A PRACTICAL IMPLEMENTATION GUIDE

Build the evaluation mechanism before you scale the automation

The goal is not to ask an AI for an answer once. The goal is to express your decision policy as a testable prompt, compare it with real professional judgment, and improve both the prompt and the review process.

Core principle: Treat the prompt as a version-controlled decision policy. Require an explanation for every score, keep humans responsible for final action, and test on representative data before applying the process to the full portfolio.

1

STEP 1

Use AI prompts to build the evaluation mechanism

Start in plain English. Describe how a knowledgeable reviewer should evaluate an asset, the evidence the reviewer should consider, and the actions tied to each score.

- Define the permitted outcomes—for example: High / recommend keep, Medium / review, and Low / recommend do not keep.
- List the evidence hierarchy: jurisdiction, remaining term, product alignment, detectability, cost, internal use, competitor read-on, licensing and enforcement signals.
- Specify the output format: score, evidence considered, missing data, explanation, confidence, and recommended action.
- Make the instructions concise, ordered, and internally consistent; avoid undefined terms and conflicting rules.

OUTPUT: **Prompt v1, scoring rubric, required evidence, and a standard explanation format.**

PROMPT BLUEPRINT

"Evaluate each patent asset as High, Medium, or Low using the approved criteria. Consider jurisdiction, remaining term, product alignment, detectability, cost, internal use, competitor read-on, and enforcement or licensing evidence. Return the score, evidence considered, missing information, rationale, confidence, and recommended action."

Use only enterprise-approved tools for confidential portfolio information.

2

STEP 2

Iterate with real data and compare human reasoning with AI reasoning

Select a representative pilot set, not just the easiest assets. Have the AI and one or more human reviewers score the same assets independently before revealing one another's conclusions.

- Include different jurisdictions, ages, technologies, product relationships, cost profiles, and likely enforcement value.
- Compare both the final label and the explanation—not merely the percentage of identical scores.
- Pay special attention to severe disagreements, such as AI = High while the human = Low, or the reverse.
- Normalize the human standard: reviewers may use "High," "Medium," and "Low" differently or rely on undocumented institutional knowledge.

OUTPUT: **A disagreement set showing where the AI, the human, or the underlying data may be wrong.**

3

STEP 3

Use AI to loop between the two explanations and improve the prompt

Feed the human rationale, AI rationale, and available evidence into a separate comparison step. Ask the AI to identify the cause of each disagreement and propose a narrowly targeted prompt change.

- Classify each disagreement: missing data, ambiguous criterion, inconsistent human standard, prompt conflict, or model reasoning error.
- Change one material rule at a time, preserve the prior prompt version, and retest the selected disagreement set plus a holdout sample.
- Do not blindly optimize for historical labels. A disagreement may reveal that the human standard itself needs clarification.
- Stop when the process is sufficiently reliable for the intended action—not when every historical label is reproduced.

OUTPUT: **Prompt v2+, a documented change log, and an approved validation threshold.**

FROM VALIDATED PILOT TO CONTINUOUS OPERATION

Scale the process—and then keep the evidence current

Once the scoring logic and explanation quality are accepted, the same mechanism can be applied across the portfolio and re-run as legal, business, financial, and market evidence changes.

Human evaluation	Compare outputs	Diagnose disagreement	Revise prompt	Retest on holdout data
→	→	→	→	↺

4

STEP 4

Once satisfied, scale to the whole portfolio

Freeze the approved prompt and rubric as a controlled version, run the portfolio in batches, and route uncertain or high-impact results to a human exception queue.

- Apply the validated prompt to all in-scope assets, while preserving the evidence snapshot and prompt version used for each score.
- Prioritize review of Medium results, low-confidence results, missing-data cases, and any recommendation that could trigger abandonment or material spend.
- Allow a human to override the score and explanation, but require the reviewer to record the reason and supporting evidence.
- Use rank ordering, cost exposure, and review capacity to sequence decisions. With a ranking system, it's possible to enforce arbitrary percentages and can be useful in certain cases (i.e. have to cut 20% of your portfolio). In general, the AI can struggle putting arbitrary percentages into each category, unless they are clearly defined and must be human reviewed. Humans will be better at splitting after ranking.

OUTPUT: **Portfolio-wide recommendations, an exception queue, and a defensible decision record.**

5

STEP 5

Use tools to automate recurring evaluation as datasets change

Portfolio value is dynamic. Automate data ingestion, monitoring, rescoring, alerts, and audit history so the recommendation evolves when the underlying facts evolve.

- Re-evaluate on a schedule and on material triggers—for example: new product priorities, budget changes, legal-status events, competitor filings, usage evidence, or new infringement signals.
- Feed current datasets into the same approved mechanism rather than rebuilding the analysis manually each cycle.
- Show what changed, why the score changed, and which evidence caused the change; preserve prior versions for comparison.
- Maintain human approval for consequential decisions and retain overrides as future training and prompt-improvement data.

OUTPUT: **A continuously monitored portfolio with explainable recommendations, alerts, overrides, and version history.**

DATA FEEDS THAT KEEP THE MECHANISM CURRENT

Legal & term	Business	Market	Financial	Human input
Status, jurisdiction, remaining term, claim scope	Product priorities, internal use, strategic tags	Competitors, read-on evidence, new filings, licensing signals	Renewal fees, prosecution cost, budget thresholds	Overrides, explanations, approvals, exception outcomes

WHAT "READY TO SCALE" LOOKS LIKE

✓ Human reviewers agree on the rubric and category definitions.

✓ Severe disagreement rates are understood and acceptable for the use case.

✓ Consequential recommendations remain subject to human approval.

✓ The AI provides a usable explanation and identifies missing evidence.

✓ Prompt, data, model, and reviewer versions can be reconstructed later.

✓ The process has an owner, review cadence, escalation path, and audit trail.

IP COPILOT IPWatchdog

AI-ASSISTED PATENT PRUNING

Deciding What to Keep with Precision

A repeatable, defensible framework for using AI to structure portfolio review - while preserving experienced human judgment.

September 8, 2026 | 12:00-1:00 PM ET

Mike Scannell, Ph.D.
Director, Legal Counsel – IP
Resmed

Austin Walters
Founder & CEO
IP Copilot

Gene Quinn
Moderator · Founder & CEO
IPWatchdog

STRUCTURE

1 Evidence
Use the same legal, cost, business and market inputs across the portfolio.

2 Calibration
Compare AI and attorney reasoning on a representative review set.

3 Judgment
Treat AI as decision support; document human approval and overrides.

CLE educational program | Not legal advice

PROGRAM OUTCOMES

Learning objectives

After completing this program, attendees should be able to:

01

Identify the evidence

Select legal-status, cost, business, claim-scope and market information needed to evaluate patent families.

02

Build the workflow

Create a repeatable process from data normalization and AI analysis through human approval and audit logging.

03

Design the rubric

Write a constrained prompt and output schema that produces explainable recommendations and flags uncertainty.

04

Validate recommendations

Calibrate AI output against normalized attorney judgment and investigate disagreements before scaling.

05

Apply guardrails

Address competence, confidentiality, verification, supervision, versioning and change-management requirements.

THE PROBLEM

Patent pruning is an investment and governance problem

Portfolios grow, costs recur and the filing context fades.

Portfolio growth

New filings and acquisitions add families faster than most teams can reassess them. Redundant or weak coverage becomes harder to see.

Recurring cost clock

U.S. utility and reissue utility patents have maintenance-fee payment windows at 3-3.5, 7-7.5 and 11-11.5 years after issue. In some jurisdictions (e.g. EP), costs are incurred annually, depending on jurisdiction, even before grant.

Context drift

The people reviewing an asset may not know why it was filed, which product it protects, or how the competitive landscape has changed.

A defensible process replaces deadline triage with evidence, consistent criteria and documented judgment.

Maintain | Monetize | Abandon | File more

DECISION FRAMEWORK

Six inputs to a defensible recommendation

Score evidence consistently, but preserve uncertainty and reviewer discretion.

1 Legal status & term

Status, jurisdiction, remaining term, family relationships, prosecution posture and next deadline.

2 Projected cost

Maintenance, prosecution, translation, validation, annuity and outside-counsel costs by jurisdiction.

3 Business alignment

Current products, roadmap, strategic priorities, revenue relevance and technology reuse.

4 Claim detectability

Whether meaningful claim limitations can be observed, tested or proven in a product or process.

5 Infringement / licensing signals

Evidence that may support enforcement, licensing, defensive value or a monetization review.

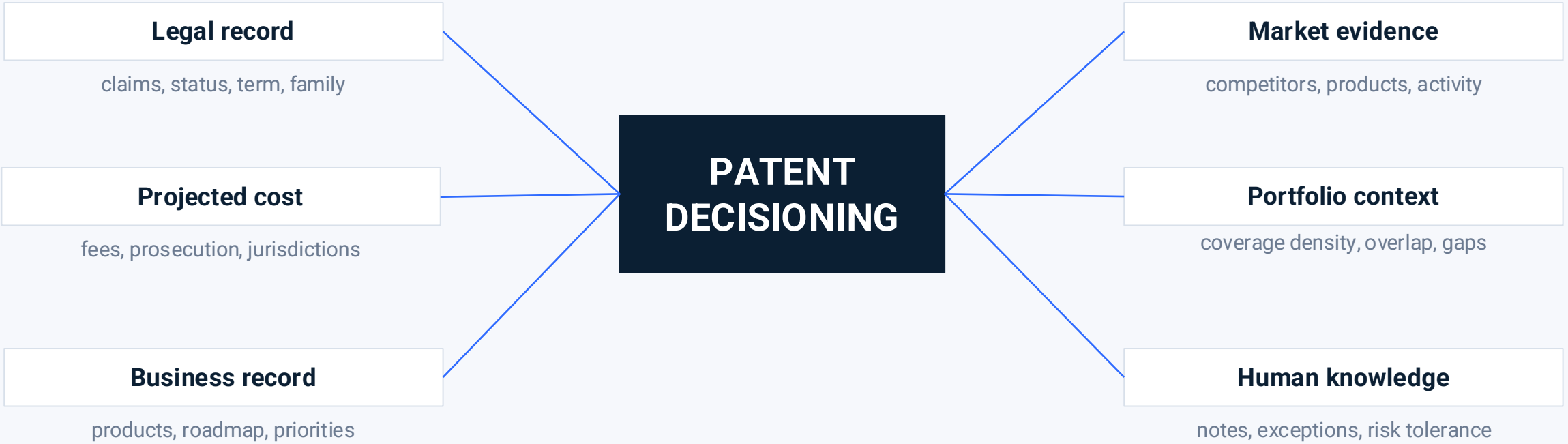
6 Map to your internal products

Patent-to-product relationships based on product documentation and technical similarity, revealing coverage, overlap and potential gaps.

DATA FOUNDATION

The model can only use the evidence it receives

Make the permitted sources explicit and expose missing or stale data.



Data-quality rule

A missing field should lower confidence or trigger review - not be silently inferred.

OPERATING MODEL

A seven-step AI-assisted review workflow

Start in shadow mode; move to operational use only after calibration and governance review.

1 Define decision

Family-level outcome, reviewer and timing.

2 Normalize data

Resolve status, identifiers, costs and tags.

3 Set rubric

Allowed evidence, factor definitions and output schema.

4 Build gold set

Representative cases with reasoned human review.

5 Run + version

Store data snapshot, prompt, model and output.

6 Adjudicate gaps

Review severe disagreements and missing evidence.

7 Approve + monitor

Human decision, override reason and re-review trigger.

EXAMPLE

Designing a prompt

Title	Idea connections	Potential infringements	Business alignment	Detectability	Recommendation
ARTIFICIAL INTELLIGENCE-BASED IDEA DISCOVERY AND FEEDBACK ...	3	0	Highly aligned	Hard to detect	Low
SYSTEM FOR AUTOMATED KNOWLEDGE ASSET DISCOVERY, GRAPH-...	5	1	Highly aligned	Hard to detect	High
System and Method for Protsdsdsection During Inverter Shutdown in ...	0	0	Minimally aligned	Hard to detect	Low

```
EVALUATE:

- STRATEGIC RELEVANCE - to products, platforms, workflows, or growth areas
- POTENTIAL INFRINGEMENT - Competitive differentiation or leverage
- SCALABLE & DURABLE - useful within a broader portfolio
- ENFORCEABLE - Any weak characteristics such as poor detectability, limited
enforceability, narrow standalone use, weak commercial relevance, or better
treatment as a trade secret

WEIGHTED PRIORITY:
  1) STRATEGIC RELEVANCE - portfolio asset patentability signals
  2) ENFORCEABLE - disclosures patentability signals (if present)
  3) SCALABLE & DURABLE - Ideas patentability signals (if present)
  4) POTENTIAL INFRINGEMENT - Are others using it (if present)
```

VALIDATION

Do not treat attorney labels as unquestioned ground truth

Normalize the human rubric first, then learn from the disagreements.

Build a representative gold set

Include a cross-section of technologies, jurisdictions, ages, costs, statuses and expected decisions. Avoid a sample dominated by one vintage or one outcome.

Rank first

Calibrate the reviewers

Define what low, medium, high or other categories mean. Compare multiple reviewers and adjudicate inconsistent thresholds before judging the model.

Set threshold second

Analyze disagreement reasons

Review the evidence and explanations behind each material difference. Separate prompt gaps, missing data, model error and human inconsistency.

Rank and threshold results

IMPLEMENTATION CASE STUDY

Lessons from an in-house medtech pilot

The following lessons are generalized; confidential asset-level data and quantitative results are not included.

1	**Start with reviewer alignment**	Multiple attorneys may apply the same category names differently. Calibrate definitions and examples first.
2	**Inspect explanations**	A disagreement may reveal a prompt gap, missing business knowledge or a stronger AI-supported conclusion.
3	**Rescore selected cases**	Test prompt changes on a stable representative set before rerunning the full portfolio.
4	**Expose the actual taxonomy**	Ensure hidden defaults, score buckets and system instructions match the rubric shown to reviewers.
5	**Preserve human judgment**	AI can reduce inconsistency and emotional attachment, but counsel owns the decision and the record.

The objective is not to prove that AI agrees with every reviewer. It is to produce a better-informed, more consistent and more reviewable decision process.

GOVERNANCE AND ETHICS

Guardrails for AI-assisted legal decision support

Existing professional obligations continue to apply when AI is used.

Competence

Understand the tool's capabilities and limitations; train users and review the output.

Confidentiality

Assess data handling, retention, access, vendor terms, hosting and approved-use boundaries.

Verification

Do not rely on AI output without a reasonable human review of facts, law and evidentiary support.

Auditability

Version prompts, models and data; preserve explanations, overrides, approvals and rerun triggers.

GOVERN policy + ownership

MAP context + impact

MEASURE testing + monitoring

MANAGE response + improvement

DECISION AND CLOSE

AI structures the review; humans own the decision

A complete record should show the evidence, recommendation, uncertainty and approving judgment.

MAINTAIN

Continue investment

Strong alignment, differentiated coverage or strategic defensive value.

MONETIZE

License or enforce

Credible market evidence, detectable claims or a focused transaction opportunity.

ABANDON

Stop or narrow spend

Weak alignment, duplicative coverage, low enforceability or cost outweighing value.

FILE MORE

Close a coverage gap

Important product or competitor activity is not adequately covered by current claims.

Three closing rules

1. Rank recommendations before applying a budget or portfolio cutoff.
2. Escalate missing data and material disagreement; do not hide them in a score.
3. Re-run the review when strategy, evidence, costs, claims, prompts or models materially change.

QUESTIONS

SCALING THE WORKFLOW

Tools can scale evaluation and monitoring

Platforms such as IP Copilot can apply the rubric continuously, surface new evidence and preserve human control.

PORTFOLIO-WIDE SCORING + EXPLANATION

Title	Business alignment	Detectability	Recommendation
ARTIFICIAL INTELLIGENCE-BASED IDEA DISCOVERY AND FEEDBACK ...	Highly aligned	Hard to detect	Low
SYSTEM FOR AUTOMATED KNOWLEDGE ASSET DISCOVERY, GRAPH-...	Highly aligned	Hard to detect	Medium
System and Method for Protsdsdsection During Inverter Shutdown in ...	Minimally aligned	H	
Current sensing on a MOSFET	Minimally aligned	F	
Zero voltage shibwitching	Minimally aligned	H	
uiiniPower converter for a solar panel	Minimally aligned	F	
Method for distributed power harvesting using DC power sources	Minimally aligned	H	
Apparatus and method for determining an order of power devices in p...	Minimally aligned	H	

Medium

The asset is moderately recommended due to strong business alignment and high strategic value in automating IP workflows. However, the recommendation is tempered by significant risks regarding patent eligibility and enforceability. The core invention relies on backend algorithms and internal data processing, which are classified as hard to detect and potentially ineligible abstract ideas, limiting its utility as a defensive patent asset.

AI generated

MONITOR | RESCORE | EXPLAIN | OVERRIDE

Scale the evidence and consistency - not the authority to make the final legal and business decision.

WHAT THE TOOLING LAYER SHOULD DO

1 **Monitor competitors**
Refresh filings, product documents and likely claim read-on signals.

2 **Monitor internal usage**
Track current products, roadmap relevance and technology reuse.

3 **Recommend over time**
Re-score when term, cost, claims, evidence or objectives change.

HUMAN OVERRIDE + AUDIT TRAIL

Counsel can change the score and explanation, with the reviewer and timing preserved.