

AI-Driven Prioritization in Enterprise Product Management: A Framework Integrating Customer Insights and Business Value

Sreehari Prasad Jinka Venugopal

Abstract: Enterprise product managers must continually decide which features, initiatives and fixes deserve scarce engineering capacity. Prevailing scoring methods such as RICE, WSJF and MoSCoW depend on subjective estimates, are easily dominated by internal opinion, and treat customer evidence and business value as two separate conversations. This paper proposes the Integrated Customer-Value Prioritization Framework (ICVPF), a five-layer architecture in which AI competencies (infrastructure, business spanning and proactive stance) are converted into three product-management capabilities (information management, planning and implementation) that ultimately support customer lifetime value (CLV). ICVPF mines multichannel customer feedback with language models to produce a calibrated Customer Insight Score (CIS), estimates a Business Value Score (BVS) from CLV uplift, strategic alignment and risk, fuses the two through a learned composite value estimate, and allocates capacity with a constrained knapsack optimization wrapped in human oversight and explainability controls. Seven hypotheses (H1–H7) are derived from the conceptual model for future field validation. As a first test of the decision logic, we ran a Monte Carlo simulation of 1,000 planning cycles with 120 backlog items each. The integrated model captured 95.3% of the oracle value against 87.1% for manual RICE-style scoring (a 9.4% relative gain; paired t-test, $p < 0.001$), and its rank correlation with realized value was 0.872 versus 0.643. These results are synthetic: they demonstrate the internal logic of the framework, not field performance. We close with governance guidance and a research agenda.

Introduction

Every enterprise product organization faces the same structural problem: demand for change is effectively unlimited, while engineering, design and go-to-market capacity is fixed. A backlog of several hundred candidate items is normal in a mature B2B software company, and each planning cycle forces a small subset to be funded at the expense of everyone else. The quality of that selection is one of the most consequential and least measured decisions in the firm. A mistaken choice does not merely waste effort; it displaces a better alternative, delays learning, and erodes customer trust when a promised capability never arrives.

Practitioners have long relied on lightweight scoring models to bring discipline to this decision. RICE (reach, impact, confidence, effort), Weighted Shortest Job First (WSJF), MoSCoW and the Kano model are all popular because they are simple and transparent. Yet they share three weaknesses. First, their inputs are human estimates, and these estimates are noisy, anchored on recent events and biased toward whichever stakeholder speaks loudest.

MBA, Product Management

Enterprise Product Management, Data Protection, Cyber Resilience, Trustworthy AI Governance
Phoenix, Arizona, USA

mail09hari@gmail.com, ORCID: 0009-0002-5969-9357

Keywords: Product Management, AI-Driven Prioritization, Customer Insights, Business Value, Customer Lifetime Value, Decision Support, Enterprise Software

Second, the evidence base is thin: a product manager typically reads a few dozen support tickets or interview notes, while the organization actually holds thousands of signals in tickets, call transcripts, usage telemetry, community forums, sales notes and contract renewals. Third, customer value and business value are usually handled in different forums. Customer success argues for what accounts are asking for, finance argues for what moves revenue, and the roadmap becomes the outcome of negotiation rather than analysis.

Recent advances in machine learning and language models make it feasible to address all three weaknesses. Transformer-based models can classify and summarize large volumes of unstructured feedback, predictive models can estimate the revenue and retention consequences of changes, and optimization methods can allocate capacity under constraints. However, adopting AI in a prioritization setting is not simply a matter of plugging a model into a spreadsheet. Enterprise decisions are high-stakes, accountable and political. Models that are opaque, miscalibrated or trained on unrepresentative feedback can produce confident but wrong rankings, and organizations that lack data infrastructure, cross-functional integration or a proactive stance toward AI are unlikely to realize its benefits.

This paper therefore treats AI-driven prioritization as an organizational capability and a technical architecture at the same time. The primary objective is to develop and test the logic of the Integrated Customer-Value Prioritization Framework (ICVPF). The specific tasks are: (i) to position AI use in product management within a capability taxonomy based on coverage and complementarity; (ii) to specify a conceptual model linking AI competencies, product-management capabilities and customer lifetime value; (iii) to formalize a Customer Insight Score and a Business Value Score and fuse them into a single decision metric; (iv) to define a constrained allocation procedure with governance controls; and (v) to evaluate the decision logic against conventional scoring in a controlled simulation. The contribution is a coherent, testable and auditable blueprint that reduces dependence on opinion without removing human accountability.

The remainder of the paper is organized as follows. Section II reviews the literature. Section III summarizes related work. Section IV presents the conceptual model and hypotheses. Section V describes the framework methodology. Section VI details the experimental setup, Section VII reports results, Section VIII discusses implications, Section IX addresses limitations, and Section X concludes.

Literature Review

A. Prioritization in product and requirements management

Prioritization research has deep roots in requirements engineering. Karlsson and Ryan proposed a cost-value approach in which stakeholders compare requirements pairwise on value and cost using the Analytic Hierarchy Process, an idea derived from Saaty's multi-criteria decision theory [4], [5]. Achimugu and colleagues later surveyed the field and found a long list of techniques (numerical assignment, ranking, bubble sort, hierarchy analysis, planning games and others), together with recurring problems of scalability, dependence on stakeholder judgment and weak handling of uncertainty [6]. Cooper's stage-gate work added a portfolio view in which projects are screened against scorecards and strategic fit [7]. In agile practice, Reinertsen's cost-of-delay reasoning, popularized as WSJF, ranks work by the ratio of delay cost to duration [8], while RICE and MoSCoW offer simple heuristics for product teams.

These methods are valuable because they make criteria explicit. Their limitation is the quality and volume of input data. Pairwise comparison scales poorly, since the number of comparisons grows quadratically with the number of items. Heuristic scores depend on a single person's estimate of reach and impact, which is difficult to verify. Almost none of the classical methods describe how continuous customer evidence should enter the score, how that evidence should be weighted by commercial importance, or how the resulting score should be validated against realized outcomes.

B. Customer insight as a decision input

The Kano model distinguishes must-be, one-dimensional and attractive qualities and explains why satisfaction does not rise linearly with feature count [9]. Net-promoter style metrics and voice-of-customer programs aim to link sentiment to growth [10]. At scale, customer insight has become a text-mining problem: Liu surveys sentiment analysis and opinion mining [11], topic models such as latent Dirichlet allocation uncover themes in large corpora [12], and pre-trained transformers such as BERT have greatly improved classification and extraction from short, noisy text [13], [14]. Applied to product feedback, these tools can cluster requests, estimate how many distinct accounts mention a problem, and gauge the intensity of dissatisfaction. Two cautions are important. Raw frequency over-represents vocal, accessible customers, and sentiment scores are unreliable in domain-specific language. Insight therefore has to be weighted by account value and calibrated against outcomes before it can serve as a priority signal.

C. Business value and customer lifetime value

Business value in enterprise settings is multi-dimensional: revenue expansion, retention, cost reduction, strategic positioning and risk. Customer lifetime value (CLV), the discounted stream of margin expected from a customer relationship, provides a unifying quantity because it links product decisions to retention and expansion [15], [16]. Reinartz and Kumar showed that long-lived customers are not always the most profitable, which implies that value must be modelled per segment rather than assumed [16]. Treating CLV uplift as the common currency lets a prioritization model compare a retention fix for a strategic account with a new capability aimed at prospects on one scale.

D. AI capability and the complementarity of humans and machines

Research on AI capability emphasizes that value comes from organizational resources and not from algorithms alone. Mikalef and Gupta conceptualize AI capability as a bundle of tangible resources (data, technology, basic resources), human skills and intangible resources, and show its relationship with creativity and performance [17]. Davenport and Ronanki observe that enterprise AI projects succeed when they target well-defined business processes and augment decisions [18]. Jarrahi argues that human and machine strengths are complementary, with machines handling complexity and humans handling ambiguity and judgment [19], and Raisch and Krakowski describe the automation-augmentation paradox, in which the two are intertwined and cannot be managed as alternatives [20].

A useful way to apply this thinking to product management is to classify AI-enabled tasks along two dimensions: *coverage*, the share of a task that AI can currently handle, and *complementarity*, the extent to which human contribution adds to AI's. This produces four capability types (Figure 1). *AI-substitutive* capabilities have high coverage and low complementarity, and include data entry, assembly-style routine work and customer-service-center handling. *AI-integrated* capabilities have high coverage and high complementarity, and include R&D

design, radiologic diagnosis and data analysis. *AI-augmented* capabilities have low coverage but high complementarity, such as psychotherapy, anomaly handling and monitoring and maintenance. *AI-assisted* capabilities have low coverage and low complementarity, such as decision-making, geriatric care and reception services.

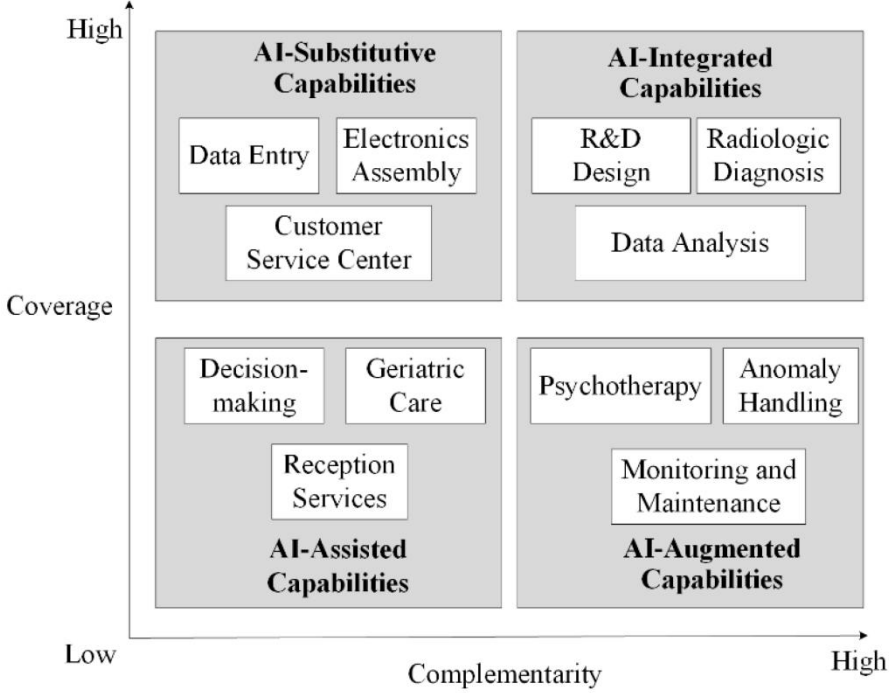


Figure 1. Classification of AI capabilities by coverage (vertical axis) and complementarity (horizontal axis).

Product prioritization sits most naturally in the AI-integrated quadrant. Analyzing thousands of signals is a high-coverage task suited to machines, while interpreting strategy, negotiating trade-offs and carrying accountability are complementary human strengths. Treating prioritization as AI-substitutive would remove judgment where it is most needed, whereas treating it as merely AI-assisted would waste the available analytical capacity. This positioning underpins the human-in-the-loop design of ICVPF.

Related Work

Several strands of work approach AI-supported prioritization from different directions. In requirements engineering, researchers have applied machine learning to classify requirements, predict their implementation effort, and cluster user-story backlogs. Search-based software engineering formulates the next-release problem as a constrained optimization in which a subset of requirements is chosen to maximize stakeholder satisfaction under a cost budget, a formulation that anticipates the knapsack structure used here. These studies typically take stakeholder value as given and do not model where the value estimate comes from.

In the app-review and user-feedback literature, classifiers extract feature requests, bug reports and sentiment from public reviews, and some work links review volume to subsequent rating changes. These studies show that unstructured feedback holds actionable signal, yet they usually stop at classification. They rarely connect the signal to account value, to revenue outcomes or to a capacity-constrained plan, and they are rarely evaluated against realized value.

In the management literature, AI capability studies relate organizational resources to performance [17], and work on marketing capabilities relates information management, planning and implementation to business outcomes. A recurring finding is that capabilities mediate the effect of technology on performance: technology does not create value by itself, but it improves how firms manage information, plan and execute. We draw on this mediation logic to structure the conceptual model below.

Finally, the explainable-AI literature stresses that high-stakes decision support must be interpretable and auditable. Post-hoc attribution methods such as LIME and SHAP explain individual predictions [21], [22], while Rudin argues that where possible, inherently interpretable models should be preferred in high-stakes settings [23]. We adopt this position by using a low-dimensional, interpretable value model at the core of ICVPF.

Taken together, existing work leaves a gap. No widely used approach jointly (a) converts large-scale customer evidence into an account-weighted score, (b) expresses business value on a CLV basis, (c) learns how to combine them from realized outcomes, (d) allocates capacity optimally under constraints, and (e) embeds the organizational capabilities and governance needed to operate it. ICVPF is designed to fill this gap.

Conceptual Model and Hypotheses

Figure 2 presents the conceptual model. AI competencies is a second-order construct formed by three first-order dimensions: infrastructure (data platforms, integration, compute and tooling), business spanning (the degree to which technical and business staff share understanding and collaborate), and proactive stance (leadership willingness to experiment with and invest in AI ahead of competitors). AI competencies influence three product and B2B marketing capabilities: information management (collecting, integrating and interpreting market and customer information), planning (translating information into priorities, roadmaps and resource allocation), and implementation (executing plans through delivery, launch and customer engagement). These capabilities in turn affect the outcome construct, customer lifetime value, and a direct path from AI competencies to CLV is also specified.

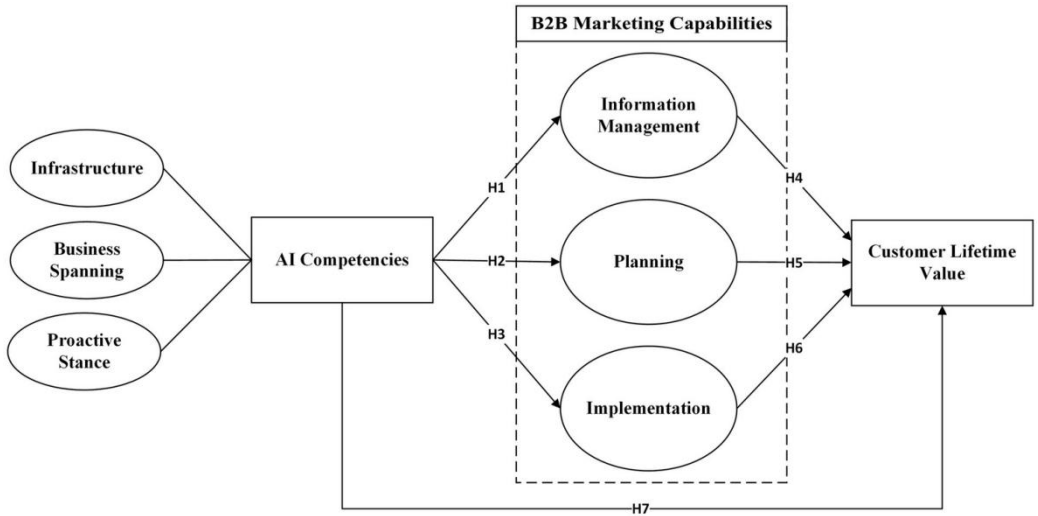


Figure 2. Conceptual model linking AI competencies, B2B marketing capabilities and customer lifetime value (H1–H7).

The hypotheses follow the paths in the figure:

H1: AI competencies are positively related to information management capability.

H2: AI competencies are positively related to planning capability.

H3: AI competencies are positively related to implementation capability.

H4: Information management capability is positively related to customer lifetime value.

H5: Planning capability is positively related to customer lifetime value.

H6: Implementation capability is positively related to customer lifetime value.

H7: AI competencies are directly and positively related to customer lifetime value.

The rationale is as follows. Better infrastructure and cross-functional understanding allow an organization to collect and interpret more customer evidence (H1), turn it into more defensible priorities (H2) and deliver against them more reliably (H3). Each capability then improves the customer experience and commercial outcomes, which accumulate as higher CLV (H4–H6). A direct effect (H7) is plausible because AI competencies may also influence value through channels not captured by the three capabilities, such as pricing intelligence or churn prediction. Because these are conceptual hypotheses, this paper does not test them statistically. They define the organizational preconditions for ICVPF and specify the survey or panel study needed to validate it in the field, as discussed in Section IX.

Methodology: The ICVPF Framework

A. Architecture overview

ICVPF has five layers that map onto the three capabilities in the conceptual model. Layer 1 (signal ingestion) and Layer 2 (insight extraction) implement information management. Layer 3 (value modelling) and Layer 4 (fusion and allocation) implement planning. Layer 5

(governance and learning) implements the oversight and feedback needed for dependable implementation. The flow is cyclical: realized outcomes from each delivery cycle are fed back to recalibrate the models used in the next.

Layer 1: Signal ingestion. The system consolidates customer-facing and internal evidence: support tickets, chat and call transcripts, community posts, product telemetry, win/loss notes, renewal and churn records, and CRM data on account value. Each record is linked to an account identifier and a segment so that evidence can later be weighted by commercial importance. Personally identifiable information is masked at ingestion.

Layer 2: Insight extraction. A language model maps each unstructured record to a controlled taxonomy of need statements. A pre-trained transformer encoder produces embeddings, near-duplicate statements are clustered into *need clusters*, and each cluster is linked to backlog items through semantic similarity and human confirmation. For each cluster the system derives distinct-account demand, sentiment-weighted pain intensity, and association with churn or contraction events.

Layer 3: Value modelling. For each backlog item, a business model estimates incremental CLV, strategic alignment and risk, as formalized below. Estimates carry uncertainty intervals.

Layer 4: Fusion and allocation. The Customer Insight Score and Business Value Score are combined by a learned value function, and capacity is assigned by a constrained optimization that respects effort budgets and dependencies.

Layer 5: Governance and learning. Explanations, override logging, fairness checks, drift monitoring and post-release outcome measurement keep the system accountable and allow it to improve.

B. Customer Insight Score

Let backlog item i be linked to need cluster $k(i)$. Raw mention counts over-represent large accounts that generate many tickets and under-represent quiet but valuable ones, so demand is weighted by segment value. With m_{is} denoting the number of distinct accounts in segment s that express the need and n_s the segment size, the account-value-weighted demand is

$$d_i = \frac{\sum_s \text{CLV}_s \cdot (m_{is}/n_s)}{\sum_s \text{CLV}_s} \quad (1)$$

where CLV_s is the mean customer lifetime value of segment s . Pain intensity p_i is the mean sentiment-weighted severity of the need statements, rescaled to $[0,1]$ and calibrated against historical outcomes. Churn linkage r_i is the share of churned or contracting accounts in the preceding period that mentioned the need. Competitive pressure q_i is the share of lost deals in which the need appears. Each component is standardized (z-scored) across the backlog, and the Customer Insight Score is

$$\text{CIS}_i = \alpha_1 \cdot z(d_i) + \alpha_2 \cdot z(p_i) + \alpha_3 \cdot z(r_i) + \alpha_4 \cdot z(q_i) \quad (2)$$

with weights α estimated in Layer 4 rather than set by hand. Calibration is checked with reliability curves comparing predicted and realized adoption, and clusters supported by fewer

than a minimum number of distinct accounts are flagged as low-confidence rather than scored.

C. Business Value Score

The customer lifetime value of segment s is modelled as a discounted margin stream:

$$\text{CLV}_s = \sum_{t=0}^T \frac{m_s \cdot \rho_s^t}{(1 + \delta)^t} \quad (3)$$

where m_s is the periodic margin, ρ_s the retention probability, δ the discount rate and T the horizon. A feature can change CLV by shifting retention ($\Delta\rho$), expansion (Δm), or acquisition volume. The expected CLV uplift for item i is the sum, over affected segments, of the expected change in CLV multiplied by the number of affected accounts and the probability of success (confidence) c_i :

$$\Delta\text{CLV}_i = c_i \cdot \sum_s N_{is} \cdot (\text{CLV}'_s - \text{CLV}_s) \quad (4)$$

Strategic alignment SA_i is rated against explicit objectives (for example platform extensibility or regulatory readiness) on a bounded scale and is the one component that is deliberately human-assigned. Risk RISK_i captures technical, regulatory and delivery uncertainty. The Business Value Score is

$$\text{BVS}_i = z(\Delta\text{CLV}_i) + \lambda \cdot z(\text{SA}_i) - \mu \cdot z(\text{RISK}_i) \quad (5)$$

where λ and μ are set by leadership, because they encode strategy and risk appetite and should be explicit decisions rather than model outputs.

D. Fusion: learning how customer insight and business value combine

Customer demand and business value are correlated but not identical. A frequently requested item may have little commercial effect, while a niche request from strategic accounts may be decisive. Moreover, the two can interact, since an item that is both widely demanded and commercially significant tends to be worth more than the sum of its parts. We therefore estimate realized value with a small interpretable model fitted to historical items whose outcomes are known:

$$\hat{V}_i = \theta_0 + \theta_1 \cdot \text{CIS}_i + \theta_2 \cdot \text{BVS}_i + \theta_3 \cdot (\text{CIS}_i \times \text{BVS}_i) \quad (6)$$

The realized-value label V_i is a composite of post-release adoption, retention change and revenue influence over a fixed window. Parameters θ are estimated by regularized least squares [3], and ensembles or gradient-boosted trees [1], [2] may be substituted when enough history exists, at the cost of interpretability. Prediction uncertainty σ_i is estimated by bootstrapping, and decisions use a conservative lower confidence bound

$$\hat{V}_i^{\text{LCB}} = \hat{V}_i - \kappa \cdot \sigma_i \quad (7)$$

where $\kappa \geq 0$ expresses risk aversion. This discourages the model from favoring items whose apparent value rests on sparse or noisy evidence.

E. Capacity allocation

Given effort estimates e_i and a cycle budget B , selection is formulated as a 0–1 knapsack problem with precedence constraints:

$$\max_{\sum_i x_i} \sum_i x_i \cdot \hat{v}_i^{\text{LCB}} \text{ subject to } \sum_i x_i \cdot e_i \leq B, x_j \leq x_i \text{ for every dependency } (i \rightarrow j), x_i \in \{0,1\} \quad (8)$$

For realistic backlog sizes this can be solved exactly with integer programming, and a greedy ranking by value density $\hat{v}_i^{\text{LCB}}/e_i$ is a strong, transparent approximation that product managers can inspect. Additional constraints can reserve capacity for mandatory work (security, compliance) or protect a share of the budget for exploratory bets, which prevents the optimizer from starving long-term investment. Effort estimates are themselves uncertain; where historical estimation error is available, buffers are applied to e_i .

F. Human oversight, explainability and governance

Because prioritization is an AI-integrated task, ICPVF specifies how humans and machine interact. The model proposes; accountable product leaders decide. Each recommendation is accompanied by its component scores, the account segments driving demand, representative anonymized evidence, and the sensitivity of its rank to λ , μ and κ . Overrides are permitted and logged with a reason. Override outcomes are later compared with model recommendations, which turns disagreement into a source of learning and not of conflict.

Several safeguards address known failure modes. Representation checks compare the demographic and segment distribution of feedback sources with the customer base to detect silent segments. Drift monitors track shifts in language and in the relationship between scores and outcomes, triggering recalibration. Goodhart protections prevent teams from gaming inputs, for example by inflating ticket tagging, by auditing sample records and by rotating validation metrics. Finally, data governance enforces privacy, retention limits and access control for customer communications.

Experimental Setup

The purpose of the experiment is limited and should be read as such. Real product organizations do not observe the counterfactual value of items they did not build, so the decision logic of ICPVF cannot be benchmarked against an oracle in field data. We therefore use a controlled Monte Carlo simulation in which the true value of every backlog item is known, and ask whether each method, given realistically noisy inputs, selects a set of items that captures more of the attainable value. The simulation was implemented in Python (NumPy, SciPy) with a fixed random seed, and 1,000 independent planning cycles were run.

Data-generating process. Each cycle contains 120 backlog items. For item i , latent customer demand c_i is drawn from a standard normal distribution. Latent business value b_i is correlated with demand ($\rho = 0.4$) but not identical to it. Effort e_i follows a log-normal distribution (mean of the log = 1.0, standard deviation = 0.6). True realized value is

$$V_i = \max \left(0.05, 1 + 0.55 \cdot c_i + 0.55 \cdot b_i + 0.30 \cdot c_i \cdot b_i + 0.25 \cdot \varepsilon_i \right), \varepsilon_i \sim \mathcal{N}(0,1) \quad (9)$$

so that value depends on both signals and on their interaction, and contains irreducible randomness. The capacity budget is $B = 30\%$ of the total backlog effort, which forces substantial rejection.

Compared methods. All methods select items greedily by estimated value density (estimated value divided by effort) until the budget is exhausted. They differ only in how value is estimated.

- Manual scoring (RICE-style): estimated value equals true value plus Gaussian noise with standard deviation 1.0, plus an upward bias of 0.8 on a random 20% of items to represent loud-voice or recency effects.
- WSJF (value proxy): estimated value is $1 + 0.8 \cdot b_i$ plus noise with standard deviation 0.7, capturing a cost-of-delay view that considers business value only.
- Customer-only AI: estimated value uses the model-derived demand signal $\hat{c}_i = c_i + \text{noise}$ (standard deviation 0.5) and ignores business value.
- Business-only AI: estimated value uses $\hat{b}_i = b_i + \text{noise}$ (standard deviation 0.5) and ignores customer demand.
- Integrated AI (ICVPF): estimated value uses Equation (6), with θ fitted by ordinary least squares to a separate set of 200 historical items (with the same signal noise) whose outcomes are known. The conservative adjustment of Equation (7) was not applied ($\kappa = 0$), so results reflect the fusion step alone.

An oracle that ranks by true value density under the same budget provides the ceiling.

Metrics. The primary metric is the *value capture ratio* (VCR), the total true value of the selected set divided by that of the oracle’s set. Secondary metrics are the Spearman rank correlation between estimated and true value density, and *top-k recall*, the share of the true top 20% of items (by value density) that appear in a method’s top 20%. Methods are compared with a paired t-test over the 1,000 cycles.

Results

A. Overall decision quality

Table 1 reports the mean and standard deviation of the value capture ratio, mean rank correlation and mean top-k recall over 1,000 planning cycles. All figures are outputs of the simulation described above and are synthetic.

TABLE 1. Simulated Prioritization Performance (1,000 planning cycles, 120 items each)

Method	Value capture (% of oracle)	Std. dev. (pp)	Rank correlation with true value	Top-20% recall (%)
Manual scoring (RICE-style)	87.1	3.4	0.643	66.6
WSJF (value proxy)	84.2	3.8	0.636	60.8
Customer-only AI	89.2	3.0	0.736	67.6
Business-only AI	89.0	3.1	0.733	67.1
Integrated AI (ICVPF)	95.3	1.6	0.872	79.1

The integrated model captured 95.3% of the attainable value, compared with 87.1% for manual scoring, a relative improvement of 9.4%. The paired comparison was significant ($t = 72.9$, $df = 999$, $p < 0.001$), and the integrated approach outperformed manual scoring in 99.5% of cycles. The spread across cycles also narrowed: the standard deviation fell from 3.4 to 1.6 percentage points, indicating that quality became more consistent as well as higher. Rank correlation rose from 0.643 to 0.872, and the integrated model recovered 79.1% of the truly best items against 66.6% for manual scoring.

B. Contribution of each information source

Using either customer insight alone or business value alone yielded modest gains over manual scoring (about two percentage points in value capture each), and both single-source AI variants substantially improved rank correlation (0.736 and 0.733). The largest step, however, came from combining them: moving from the best single-source variant to the integrated model added roughly six more percentage points of value capture and 11 points of top-k recall. In this setting, each signal explains a different part of true value, so each alone leaves a large part unexplained, and fusing the two also averages out independent estimation noise. The learned interaction term lets the model reward items that are both demanded and commercially significant. WSJF with a business-value proxy performed worst, which is consistent with its reliance on a single dimension and with the absence of any customer evidence.

C. Illustrative interpretation for an enterprise planning cycle

To translate the numbers into planning terms, consider a quarter with a 100-person-week delivery budget. If manual scoring yields 87.1% of the attainable value and ICVPF yields 95.3%, the difference of 8.2 percentage points is equivalent to the value of roughly eight person-weeks of work, delivered with no additional capacity, simply by choosing better. This is a scale illustration only. In practice the gain depends on the data quality, the signal-to-noise ratio of customer evidence and the maturity of the underlying capabilities described in Section IV.

Discussion

The simulation supports the central design argument of ICVPF: prioritization improves when high-coverage analytical tasks are delegated to models while accountable humans retain decisions about strategy and risk. The gains came from three mechanisms. First, language-model-derived signals replace a small, biased sample of evidence with a broad, account-weighted one, which lowers the noise of the demand estimate. Second, expressing business value in CLV terms provides a second, partly independent estimate of worth. Third, a learned value function weights the two according to what actually predicted outcomes, instead of leaving their combination to negotiation.

The results also have organizational implications that connect to the conceptual model. Insight extraction (information management) is only as good as the data infrastructure beneath it, which is the AI-infrastructure dimension of H1. The fusion and allocation step (planning) requires finance, customer success and product to agree on a shared CLV model, which is the business-spanning dimension of H2. Realizing the plan (implementation) depends on delivery teams trusting recommendations enough to act on them, which is helped by the explainability and override mechanisms and by a proactive leadership stance. Seen this way, a prioritization tool does not create value by itself. It is the visible product of a set of capabilities, and organizations that lack those capabilities should expect the tool to under-deliver.

For practitioners we suggest an incremental path. Begin with Layers 1 and 2 on a single product line, using the model only to summarize and rank evidence and keeping scoring manual. Add account-weighting and the CLV model next, then introduce the learned fusion once at least a few cycles of outcome data exist. Retain the human decision at every stage, review overrides regularly, and track the value-capture gap between the model and human selections as a management metric.

Finally, the framework is deliberately conservative about autonomy. Prioritization determines which customers get what and which teams work on what, and errors may fall unevenly across segments. The representation checks, drift monitoring and logged overrides are therefore integral to the method and not optional extras. A model that raises average value while silently ignoring a small but strategically important segment would be a failure by the framework's own standards.

Limitations and Future Work

Several limitations qualify the findings. Most importantly, the evaluation is synthetic. The data-generating process was constructed so that realized value depends on customer demand, business value and their interaction, and the integrated model is correctly specified to exploit exactly that structure. The measured advantage is therefore best regarded as an indication of what is possible when both signals are informative and noise levels are as assumed, not as an expected field improvement. The noise levels assigned to manual and AI estimates (1.0 and 0.5) are assumptions and were not varied in a sensitivity analysis; real-world gains could be smaller if language-model signals are noisier, or if value is driven by factors that neither signal captures. The uncertainty-aware selection of Equation (7), dependency constraints and reserved-capacity rules were specified but not exercised in the simulation.

Second, the hypotheses H1–H7 are untested. A field validation would survey product and marketing organizations on AI competencies and capabilities, collect CLV or proxy outcomes, and estimate the structural model with partial least squares or covariance-based structural equation modelling, ideally with multi-informant data to reduce common-method bias.

Third, outcome labels in real settings are delayed, confounded and sparse. Estimating the realized value of an item requires attributing changes in adoption and retention to a single release among many. Quasi-experimental designs, staged rollouts and holdout groups should be used wherever possible.

Future work should therefore include: (i) sensitivity analyses over noise, correlation and interaction structure; (ii) a pilot in at least one enterprise product line with prospective comparison to the incumbent method; (iii) evaluation of explanation quality and its effect on trust and override behavior; (iv) fairness analysis across customer segments; and (v) extension to portfolio-level decisions that span multiple products and multi-quarter dependencies.

Conclusion

This paper presented the Integrated Customer-Value Prioritization Framework, an approach to enterprise product prioritization that joins account-weighted customer insight mined from unstructured feedback with a CLV-based estimate of business value, fuses them through an interpretable learned model, allocates capacity by constrained optimization, and surrounds the process with human oversight and governance. A

conceptual model linked AI competencies to information management, planning and implementation capabilities and to customer lifetime value, yielding seven testable hypotheses. In a controlled simulation of 1,000 planning cycles, the integrated model captured 95.3% of the attainable value against 87.1% for manual scoring, with markedly higher rank agreement and lower variability, and the combination of customer and business signals produced far more benefit than either alone.

The simulation is evidence of coherence and not of field effectiveness. The next step is empirical: pilot the framework against current practice, measure realized value, and test the capability hypotheses in real organizations. If those results are consistent with the logic shown here, AI can help product leaders replace contested opinion with auditable evidence while keeping the judgment and accountability that enterprise decisions require.

References

- [1] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2016, pp. 785–794.
- [2] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [3] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, 2nd ed. New York, NY, USA: Springer, 2009.
- [4] J. Karlsson and K. Ryan, "A cost-value approach for prioritizing requirements," *IEEE Software*, vol. 14, no. 5, pp. 67–74, 1997.
- [5] T. L. Saaty, *The Analytic Hierarchy Process*. New York, NY, USA: McGraw-Hill, 1980.
- [6] P. Achimugu, A. Selamat, R. Ibrahim, and M. N. Mahrin, "A systematic literature review of software requirements prioritization research," *Information and Software Technology*, vol. 56, no. 6, pp. 568–585, 2014.
- [7] R. G. Cooper, "Stage-gate systems: A new tool for managing new products," *Business Horizons*, vol. 33, no. 3, pp. 44–54, 1990.
- [8] D. G. Reinertsen, *The Principles of Product Development Flow: Second Generation Lean Product Development*. Redondo Beach, CA, USA: Celeritas Publishing, 2009.
- [9] N. Kano, N. Seraku, F. Takahashi, and S. Tsuji, "Attractive quality and must-be quality," *Hinshitsu (Quality)*, vol. 14, no. 2, pp. 39–48, 1984.
- [10] F. F. Reichheld, "The one number you need to grow," *Harvard Business Review*, vol. 81, no. 12, pp. 46–54, 2003.
- [11] B. Liu, *Sentiment Analysis and Opinion Mining*. San Rafael, CA, USA: Morgan & Claypool, 2012.
- [12] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003.
- [13] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. NAACL-HLT, 2019, pp. 4171–4186.
- [14] A. Vaswani et al., "Attention is all you need," in *Advances in Neural Information Processing Systems* 30, 2017.
- [15] S. Gupta et al., "Modeling customer lifetime value," *Journal of Service Research*, vol. 9, no. 2, pp. 139–155, 2006.
- [16] W. Reinartz and V. Kumar, "On the profitability of long-life customers in a noncontractual setting: An empirical investigation and implications for marketing," *Journal of Marketing*, vol. 64, no. 4, pp. 17–35, 2000.

- [17] P. Mikalef and M. Gupta, “Artificial intelligence capability: Conceptualization, measurement calibration, and empirical study on its impact on organizational creativity and firm performance,” *Information & Management*, vol. 58, no. 3, art. 103434, 2021.
- [18] T. H. Davenport and D. D. D.J. Ronanki, “Artificial intelligence for the real world,” *Harvard Business Review*, vol. 96, no. 1, pp. 108–116, 2018.
- [19] M. H. Jarrahi, “Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making,” *Business Horizons*, vol. 61, no. 4, pp. 577–586, 2018.
- [20] S. Raisch and S. Krakowski, “Artificial intelligence and management: The automation–augmentation paradox,” *Academy of Management Review*, vol. 46, no. 1, pp. 192–210, 2021.
- [21] M. T. Ribeiro, S. Singh, and C. Guestrin, “‘Why should I trust you?’: Explaining the predictions of any classifier,” in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, 2016, pp. 1135–1144.
- [22] S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” in *Advances in Neural Information Processing Systems* 30, 2017.
- [23] C. Rudin, “Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead,” *Nature Machine Intelligence*, vol. 1, no. 5, pp. 206–215, 2019.