract value increased by 88% to £7,900,000 (all as of 31 December 2023 compared to previous periods).
- The Group's principal risks include performance risk, reliance on a few customers, reliance on key individuals, market change, cybersecurity, technology competition, liquidity, legal/regulatory/political risk, natural disasters, and interest rate changes; mitigation strategies are in place but no system can fully eliminate risk.
- The Company consumed less than 40,000 kWh of energy during the period and therefore has not made separate disclosures relating to energy consumption and efficiency; there are no post balance sheet events or contingent liabilities to report.

T2 Evidence Q&A — verdict: Tie

The substantive answers tie. UPDF AI follows the short-quotation format more closely; both use a recoverable printed-page reference and neither cites both needed pages.

Verified source location: **PDF pp. 4-5; printed pp. 2-3**.

UPDF AI answered:

1. Direct answer: The sales pipeline at annual contract value increased from £4.2m at 31 March 2023 to £7.9m at 31 December 2023, an 88% increase. Sales leads are included in the sales pipeline using an industry standard sales qualification methodology.
2. Page number(s): Page 3 of the PDF.
3. Evidence: “Sales leads are included in the sales pipeline using an industry standard sales qualification methodology.”

Adobe Acrobat AI answered:

Reviewing the PDF for sales pipeline changes and lead inclusion details.

1. The sales pipeline at annual contract value increased by 88%, from £4,200,000 on 31 March 2023 to £7,900,000 on 31 December 2023. The report states that sales leads are included in the sales pipeline using an industry standard sales qualification methodology.
2. Supported by page 2 of the PDF (Chief Executive’s Report).
3. Quoted evidence: "Sales leads are included in the sales pipeline using an industry standard sales qualification methodology. Sales pipeline is not an IFRS measure."

T3 Challenge — verdict: Acrobat

Acrobat AI wins. UPDF AI miscopies one of the four requested cells (The 2022 Revenue cell is 60, not 1); this is a factual extraction error, not an invented subject.

Verified source location: **PDF p. 32; printed p. 30**.

UPDF AI answered:

Page: 32

Table of amounts (all values in £'000)

Row	Year ended 31 December 2023	Year ended 31 December 2022
Revenue	1,366	1
Gross profit	1,254	60

Adobe Acrobat AI answered:

Okay, extracting exact revenue and gross profit figures from page 32.

From the Group statement of comprehensive income, the exact figures are as follows.

Row	Year ended 31 December 2023 (£'000)	Year ended 31 December 2022 (£'000)
Revenue	1,366	60
Gross profit	1,254	60

Page: 32.

F03 — Acuity RM Group — Annual Report 2022

T1 Summary — verdict: UPDF

UPDF AI is stronger: it retains the unaudited qualifier and exact KPI comparison dates. Acrobat AI is factually sound on the counted claims but is less precise about source qualifications (the KPIs' "unaudited management information" qualifier is missing).

Verified source location: **PDF pp. 3, 6, 9, 35–36, 42, 49, 56 (printed pages = PDF minus 2)**.

UPDF AI answered:

- For the year ended 31 December 2022 the Group reported revenues of £60,000 (2021: £44,000), an operating loss of £256,000 (2021: loss £239000) and a loss before taxation of £341,000 (2021: loss of £422,000), with no dividend declared for the year (2021: £nil).
- At 31 December 2022 the Group's principal asset was a 25 per cent equity stake in Acuity Risk Management Limited valued at cost of £625,000, and the Group also held a legacy 5.8% holding in KCR Residential REIT plc valued at £305,000 (cost £1,705,000) which produced a fall in value of £85,000 in 2022 (2021: loss of £183,000).
- Group net assets fell to £1,227,000 at 31 December 2022 (2021: £1,547,000), net assets per ordinary share were 0.29p (2021: 0.37p), cash and cash equivalents were £222,000 (2021: £561,000), and total investments at fair value through profit or loss were £930,000 (2021: £1,015,000).
- Post balance sheet event: in April 2023 Drumz completed the acquisition of all outstanding shares in Acuity (acquisition approved by shareholders on 24 April 2023), the proposed total consideration was approximately £3.6 million comprising £0.5

million cash and issue of 45,709,570 New Ordinary Shares at 6.75p, together with a conditional placing and subscription to raise £1.45 million by issue of 32,222,222 New Ordinary Shares at 4.5p (professional costs of £78,000 incurred during the year treated as prepayments; further costs of £575,000 anticipated in 2023).
- Key KPIs for Acuity from unaudited management information: monthly recurring revenue £0.14 million as at 28 February 2023 (30 September 2022: £0.13 million, 31 March 2022: £0.10 million); forward contracted revenue £2.17 million as at 28 February 2023 (30 September 2022: £1.99 million, 31 March 2022: £2.17 million); renewal rate 96 per cent. as at 28 February 2023 (30 September 2022: 96 per cent., 31 March 2022: 82 per cent.); sales pipeline £3.83 million as at 22 March 2023 (30 September 2022: £1.67 million, 31 March 2022: £1.36 million).

Adobe Acrobat AI answered:

Let me check the PDF to identify the key points accurately.

- For the year ended 31 December 2022, Drumz plc reported revenues of £60,000 (2021: £44,000), an operating loss of £256,000 (2021: loss £239,000), and a loss before taxation of £341,000 (2021: loss of £422,000); no dividend was declared for the year (2021: £nil).
- As of year-end, Drumz plc owned a 25% equity stake in Acuity Risk Management Limited valued at cost of £625,000, and a 5.8% holding in KCR Residential REIT plc valued at £305,000 (cost: £1,705,000), with the KCR holding declining in value by £85,000 during the year.
- Post balance sheet, in April 2023, Drumz plc completed the acquisition of all outstanding shares in Acuity, making it a wholly owned subsidiary and reclassifying Drumz as a trading company under AIM Rules; the acquisition of the remaining c.75% of Acuity was for approximately £3.6 million (comprising £0.5 million in cash and 45,709,570 new ordinary shares at 6.75p per share).
- Key performance indicators for Acuity as at 28 February 2023 included monthly recurring revenue of £0.14 million, forward contracted revenue of £2.17 million, a renewal rate of 96%, and a sales pipeline of £3.83 million as at 22 March 2023.
- The Group's net assets at 31 December 2022 were £1,227,000 (2021: £1,547,000), and shareholders' funds have fallen due to losses incurred; it is not stated that any new standards or interpretations adopted during the year had a significant effect on the financial statements.

T2 Evidence Q&A — verdict: Acrobat

Acrobat AI wins on citation reliability. Both get the three required values right; UPDF AI's extra values are real, but its page references are wrong.

Verified source location: **PDF p. 3; printed p. 1**.

UPDF AI answered:

1. Direct answer: The value of the Group's KCR holding declined by £85,000 in 2022, from £390,000 to £305,000. The closing value was also stated as £304,714 at 31 December 2022.
2. Page number(s): PDF pages 4 and 52.

3. Evidence: “The share price performance of KCR once again has been extremely disappointing and as a result, the value of the KCR holding has declined further from £390,000 to £305,000, equating to a loss of £85,000 (2021: loss of £183,000).”

Adobe Acrobat AI answered:

Reviewing the PDF for specific figures on the Group's KCR holding decline.

1. The Group’s KCR holding declined by **£85,000** in 2022, from an opening value of **£390,000** to a closing value of **£305,000**.
2. Pages: **1** and **7**.
3. “the value of the KCR holding has declined further from £390,000 to £305,000, equating to a loss of £85,000””

T3 Challenge — verdict: Tie

Tie on all substantive metrics. UPDF AI is more concise; both report the exact signed values and correct page.

Verified source location: PDF p. 35; printed p. 33.

UPDF AI answered:

• Page: 35

Year	Basic and diluted earnings per share (from total and continuing operations)
2022	(0.8)p
2021	(1.2)p

Adobe Acrobat AI answered:

Okay, I'm extracting the exact earnings per share values from page 35.

1. Basic and diluted earnings per share from total and continuing operations were **(0.8)p** in the **2022** column and **(1.2)p** in the **2021** column.
2. Page: **35**.

F04 — Dietary Guidelines for Americans

T1 Summary — verdict: Acrobat

Acrobat AI is safer on factual grounding. UPDF AI preserves more population-specific details but introduces the unsupported “lower-heat” recommendation. Both lose some important dietary qualifiers.

Verified source location: **PDF pp. 2–10**; principal numeric guidance on **pp. 3–8**.

UPDF AI answered:

- The Dietary Guidelines for Americans, 2025–2030, emphasize returning to whole, nutrient-dense "real food"—protein, dairy, vegetables, fruits, healthy fats, and whole grains—while dramatically reducing highly processed foods with refined carbohydrates, added sugars, excess sodium, unhealthy fats, and chemical additives.
- Protein guidance includes prioritizing high-quality protein at every meal from varied animal and plant sources, using lower-heat cooking methods instead of deep-frying, consuming meat with no or limited added sugars/refined carbohydrates/chemical additives, and a protein serving goal of 1.2–1.6 grams of protein per kilogram of body weight per day, adjusted as needed based on individual caloric requirements.
- Dairy guidance recommends consuming full-fat dairy with no added sugars as an excellent source of protein, healthy fats, vitamins, and minerals, with a dairy serving goal of 3 servings per day as part of a 2,000-calorie dietary pattern, adjusting as needed based on individual caloric requirements.
- Specific intake targets and limits include vegetables: 3 servings per day and fruits: 2 servings per day for a 2,000-calorie dietary pattern (adjusting as needed); whole grains: 2–4 servings per day (adjusting as needed); saturated fat should not exceed 10% of total daily calories; sodium for ages 14 and above should be less than 2,300 mg per day with age-specific lower limits for children (1–3: less than 1,200 mg/day; 4–8: less than 1,500 mg/day; 9–13: less than 1,800 mg/day); and no more than 10 grams of added sugars per one meal.
- Special population recommendations include: Infancy & Early Childhood (Birth–4 Years) guidance to feed only breast milk for about the first 6 months (use iron-fortified infant formula when breast milk is not available), continue breastfeeding as long as mutually desired for 2 years or beyond, stop infant formula at 12 months and give whole milk, provide 400 IU vitamin D daily for breastfed infants or those consuming less than 32 ounces of formula per day starting shortly after birth, introduce solids at about 6 months while continuing milk/formula, and avoid added sugars during infancy and early childhood; other groups (Middle Childhood 5–10 Years, Adolescence 11–18 Years, Young Adulthood, Pregnant Women, Lactating Women, Older Adults, Individuals with Chronic Disease, Vegetarians & Vegans) have specific nutrient emphases and cautions as stated in the PDF (dates such as January 2026 and the report title 2025–2030 are provided).

Adobe Acrobat AI answered:

Let me see what the PDF reveals about its key findings and terms.

- The Guidelines for 2025–2030 emphasize eating whole, nutrient-dense foods—protein, dairy, vegetables, fruits, healthy fats, and whole grains—while dramatically reducing highly processed foods with refined carbohydrates, added sugars, excess sodium, unhealthy fats, and chemical additives.
- Protein serving goals are 1.2–1.6 grams of protein per kilogram of body weight per day, and dairy serving goals are 3 servings per day as part of a 2,000-calorie dietary pattern, both adjustable based on individual caloric requirements.
- For a 2,000-calorie dietary pattern, the recommended daily servings are: vegetables—3 servings, fruits—2 servings, and whole grains—2–4 servings; saturated fat consumption should not exceed 10% of total daily calories, and sodium intake for ages 14 and above should be less than 2,300 mg per day.
- No amount of added sugars or non-nutritive sweeteners is recommended; one meal should contain no more than 10 grams of added sugars, and specific snack limits are

set (e.g., grain snacks: ≤5 grams added sugar per ¾ ounce whole-grain equivalent, dairy snacks: ≤2.5 grams per ⅔ cup equivalent).
- For infants, exclusive breastfeeding is recommended for about the first 6 months, with introduction of solid foods at about 6 months; avoid added sugars during infancy and early childhood, and introduce potentially allergenic foods (e.g., nut butters, eggs, shellfish, wheat) at about 6 months as appropriate.

T2 Evidence Q&A — verdict: UPDF

UPDF AI is better overall. Both answer the nutrition question correctly, but Acrobat AI omits the requested page number and gives two quotations.

Verified source location: **PDF p. 3; printed p. 2**.

UPDF AI answered:

1. Direct answer: The Guidelines state a daily protein target of 1.2–1.6 grams of protein per kilogram of body weight per day and a dairy serving goal of 3 servings per day as part of a 2,000-calorie dietary pattern. Both should be adjusted as needed based on your individual caloric requirements.
2. Page number(s): PDF pages 2–3.
3. Evidence: “Protein serving goals: 1.2–1.6 grams of protein per kilogram of body weight per day, adjusting as needed based on your individual caloric requirements.”

Adobe Acrobat AI answered:

Reviewing the PDF for protein targets, dairy goals, and related adjustments.

1. The Guidelines state a daily protein target of 1.2–1.6 grams of protein per kilogram of body weight per day, and a dairy serving goal of 3 servings per day as part of a 2,000-calorie dietary pattern; both should be adjusted as needed based on individual caloric requirements.
2. Page(s): Not explicitly numbered, but found in the main dietary recommendations section under "Protein, Dairy & Healthy Fats".
3. Evidence: "Protein serving goals: 1.2–1.6 grams of protein per kilogram of body weight per day, adjusting as needed based on your individual caloric requirements." and "Dairy serving goals: 3 servings per day as part of a 2,000-calorie dietary pattern, adjusting as needed based on your individual caloric requirements."

T3 Challenge — verdict: UPDF

UPDF AI passes the strict negative control; Acrobat AI is semantically honest but fails only the exact-output requirement (Honest substantive refusal, but an extra preamble precedes the required sentence. Under this negative-control prompt’s explicit “reply exactly” condition, the strict pass is No.). Do not mislabel Acrobat AI’s preamble as a hallucination.

Verified source location: **No supporting page**; verified against the full 10-page PDF.

UPDF AI answered:

The PDF does not contain this information.

Adobe Acrobat AI answered:

Okay, checking the PDF for the recommended maximum daily caffeine intake.

The PDF does not contain this information.

F05 — Employee Choice Guide for IT

T1 Summary — verdict: Acrobat

Acrobat AI is the better programme summary despite its inaccurate blanket denial of technical requirements. UPDF AI misreads the chart and devotes most of its summary to survey statistics rather than the programme's five principles (The correct chart pairs are Innovation 32%/40%, Performance 71%/76%, Satisfaction 50%/60%).

Verified source location: **PDF pp. 2–12 and endnotes on p. 14**; chart labels visually checked.

UPDF AI answered:

- A PwC survey found that 78% of millennials believe having access to the technology they like at work makes them more effective in their job (source dated 2011).
- A Jamf Employee Choice survey showed that roughly 3 out of 4 employees would choose Apple for use at work (source dated 2016).
- A Gensler Workplace Study cited that providing employees with choice can notably impact innovation, job performance, and job satisfaction, with example percentages shown as 76% (Innovation), 71% (Performance), 60% (Satisfaction) in the document (source dated 2013 / Harvard Business Review).
- IBM reported saving up to $543 for each Mac it purchases and that only 5% of Mac users needed support versus about 40% of PC users (source: IBM / 2015 Mac@IBM, Zero to 30,000 in 6 Months).
- The document does not state exact enterprise adoption rates, overall cost models for every company, or precise timelines for rollout beyond recommending steps (get prepared, assemble the team, establish a management solution, complete a security review, test deployment, prepare to launch, continuously improve, spread the word); specific quantitative metrics for program ROI beyond the IBM example are not stated..

Adobe Acrobat AI answered:

Let me see what the PDF reveals about its key points.

- Roughly 3 out of 4 employees would choose Apple for use at work, according to a Jamf Employee Choice survey; a PwC survey found that 78% of millennials believe having access to preferred technology at work makes them more effective.
- IBM saves up to $543 for each Mac it purchases due to higher residual values and lower support costs; at IBM, only 5% of Mac users needed support versus about 40% of PC users.

- The guide recommends a new model for IT: corporate-provisioned employee choice programs, modern provisioning and deployment (including zero-touch deployment and MDM), personalized setup, shared responsibility for support, and flexible financial models such as device-as-a-service and payroll deduction.
- The document emphasizes that employee-focused experiences, including involvement in program design and communication, increase engagement, productivity, and satisfaction, and that internal alignment between IT, HR, procurement, finance, and information security is critical.
- No specific technical requirements, hardware specifications, or detailed cost breakdowns are stated in the document; if such details are needed, they are not stated.

T2 Evidence Q&A — verdict: Tie (both weak)

Neither answers correctly. UPDF AI recovers two main labels but misorders one; Acrobat AI recovers only the first. Correct sequence: Get prepared → Establish a management solution → Prepare to launch.

Verified source location: **PDF p. 9; printed p. 9**.

UPDF AI answered:

1. Direct answer: The guide's three basic steps are: • Get prepared • Assemble the team • Establish a management solution
2. Page number(s): PDF pages 9–10.
3. Evidence: "There are three basic steps involved in creating your own program."

Adobe Acrobat AI answered:

1. The three basic steps for creating an Apple at Work employee choice program are: (1) Get prepared, (2) Assemble the team, and (3) Evaluate your infrastructure.
2. Supported on page 9 of the PDF.
3. Evidence: "There are three basic steps involved in creating your own program. 1 • Get prepared 2 • Assemble the team 3 • Evaluate your infrastructure"

T3 Challenge — verdict: Tie

Tie. Both preserve the exact figures and the important "about" qualifier.

Verified source location: **PDF p. 8; printed p. 8**.

UPDF AI answered:

- Mac users: 5%
- PC users: about 40%
- Page: 8

Adobe Acrobat AI answered:

1. Checking the IBM support-rate graphic on page 8 now.

2. Mac users: 5% needed support; PC users: about 40% needed support. This information is found on page 8 of the PDF.
3. Quoted evidence: "At IBM, only 5% of Mac users needed support vs. about 40% of PC users."

F06 — Transformative Service in Healthcare (paper)

T1 Summary — verdict: Tie (both weak)

Neither is fully source-bound: both add unmarked coping-category classifications. UPDF AI retains more methodology and the conditional mechanism; Acrobat AI is more focused on the three named stressors but omits the transformational-care condition (if adequate coping strategies are available, HCPs could provide transformational care).

Verified source location: **PDF pp. 2, 5–7, 10–16 (printed journal pages = PDF plus 22**).

UPDF AI answered:

- The study introduces and defines "secondary vulnerability" as the susceptibility for a frontline employee (FLE) who is providing support of a vulnerable consumer to experience state-based harm themselves when the job demands outweigh the resources available, and proposes a theoretical framework of secondary vulnerability applied to healthcare professionals (HCPs) in end-of-life care.
- Using the Transactional Model of Stress and Coping (TMSC; Lazarus and Folkman, 1987), the research identifies key stressors for HCPs: family emotional unpreparedness, personal challenges in delivering end-of-life care, organizational constraints and delays, limited access to community resources for end-of-life care, and rapid escalation to end-of-life.
- Coping strategies observed and categorized were: promoting person-centred care and communication (emotion-focused), enhanced staff competence and confidence through education and training (problem-focused), and staff holding limited accountability (problem-focused/non-transformative), with these strategies influencing whether HCPs assume a Transformative Service Provider (TSP) role.
- Outcomes linked to coping strategy effectiveness included improved person and family involvement in decision-making, depersonalized care when coping resources were insufficient (e.g., reliance on protocols, deferred decisions), and impacts on staff well-being including emotional discomfort and professional strain; professional development initiatives were reported to yield positive outcomes for staff confidence.
- Methodological details stated: an interpretive qualitative study conducted in a 750-bed public tertiary teaching hospital in south-east Queensland, Australia; data collection via semi-structured interviews from September to November 2019 with eleven HCPs (individual n = 8, group n = 1), mean participant age 36, interviews averaged 19 min (range 10 to 39 min); sample was mostly nurses with one social worker.

Adobe Acrobat AI answered:

Let me see what the PDF reveals about its key points.

- The study introduces and defines 'secondary vulnerability' as the susceptibility of frontline healthcare professionals (HCPs) to experience state-based harm when job

demands in end-of-life care outweigh available resources, impacting both their well-being and service outcomes.

- Using a qualitative approach, 11 HCPs (mostly nurses, one social worker) from a 750-bed public tertiary teaching hospital in south-east Queensland, Australia, were interviewed between September and November 2019; interviews averaged 19 minutes, with participants' mean age being 36.
- Key stressors identified for HCPs include family emotional unpreparedness, personal challenges in delivering end-of-life care, and organizational constraints and delays, all of which contribute to heightened stress and secondary vulnerability.
- Coping strategies observed among HCPs include promoting person-centred care and communication (emotion-focused), enhancing staff competence and confidence through ongoing training (problem-focused), and, at times, holding limited accountability to manage emotional burden and professional boundaries.
- Outcomes of these coping mechanisms include improved person and family involvement in decision-making, instances of depersonalized care due to systemic barriers, and significant impacts on staff well-being; the study emphasizes the need for targeted interventions and organizational support to mitigate secondary vulnerability and enhance care quality.

T2 Evidence Q&A — verdict: Acrobat

Acrobat AI wins on source location. Both state the three required facts correctly.

Verified source location: **PDF p. 2; printed journal p. 24**.

UPDF AI answered:

1. Direct answer: Secondary vulnerability is the susceptibility of a frontline employee supporting a vulnerable consumer to experience state-based harm themselves. It arises when the job demands outweigh the resources available.
2. Page number(s): PDF pages 1 and 3.
3. Quoted evidence: “Secondary vulnerability is defined here as the susceptibility for a frontline employee (FLE) who is providing support of a vulnerable consumer to experience state-based harm themselves when the job demands outweigh the resources available.”

Adobe Acrobat AI answered:

Okay, I'm analyzing the document to find the definition and conditions of secondary vulnerability.

1. Direct answer: Secondary vulnerability for a frontline employee is defined as the susceptibility to experience state-based harm when supporting a vulnerable consumer, specifically arising when job demands outweigh the resources available to the employee.
2. Page number(s): Page 24 (Introduction section)
3. Quoted sentence: "Secondary vulnerability is defined here as the susceptibility for a frontline employee (FLE) who is providing support of a vulnerable consumer to experience state-based harm themselves when the job demands outweigh the resources available."

T3 Challenge — verdict: UPDF

UPDF AI is substantively safer because it retains the tentative conditional claim. Acrobat AI locates more of the relevant discussion but overstates what the condition proves (The study says “If adequate coping strategies were available, HCPs could provide transformational care.” It does not establish the converse or an exclusive necessary-and-sufficient rule).

Verified source location: **PDF pp. 2 and 15; printed journal pp. 24 and 37**.

UPDF AI answered:

1. External stressors: Family emotional unpreparedness, organizational constraints and delays, limited hospital or community resources, and rapid escalation to end-of-life.
2. Internal stressors: Personal challenges in delivering end-of-life care, including personal biases, discomfort with end-of-life language, and concerns about professional scope.
3. Condition for transformational care: HCPs could provide transformational care when adequate coping strategies and resources were available. These included person-centred care and communication, enhanced staff competence and confidence, and support from internal and external resources.
4. Boundary/condition: This conclusion applies specifically to HCPs providing care and support to patients at the end of life; when coping resources were unavailable, HCPs were likely to experience secondary vulnerability, which could negatively affect patient and HCP outcomes.
5. Supporting page numbers: PDF pp. 27–28.
6. Quoted evidence: “However, when HCPs do have the available coping resources available, it is proposed that they will take on the role of a TSP and provide a transformative service to the patient.”

Adobe Acrobat AI answered:

I'm extracting details on stressors and conditions for transformational care from the relevant sections.

1. Direct answer: External stressors leading healthcare professionals to perceive secondary vulnerability in end-of-life care include family emotional unpreparedness and organizational constraints and delays; internal stressors include personal challenges in delivering end-of-life care. Healthcare professionals could provide transformational care only when adequate coping strategies and resources are available to manage these stressors—otherwise, secondary vulnerability persists and transformational care is not achieved. The boundary condition is that transformational care is possible only if coping strategies are sufficient to meet the demands.
2. Page numbers: 30–32 (Findings: Stressors), 37–38 (Discussion: Condition for transformational care)
3. Quoted sentence: "Based on the research findings and applying the TMSC as a guiding framework, the research demonstrates how external stressors (such as family emotional unpreparedness, organizational constraints and delays) and internal stressors (including personal challenges in delivering end-of-life care) led HCPs to perceiving a level of secondary vulnerability in their provision of care for end-of-life

patients. ... If adequate coping strategies were available, HCPs could provide transformational care to patients..."

F07 — Nudging University Students (paper)

T1 Summary — verdict: UPDF

UPDF AI is stronger and more complete, especially on study design and implementation capacity. Acrobat AI is broadly grounded but omits several conditions needed to interpret the findings (Omits the focus-group sample/date, explicitly interlinked roles, detailed service-capacity planning and score-to-service mappings).

Verified source location: **PDF pp. 2–4, 7–9 (printed journal pages = PDF plus 56)**.

UPDF AI answered:

- The study evaluated a proactive well-being analytics intervention at a UK university that asked new and returning students during online enrolment to complete the 5-item WHO-5 Well-Being Index and consent to tailored email "nudges" signposting support, with qualitative data collected via an in-person focus group of 6 counselling and mental health (CMH) practitioners in November 2024, three weeks after a nudge email had been sent.
- Staff perceived the intervention as distinct from usual reactive outreach, noting it identified students who would typically remain unseen or be unlikely to self-refer, and they viewed the tailored, brief messages as useful "consciousness raising" that could prompt earlier engagement with low- and high-intensity services.
- Participants consistently reported that the operational task of identifying students and sending nudges should be separate from CMH service delivery because combining roles could create ethical complexities and a perceived increased duty of care, and they recommended separate but interlinked roles for data analysis and student support.
- Implementation factors raised included concerns about timing (nudges sent around half-term and close to another proactive campaign), potential email fatigue reducing immediate spikes in referrals, variable internal communication about the campaign (some staff feeling insufficiently informed), and the need to align nudge delivery with staff availability and a strategic calendar to avoid overstretching CMH capacity.
- Quantitative impact data and long-term maintenance effects are not stated in this PDF; the paper indicates future integration of these qualitative findings with quantitative impact findings but does not provide those outcome figures here.

Adobe Acrobat AI answered:

Let me see what the PDF reveals about its key points.

- The intervention used the 5-item World Health Organization Well-Being Index (WHO-5) during online enrolment, with students consenting to receive tailored email nudges signposting support services based on their WHO-5 score.
- Staff reported that the nudging approach uniquely identified students who would typically remain unseen or unlikely to self-refer, targeting both those at early stages of distress and those with severe concerns reluctant to engage.

- The intervention did not produce a clear spike in self-referrals as seen in previous periods, with staff suggesting possible reasons including 'email fatigue' and the timing of the campaign around half-term.
- Staff universally agreed that the process of identifying at-risk students and sending nudges should be separate from their direct support roles, citing ethical and workload concerns if these were combined.
- The document does not state specific quantitative outcomes (such as exact numbers of referrals or improvements), and notes that the long-term impact and quantitative results will be integrated in future research.

T2 Evidence Q&A — verdict: Acrobat

Acrobat AI wins because its citation is correct under the printed-page-offset rule. Both cover all four required facts.

Verified source location: **PDF p. 3; printed journal p. 59**.

UPDF AI answered:

1. Direct answer: Students with lower WHO-5 scores were directed to more specialist counselling; moderate scorers were signposted to guided self-help options; and high scorers were provided with generic information.
2. Page number(s): Page 63.
3. Evidence: “Those scoring at the lower extremes are expected to be experiencing mental ill health … and would be directed to more specialist counselling, whereas moderate scorers would be signposted to guided self-help options, and high scorers would be provided with generic information.”

Adobe Acrobat AI answered:

1. The intervention tailored signposting by directing students with lower WHO-5 scores to more specialist counselling, those with moderate scores to guided self-help options, and those with high scores to generic information.
2. Supported on page 59 of the PDF.
3. Quoted evidence: "those scoring at the lower extremes are expected to be experiencing mental ill health (Topp et al. 2015) and would be directed to more specialist counselling, whereas moderate scorers would be signposted to guided self-help options, and high scorers would be provided with generic information."

T3 Challenge — verdict: UPDF

UPDF AI is more complete on interlinked roles and capacity planning. Both give a recoverable but incomplete printed-page range; Acrobat AI’s evidence is a multi-passage composite.

Verified source location: **PDF pp. 8-9; printed journal pp. 64-65**.

UPDF AI answered:

1. **Unique benefit:** Staff saw the analytics-based nudges as unique because they used data in a novel way to identify students who would typically remain unseen or would be unlikely to self-refer through usual service pathways.
2. **Service-design and timing conditions:** Effective implementation required:
   - Separate but interlinked roles for data analysis/student identification and pastoral or CMH support, because identifying students through analytics requires a different skill set from supporting them through well-being difficulties.
   - Communication between the analytics team and CMH services so nudge delivery matched service capacity and parallel university campaigns.
   - Staff preparedness, including adequate education and training.
   - Timing nudges around staff availability, avoiding school holidays and periods close to other proactive campaigns, supported by a strategic calendar and wider institutional oversight.
3. **Boundary/condition:** The conclusion applies only when the intervention is implemented with sufficient CMH capacity and staff availability to respond to resulting student queries; otherwise, timely signposting may be undermined because staff may be unavailable or overstretched.
4. **Supporting pages:** PDF pages 65–66.
5. **Quoted evidence:** "Well-being analytics offers an opportunity to offer a differentiated approach, which is proactive and recognised as a positive measure by staff working in CMH teams."

Adobe Acrobat AI answered:

Reviewing discussion and conclusion sections for key insights.

1. Staff saw the unique benefit of analytics-based nudges as their ability to identify and reach students who would not typically be seen or who would not self-refer through usual service pathways; for effective implementation, the authors stated that the identification and sending of nudges must be operationally separate from direct support roles, and that timing must align with staff availability and avoid overlap with other campaigns or staff leave periods.
2. Supported on pages 65–66 of the PDF.
3. Quoted evidence: "CMHS clearly recognised that the nudging intervention was unique in how it utilised data in a novel way to identify students that typically would not be identified or who choose not to engage through usual service pathways... operationalising the identification of students at-risk and sending of nudges needed to sit separately from the service they provide... To optimise readiness and to better isolate the impact of the nudges, it is imperative that nudges are not sent around school holidays when CMHS may be taking leave, to ensure remaining staff are not overstretched and details regarding nudges are not miscommunicated."

F08 — Barriers to MH Services, Rural/Remote (review)

T1 Summary — verdict: Tie

Both reproduce the main findings accurately. UPDF AI offers fuller methods/detail; neither flags the source's database/country inconsistencies or preserves all eligibility limitations.

Verified source location: **PDF pp. 1, 3–4, 11–15**; printed and physical page numbers agree.

UPDF AI answered:

- The scoping review included a total of 30 studies from 11 countries published between 2007 and 2024, identified from 1,879 articles retrieved across six databases with screening reducing to 30 eligible studies.
- Barriers were categorized using the Socio-ecological Resilience Framework into four levels: system/policy (limited resources in 28 articles; workforce shortages in 21 articles; quality of services in 9 articles; cost in 8 articles; system complexity/technology/navigation in 3 articles; privacy/confidentiality in 2 articles; policy in 1 article), social/community (community stigma and discrimination in 28 articles; geographic and cultural differences in 5 articles; lack of community awareness in 1 article; environmental stress in 1 article), family/peer level (peer and family support in 4 articles; family stigma in 1 article), and patient level (attitude toward services in 22 articles; patient understanding of mental illness in 13 articles; quality of life in 2 articles).
- Geographic distribution of the 30 studies included Australia (n = 13), America/USA (n = 8), Canada (n = 1), South Africa (n = 1), Vietnam (n = 2), India (n = 1), Mexico (n = 1), Cambodia (n = 1), Liberia (n = 1), and Malawi (n = 1), with studies using mixed methods (13), cross-sectional surveys (16), and various assessment scales (e.g., BHSS, BSI-18, PHQ-9, SF36, SF12).
- Common specific system/policy barriers detailed included long waiting times; lack of locally available general and specialised services; poor physical infrastructure and communication support (e.g., no electricity or mobile networks); high delivery costs; and recruitment/retention difficulties for specialised mental health professionals.
- The PDF does not state precise effect sizes, pooled quantitative estimates, or a formal quality appraisal of included studies, and it states that the quality of included studies was not assessed because the purpose was to map evidence.

Adobe Acrobat AI answered:

Let me see what the PDF reveals about its key points.

- The review analyzed 30 studies from 11 countries (2007–2024) and identified barriers to accessing mental health services in rural or remote areas at four levels: system/policy, social/community, family, and individual, using the Socio-ecological Resilience Framework.
- Key system/policy barriers included limited resources (28 articles), workforce shortages (21 articles), high costs (8 articles), poor service quality (9 articles), complex systems, technology challenges, and privacy concerns; these issues led to long wait times, lack of local services, and high out-of-pocket expenses.
- Social/community barriers were dominated by community stigma and discrimination (28 articles), geographic and cultural differences (5 articles), lack of community awareness (1 article), and environmental stress (1 article), with stigma and distance being especially prominent.
- Family-level barriers included weak family or peer support (4 articles), family stigma (1 article), and poor education, with some rural residents preferring informal support over professional care.
- Individual-level barriers were negative attitudes toward mental health services (22 articles), poor understanding of mental illness (13 articles), and low quality of life (2

articles); many rural residents failed to recognize symptoms or delayed seeking help due to self-reliance or stigma.

T2 Evidence Q&A — verdict: Tie

Both are substantively correct. UPDF AI follows the requested one-sentence evidence format more closely. Do not penalise Acrobat AI for “missing 11 countries”: it appears in the quotation.

Verified source location: **PDF p. 1; printed p. 1**.

UPDF AI answered:

1. Direct answer: The review’s evidence base comprised 30 studies from 11 countries. The barriers were grouped into four socio-ecological levels: system/policy, social/community, family, and patient.
2. Page number(s): PDF pages 1 and 3.
3. Quoted evidence: “A total of 30 studies from 11 countries were reviewed.”

Adobe Acrobat AI answered:

Reviewing the PDF for evidence base size and barrier categories.

1. The review's evidence base consisted of 30 studies, and barriers to accessing mental health services were grouped into four socio-ecological levels: system/policy, social/community, family, and patient (individual).
2. Page(s): Abstract (first page), Method (Eligibility criteria), and Results sections.
3. Quoted evidence: "A total of 30 studies from 11 countries were reviewed. Barriers were grouped into four levels: system/policy, social/community, family, and individual, which are based on the Socio-ecological Resilience Framework."

T3 Challenge — verdict: UPDF

UPDF AI passes the strict negative control. Acrobat AI fails only the exact-output condition; both correctly recognise that the requested pooled effect is absent.

Verified source location: **No supporting page**; verified against the full 17-page PDF.

UPDF AI answered:

The PDF does not contain this information.

Adobe Acrobat AI answered:

Okay, I'm checking the PDF for the pooled odds ratio and confidence interval details.

The PDF does not contain this information.

F09 — 2026 National Defense Strategy

T1 Summary — verdict: Tie

UPDF AI better preserves the four strategic priorities, but Acrobat AI has no comparable numerical misreading. UPDF AI's Non-U.S.-NATO GDP error is material (PDF p. 19 graphic gives Non-U.S. NATO $26 Trillion and Russia $2 Trillion. The answer assigns Russia's number to the wrong entity); neither summary should be treated as independently verified geopolitical fact.

Verified source location: **PDF pp. 9–14, 18–20, 24–29; principal LOE list p. 24 / printed p. 16**.

UPDF AI answered:

- The document emphasizes that President Trump reentered office in January 2025 and rebuilt the American military, stating that he "took office in January 2025" to one of the most dangerous security environments in the nation's history and framing subsequent actions (e.g., Operation ABSOLUTE RESOLVE, Operation MIDNIGHT HAMMER) as demonstrations of restored U.S. military capability.
- The National Defense Strategy (NDS) sets four Lines of Effort (LOEs): 1. Defend the U.S. Homeland; 2. Deter China in the Indo-Pacific Through Strength, Not Confrontation; 3. Increase Burden-Sharing with U.S. Allies and Partners; 4. Supercharge the U.S. Defense Industrial Base.
- The NDS calls for a new global NATO defense-spending standard of 3.5% of GDP on core military spending plus an additional 1.5% on security-related spending, for a total of 5% of GDP, and advocates that allies and partners meet this standard worldwide.
- The strategy prioritizes specific homeland and hemispheric measures: sealing borders, deporting illegal aliens in coordination with DHS, countering narco-terrorists (citing Operation ABSOLUTE RESOLVE), securing access to key terrain including Greenland, the Gulf of America, and the Panama Canal, and developing "Golden Dome for America" to defend skies and counter unmanned aerial threats.
- The document states quantitative comparisons and geopolitical assessments as presented: the Non-U.S. NATO nominal GDP is 2 Trillion (2024) per World Bank (2024); it asserts China is the second most powerful country and that the Indo-Pacific will soon make up more than half of the global economy, and it describes Russia, Iran, DPRK threats and recent U.S. operations (e.g., MIDNIGHT HAMMER) as in the text; where the PDF does not state specific timelines, force sizes, or budget toplines beyond the NATO percentage standard and the 2024 GDP figures, those specifics are not stated.

Adobe Acrobat AI answered:

Let me see what the PDF reveals about its key points.

- President Trump, since reentering office in January 2025, has rebuilt the American military to be the world's absolute best, prioritizing defending the U.S. Homeland, deterring China in the Indo-Pacific, increasing burden-sharing with allies, and supercharging the U.S. defense industrial base.

- NATO allies have committed to a new global standard for defense spending: 3.5% of gross domestic product (GDP) on core military spending and an additional 1.5% on security-related spending, totaling 5% of GDP, as set at the Hague Summit.
- The strategy emphasizes a strong denial defense along the First Island Chain (FIC) in the Indo-Pacific to prevent any nation, including China, from dominating the region, aiming for a balance of power that allows all to enjoy a decent peace.
- The U.S. will guarantee military and commercial access to key terrain in the Western Hemisphere, especially the Panama Canal, Gulf of America, and Greenland, and will enforce the Trump Corollary to the Monroe Doctrine, as demonstrated in Operation ABSOLUTE RESOLVE.
- Iran's nuclear program was obliterated by Operation MIDNIGHT HAMMER, and the U.S. will continue to prevent Iran from acquiring nuclear weapons; Israel is described as a model ally, and regional partners are expected to take greater responsibility for their own defense..

T2 Evidence Q&A — verdict: Tie

Tie on facts and coverage; both fail source-location verification. UPDF AI's evidence is more concise, but neither is fully compliant.

Verified source location: **PDF p. 24; printed strategy p. 16**.

UPDF AI answered:

1. Direct answer: The Strategy's four key lines of effort are:
    - Defend the U.S. Homeland.
    - Deter China in the Indo-Pacific Through Strength, Not Confrontation.
    - Increase Burden-Sharing with U.S. Allies and Partners.
    - Supercharge the U.S. Defense Industrial Base.
2. Page number(s): PDF page 1.
3. Evidence: "The Department's strategic approach rests on the following key lines of effort (LOEs):"

Adobe Acrobat AI answered:

Reviewing the PDF for the four key lines of effort now.

1. The four key lines of effort in the Strategy's strategic approach are: Defend the U.S. Homeland; Deter China in the Indo-Pacific Through Strength, Not Confrontation; Increase Burden-Sharing with U.S. Allies and Partners; and Supercharge the U.S. Defense Industrial Base.
2. Supported on page 15 of the PDF.
3. Quoted evidence: "The Department's strategic approach rests on the following key lines of effort (LOEs): 1. Defend the U.S. Homeland 2. Deter China in the Indo-Pacific Through Strength, Not Confrontation 3. Increase Burden-Sharing with U.S. Allies and Partners 4. Supercharge the U.S. Defense Industrial Base"

T3 Challenge — verdict: Acrobat

Acrobat AI is more complete on the explicit China boundary. UPDF AI better preserves diplomatic detail. Both have one valid printed-page citation and one invalid introduction citation.

Verified source location: **PDF pp. 12 and 26**; printed strategy **pp. 4 and 18**.

UPDF AI answered:

1. Direct answer: The Strategy's Indo-Pacific objective is to establish a balance of power that prevents China—or any other country—from dominating the United States or its allies, while allowing the United States, China, and other regional countries to enjoy "a decent peace." To pursue this objective, it prescribes:

    - A strong denial defense along the First Island Chain, supported by a capable Joint Force able to conduct devastating strikes and operations against targets anywhere in the world, including from the U.S. Homeland.
    - Greater contributions to collective defense by key regional allies and partners.
    - Military-to-military communication with the PLA focused on strategic stability, deconfliction, and de-escalation.
    - Sufficient military strength for the President to negotiate from a position of strength, while avoiding unnecessary confrontation.

    Boundary or condition: The conclusion holds on the condition that the United States is strong but not unnecessarily confrontational and that the resulting balance of power prevents any nation from dominating the United States or its allies.

2. Supporting page numbers: PDF pages 1 and 18.
3. Evidence: "Our goal is simple: To prevent anyone, including China, from being able to dominate us or our allies—in essence, to set the military conditions required to achieve the NSS goal of a balance of power in the Indo-Pacific that allows all of us to enjoy a decent peace."

Adobe Acrobat AI answered:

Okay, analyzing the introduction and Line of Effort 2 sections for the balance-of-power objective.

1. The Strategy's balance-of-power objective for the Indo-Pacific is to prevent any nation, including China, from dominating the region, thereby ensuring a balance of power that allows all countries—including the United States, China, and others—to enjoy a decent peace. To pursue this, the Strategy prescribes building and sustaining a strong denial defense along the First Island Chain (FIC), engaging in military-to-military communications with China to support strategic stability and de-escalation, and working closely with regional allies and partners to incentivize and enable greater collective defense contributions. The conclusion holds under the condition that the U.S. does not seek to dominate, humiliate, or strangle China, but rather aims for a peace that is acceptable to all parties.
2. Supported on pages 1–2 (Introduction) and 18 (Line of Effort 2) of the PDF.
3. Quoted evidence: "Our goal is simple: To prevent anyone, including China, from being able to dominate us or our allies—in essence, to set the military conditions

required to achieve the NSS goal of a balance of power in the Indo-Pacific that allows all of us to enjoy a decent peace. To that end, as the NSS directs, we will erect a strong denial defense along the First Island Chain (FIC)."

F10 — National Drug Control Strategy 2026

T1 Summary — verdict: UPDF

UPDF AI is decisively stronger: its detailed targets are grounded in the report. Acrobat AI's blanket claim that no quantitative targets, dates or units are stated is false (The last bullet directly contradicts substantial parts of the report. Generic coverage of the policy pillars cannot compensate for incorrectly claiming that the performance framework has no numbers).

Verified source location: **PDF pp. 3–7, 20–25, 80–89, 109, 114, 119, 135–155**. Performance numbers checked against tables.

UPDF AI answered:

- The Strategy sets an overarching goal to "Save American Lives by Reducing Fatal Overdoses from All Drugs" with a 2024 baseline of 79,384 drug overdose deaths and targets of 71,630 drug overdose deaths in 2026 and 60,000 drug overdose deaths in 2029, using CDC NVSS data and trend-analysis/Healthy People 2030 alignment.
- It designates fentanyl as a Weapon of Mass Destruction (Executive Order 14367) and describes designating certain cartels as Foreign Terrorist Organizations (EO 14157), enabling use of counter-terrorism authorities (e.g., material support statutes such as 18 U.S.C. § 2339B) and Treasury sanctions (EO 14059/OFAC actions) to target TCOs/FTOs and commercial facilitators (example: OFAC sanction of Guangzhou Tengyue Chemical Co., Ltd., September 2025).
- The Strategy's supply-side measures include suspending de minimis duty-free treatment for all countries (Executive Order 14324) to require Advance Electronic Data for low-value parcels, expanding CTPAT/trusted trader programs (baseline: 428 chemical/pharma companies in CTPAT in 2024; targets 445 in 2026 and 471 in 2029), and tracking tableting/encapsulating machines and parts via CBP ACE/AES and DEA reporting.
- The public-health/demand-side framework commits to prevention, treatment, recovery, and rescue: examples and metrics include a Prevention Framework (Appendix G), targets to increase percent of 12–17 year-olds not using illicit drugs from 84.9% (2024 baseline) to 85.9% (2026) and 88.3% (2029), expanding evidence-based treatment access (baseline 19.3% of people with SUD received treatment in 2024; targets 20.9% in 2026 and 23.4% in 2029), and increasing federally-funded naloxone kits distributed to states from 5,240,262 (2024 baseline) to 5,345,067 (2026) and 5,502,275 (2029).
- The Data/Performance systems and institutional architecture: ONDCP leads implementation with coordinators (e.g., Emerging Drug Threats Coordinator, Global Supply Reduction Coordinator, USIC, SLTAC, Demand Reduction Coordinator), a Performance Review System (PRS) linking chapter objectives to measures and 2- and 5-year targets (e.g., SUDORS jurisdiction reporting baseline 43 in 2024; targets 47 in 2026 and 50 in 2029), and an Appendix C Data Plan emphasizing AI, wastewater

monitoring, DEA/TOX, DOSE, NEMSIS, SUDORS, NFLIS, and other named federal data systems..

Adobe Acrobat AI answered:

Let me see what the PDF reveals about its key findings and terms.

- The strategy prioritizes disrupting the supply of illicit fentanyl and its precursors, with a focus on strengthening border security and international cooperation.
- There is a commitment to expanding access to evidence-based prevention, treatment, and recovery support services for individuals with substance use disorder.
- The document highlights the importance of reducing the supply of illicit drugs through enhanced law enforcement efforts and targeting drug trafficking organizations.
- The strategy emphasizes supporting public health and safety by increasing the availability of naloxone and other overdose reversal medications.
- Specific quantitative targets, date ranges, or units for reductions, expansions, or interventions are not stated in the document.

T2 Evidence Q&A — verdict: Tie (both weak)

Neither is complete: both omit the 80.6% gap. UPDF AI is stronger on the need denominator; Acrobat AI provides a usable, though non-numeric, locator. Both should cite p. 160.

Verified source location: **PDF p. 160; printed p. 160**.

UPDF AI answered:

1. Direct answer: The 2024 NSDUH estimated that 52.6 million people, or 18.2% of people aged 12 or older, needed substance use treatment. Of these, 10.2 million, or 19.4%, received treatment. This estimate is limited to treatment received in specified settings: inpatient or outpatient facilities, telehealth, prisons, jails, juvenile detention centers, or medications for alcohol or opioid use disorder received in any setting, including general medical settings such as primary care or hospitals.
2. Page number(s): The relevant page number is not stated in the provided excerpt.
3. Evidence: “People who are prescribed MOUD or may need substance use treatment is estimated to be 52.6 million people, or 18.2 percent of people aged 12 or older.”

Adobe Acrobat AI answered:

Reviewing the PDF for specific 2024 NSDUH treatment statistics now.

1. According to the 2024 NSDUH figures cited in the National Treatment Plan, an estimated 52.6 million people aged 12 or older needed substance use treatment; only 10.2 million (19.4%) received it, and this estimate is limited to treatment received in specialty clinical settings as defined by NSDUH.
2. Supported on page: Appendix F: National Treatment Plan (see paragraph beginning "According to the most recent National Survey on Drug Use and Health (NSDUH)...").

3. Quoted evidence: "The higher number reflects the addition of people who take medications for opioid use disorder. Of those, the NSDUH reports that only 10.2 million, or 19.4 percent received treatment for an alcohol or drug use disorder. NSDUH definition of treatment is limited to specialty care received in specific clinical settings: an inpatient or outpatient facility; via telehealth; in a prison, jail, or juvenile detention center; or through medications for alcohol or opioid use disorder received in any setting, including general medical settings such as primary care or hospitals."

T3 Challenge — verdict: UPDF

UPDF AI is substantially more reliable but incomplete. Acrobat AI supplies three incorrect prevalence counts and a fabricated quotation; high topic coverage does not compensate for these errors.

Verified source location: **PDF pp. 160 and 162; printed pp. 160 and 162**.

UPDF AI answered:

1. Direct answer: In 2024, for the first time, the number of Americans with a drug use disorder surpassed the number with an alcohol use disorder. The shift was driven primarily by increased rates of marijuana use and addiction. The PDF gives cannabis use disorder as affecting 20.6 million, or 7.1 percent, of Americans over the age of 12 in 2024. The conclusion is bounded to Americans over the age of 12 and to 2024; the PDF also states that the comparison was based on the 2024 NSDUH.
2. Supporting page numbers: The exact PDF page numbers are not stated in the provided excerpt. The relevant passages are in the National Treatment Plan section under “The Treatment Gap” and “Awareness.”
3. Quoted evidence: “In 2024, for the first time, the number of Americans with a drug use disorder surpassed the number with an alcohol use disorder.”

Adobe Acrobat AI answered:

Okay, I'm analyzing the PDF to combine the statistics and discussion for your question.

1. In 2024, the number of Americans with a drug use disorder surpassed those with an alcohol use disorder for the first time, primarily due to an increase in cannabis use disorder; the exact prevalence given for cannabis use disorder is 7.7 million people. This conclusion holds under the NSDUH's definition of substance use disorder and its methodology for estimating prevalence.
2. Supported on pages: Appendix F: National Treatment Plan (introductory statistics and subsequent discussion).
3. Quoted evidence: "For the first time, the number of Americans with a drug use disorder (29.5 million) exceeded those with an alcohol use disorder (28.8 million), driven primarily by an increase in cannabis use disorder (7.7 million)."

F11 — Scanned sample document (OCR)

T1 Summary — verdict: UPDF

UPDF AI is stronger: it preserves the subgroup-analysis limitation and keeps the rhinitis study distinct from the unrelated extract on the second page.

Acrobat AI: The PDF contains a separate, unrelated APP staffing extract. Conflating it with the allergy-app study misattributes the material, although the text is genuinely in this PDF

Verified source location: Scanned PDF **pp. 1–2**, visually checked; no printed page numbers.

UPDF AI answered:

- The pilot study included 1,136 users who filled in 5,789 days of VAS, which allowed comparisons among outcomes but not subgroup analyses.
- The App collected country, language, age, sex and date of entry of information and used very simple questions translated and back-translated into 15 languages.
- Over 98% of users reported "AR", the App was not designed to compare AR patients with control subjects, and none of the users was enrolled in a clinical study.
- The study period for included consecutive users was from June 1, 2016 to October 31, 2016, and there was no specific advertisement or other recruitment campaign.
- The PDF does not state detailed subgroup analysis results, exact baseline characteristic proportions, or specific measures of work productivity in rhinitis beyond noting that none assessed work productivity prior to this and that the App study itself did not assess non-AR users.

Adobe Acrobat AI answered:

Let me see what the PDF reveals about its key findings.

- The pilot study used a mobile app to assess allergy symptoms and work productivity in rhinitis, involving 1,136 users who filled in 5,789 days of visual analogue scale (VAS) data.
- All consecutive users from June 1, 2016 to October 31, 2016 were included, with demographic data collected on country, language, age, sex, and date of entry; no personal identifiers were gathered due to anonymization.
- The app questions were translated and back-translated into 15 languages, and the study was not a clinical trial; over 98% of users reported "AR" (allergic rhinitis), and responses from non-AR users could not be assessed.
- There was no specific advertisement or recruitment campaign; most users downloaded the app from app stores or internet sources, with a few clinic patients asked by physicians to use it.
- The study also discusses advanced practice providers (APPs) in emergency departments, noting that greater APP coverage allows physicians to focus on higher-acuity cases, but APP use did not result in economies of scale due to higher physician productivity; implications for APP training and quality assurance are noted.

T2 Evidence Q&A — verdict: UPDF

UPDF AI wins on instruction compliance and location. Both accurately recover the scanned figures and the subgroup-analysis limitation.

Acrobat AI: Does not provide the required PDF page number; absence of a printed page label is not a reason to omit physical page 1.

Verified source location: **PDF p. 1**; no printed page number.

UPDF AI answered:

1. Direct answer: The pilot study included 1,136 users and 5,789 days of VAS entries. The authors said these data supported comparisons among outcomes but did not support subgroup analyses.
2. Page number(s): Page 1.
3. Evidence: “This pilot study was based on 1,136 users who filled in 5,789 days of VAS allowing us to perform comparisons among outcomes, but not to make subgroup analyses.”.”

Adobe Acrobat AI answered:

Reviewing the PDF for sample size, observation volume, and analysis details.

1. The pilot study reported a sample size of 1,136 users and an observation volume of 5,789 days of VAS data; the authors stated that the data allowed for comparisons among outcomes but did not support subgroup analyses.
2. This information is supported on the page discussing the pilot study's methodology and baseline characteristics (exact page number not stated in the PDF text provided).
3. Quoted evidence: "This pilot study was based on 1,136 users who filled in 5,789 days of VAS allowing us to perform comparisons among outcomes, but not to make subgroup analyses."

T3 Challenge — verdict: UPDF

UPDF AI is fully compliant. Acrobat AI reads the date correctly but omits the requested page number.

Verified source location: **PDF p. 1**; no printed page number.

UPDF AI answered:

- Exact inclusion date range: “from June 1, 2016 to October 31, 2016”
- Page: 1
- Readability: No part of this date range is unreadable.

Adobe Acrobat AI answered:

Okay, I'm checking the 'Demographic Characteristics' section for the inclusion date range.

The exact study inclusion date range stated under “Demographic Characteristics” is: "All consecutive users from June 1, 2016 to October 31, 2016 were included in the study." This appears on the page under the heading "Demographic Characteristics." No part of this statement is unreadable.

F12 — Scanned multi-page document (OCR)

T1 Summary — verdict: UPDF

UPDF AI is closer to the study, but it loses a minus sign and a percent unit. Acrobat AI has more substantial errors in the testing sequence, Kappa maximum, "all higher" qualification and claim that generalisability is not discussed.

Verified source location: Scanned PDF **pp. 1–3, 11, 13–20, 23**; **printed journal page = PDF plus 191**.

UPDF AI answered:

- The study developed a parsimonious, 38-item instrument (with a suggested short 25-item version) comprising eight scales—Voluntariness, Image, Relative Advantage, Compatibility, Ease of Use, Trialability, Result Demonstrability, and Visibility—to measure individuals' perceptions of adopting an information technology innovation, tested across three field tests including a final field test with 540 usable questionnaires returned (response rate 68%).
- Scale reliabilities across pilot and field tests show Cronbach's ALPHA and Guttman's Lower Bound (GLB) results: e.g., Relative Advantage ALPHA = 0.95 (GLB = 0.95) in Field Test Sample #1 (n = 270), Voluntariness ALPHA = 0.82–0.93 across tests, and Trialability near the lower bound (GLB = 0.72) in the field test; one scale's reliability is noted as relatively lower than the desired 0.80 standard but improvable by re-including deleted items.
- Factor analysis (Principal Components with VARIMAX) indicated a seven-factor solution accounting for approximately 63% of variance, with Relative Advantage and Compatibility loading together (correlated at 0.99 in an eight-factor test), and Observability splitting into Result Demonstrability and Visibility during scale development.
- Adopters (n = 418) versus nonadopters (n = 122) comparisons using Mann-Whitney U-tests found significant differences on all scales with reported medians/means such as Relative Advantage: Adopters 5.9 vs Nonadopters 4.8 (U-Test Z = 9.39, significance 0.0000), and Voluntariness: Adopters 3.3 vs Nonadopters 5.0 (U-Test Z = -10.01, significance 0.0000).
- A discriminant function using all perceptions (excluding Voluntariness) correctly classified 85 of cases [x2(7) = 132, p < 0.001]; the best predictors for distinguishing adopters from nonadopters were Relative Advantage, Result Demonstrability, and Visibility, while Trialability and Image were weak predictors in this organizational context.

Adobe Acrobat AI answered:

Let me see what the PDF reveals about its key findings and terms.

- The study developed and validated a measurement instrument for perceptions of adopting information technology innovations, resulting in a final 38-item instrument (with a suggested short 25-item version) covering eight scales: voluntariness, image, relative advantage, compatibility, ease of use, result demonstrability, trialability, and visibility.

- The instrument's reliability was confirmed through multiple rounds of judge sorting (with Cohen's Kappa scores up to 0.86 and placement ratios up to 100%), three pilot tests, and a field test with 540 usable responses (68% response rate) from seven companies, with most scale reliabilities at or above 0.80 (Cronbach's Alpha), except trialability and visibility, which were near 0.70.
- Factor analysis showed that seven factors explained approximately 63% of the variance, but compatibility and relative advantage did not emerge as separate empirical factors, despite being conceptually distinct in sorting rounds; their correlation was 0.99.
- Adopters of personal work stations (PWS) scored significantly higher than nonadopters on all perception scales (e.g., relative advantage: 5.9 vs. 4.8; compatibility: 5.7 vs. 4.4; all differences significant at $p < 0.05$ or better), and discriminant analysis correctly classified 85% of cases based on these perceptions.
- The document does not state any specific date range for the field tests, nor does it provide information on the instrument's application outside the context of PWS adoption in organizations; generalizability to other innovations or contexts is not stated and requires further investigation.

T2 Evidence Q&A — verdict: UPDF

UPDF AI is more focused and better aligned with the first-page scope. Both get the four required facts right; both exceed the one-sentence quotation format.

Verified source location: **PDF p. 1; printed journal p. 192**.

UPDF AI answered:

1. Direct answer: The validation sequence involved four rounds of item sorting by judges to establish scale membership, followed by three separate field tests and validation via factor analysis and discriminant analysis; the final instrument was a parsimonious 38-item instrument comprising eight scales.
2. Page number(s): page 1
3. Quoted evidence: "For this study, both newly created and existing items were placed in a common pool and subjected to four rounds of sorting by judges ... Scales for the resulting constructs were subjected to three separate field tests. ... The result is a parsimonious, 38-item instrument comprising eight scales."

Adobe Acrobat AI answered:

Reviewing the first page for validation sequence and instrument details.

1. The first page describes a validation sequence involving four rounds of judge sorting to establish convergent and discriminant validity, followed by three separate field tests and further checks using factor and discriminant analysis; the final instrument consists of 38 items grouped into eight scales (voluntariness, image, relative advantage, compatibility, ease of use, result demonstrability, trialability, and visibility), with a suggested short version of 25 items.
2. Page 1 of the PDF.
3. "For this study, both newly created and existing items were placed in a common pool and subjected to four rounds of sorting by judges to establish which items should be

in the various scales... Scales for the resulting constructs were subjected to three separate field tests... The result is a parsimonious, 38-item instrument comprising eight scales which provides a useful tool for the study of the initial adoption and diffusion of innovations. A short, 25-item, version of the instrument is also suggested."

T3 Challenge — verdict: Tie

Tie on required content and citation. UPDF AI is the cleaner, more concise response.

Verified source location: **PDF p. 1; printed journal p. 192**.

UPDF AI answered:

- Article title: “Development of an Instrument to Measure the Perceptions of Adopting an Information Technology Innovation”
- Authors: Gary C. Moore and Izak Benbasat
- Page: 1
- Unreadable parts: None.

Adobe Acrobat AI answered:

Okay, extracting the article title and author names from the first page.

1. The exact article title printed on the first page is: **"Development of an Instrument to Measure the Perceptions of Adopting an Information Technology Innovation"**. The two author names printed are **Gary C. Moore** and **Izak Benbasat**.
2. Page 1 of the PDF.
3. Quoted evidence: "Development of an Instrument to Measure the Perceptions of Adopting an Information Technology Innovation" and "Gary C. Moore... Faculty of Management The University of Calgary... Izak Benbasat... Faculty of Commerce and Business Administration University of British Columbia...""