

Supplementary information

Supplementary Table 1. Scoring outputs under three weighting schemes for the three pilots with relatively uniform domain profiles.

Pilot	Equal weights	Expert weights	AHP weights	Max gap
SMT	78.75	79.00	78.95	0.25
KPIN	80.00	80.50	80.82	0.82
Ambient Documentation	71.25	71.50	71.71	0.46

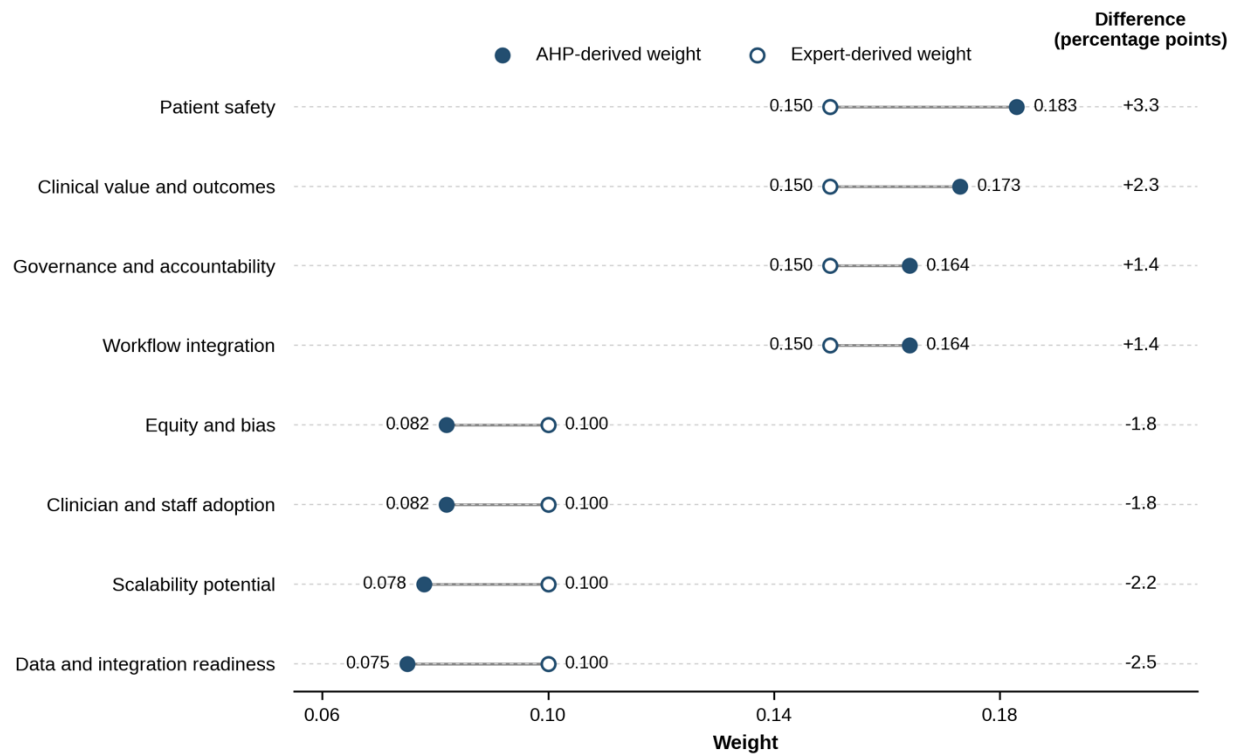
Supplementary Table 2. Monte Carlo summary under C-TACT expert weights (10,000 draws per pilot, seed 42).

Pilot	Base score	MC mean	5th–95th	P(≥ 70)	P(≥ 85)
SMT	79.0	78.8	72.5 – 85.0	0.99	0.06
KPIN	80.5	80.4	74.0 – 86.3	1.00	0.13
Ambient Documentation	71.5	71.5	64.9 – 77.9	0.65	0.00

Supplementary Table 3. Illustrative Retrospective Governance Decision Pathways for Comparator Technologies and Build-versus-Buy Outcomes

Comparator Technology Category	Reviewer Weighted Scores (Mean Range, Standard deviation)	Example Governance Question Triggered	Organizational Decision	Reported Outcome Following Decision
Comparator A	Mean: 51.1 (Range 44.0 – 55.0; SD = 4.8)	Are the safety and clinical-value limitations architectural or remediable?	Build	SMT developed internally. Published 77.2% reduction in time-to-first-read and 81.0% accuracy vs. 44.0% baseline.
Comparator B	Mean: 50.9 (Range: 41.5 – 57.0, SD = 6.7)	Can academic prototype classifiers meet enterprise governance and workflow requirements?	Build (confirmed)	Academic baselines served as validation benchmarks, not deployment candidates. SMT superiority confirmed against these comparators.
Comparator C	Mean: 54.5 (Range 45.0 – 66.0, SD = 9.0)	Is the platform scalable yet misaligned with integrated VBC workflows?	Build (replace)	Extended pilot identified VBC alignment limitations. Subsequent platform retirement supported replacement. KPIN achieved AUC 0.977 and 53.68% adjusted booking rate.
Comparator D	Mean: 33.5 (Range 10.0 – 52.5, SD = 17.6)	Does the disagreement indicate absence of a viable commercial solution?	Build (novel)	No commercial platform met integrated VBC requirements. KPIN exceeded 30% industry benchmark with 53.68% adjusted booking rate.

Abbreviations: VBC = value-based care; NLU = natural language understanding; CAS = Clinical Alert System; SMT = Smart Messaging Tool; KPIN = Kaiser Permanente Intelligent Navigator; AUC = area under the receiver operating characteristic curve.



Supplementary Figure 3. Comparison of AHP-derived and expert-derived weights across Pilot Score Card domains. Filled and open circles represent AHP-derived and expert-derived weights, respectively. Lines connect paired estimates, with differences shown in percentage points.

Supplementary Data 1. Vendor Evaluation Score Card (Excel)

Supplementary Data 2. Pilot Score Card (Excel)

Supplementary Data 3. Scale Score Card (Excel)

Supplementary Data 4. Maintenance Score Card (Excel)