



## ABSTRACT

Corporate R&D teams need better ways to turn large patent portfolios into usable knowledge for **idea discovery** and **evaluation**. In partnership with **Honda Research Institute**, we developed a **structured framework** that makes patent data **comparable** and **reusable** for innovation. Rather than treating innovation as only a **generation problem**, the framework focuses on how ideas can be judged more consistently through **semantic similarity**, **novelty**, and **business relevance**.

The framework reads and filters a **patent portfolio**, maps its structure using **semantic embeddings** and **clustering**, extracts **idea ingredients**, generates new concepts through **structured recombination**, and evaluates them through **novelty** and **business value** signals.

**5,967**

HRI patents analyzed

**8**

Technology clusters discovered

**119**

Frontier patents (24% white space)

**50**

Novel concepts scored and ranked

## INTRODUCTION

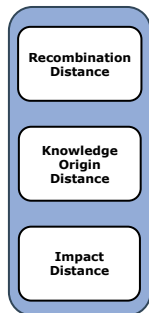
Generative AI has shifted innovation to **abundance**. LLMs can generate endless variations, but they cannot reliably tell whether an idea is genuinely **new** or merely familiar content in different words. This made the project less a generation problem than a **judgment** problem. Before evaluating ideas, we first had to define what **"new"** meant. We defined **novelty** at the level of **meaning**, not wording, and treated it as relational distance from what already exists. In this view, newness comes from **editing** and **recombination**.

To make that judgment more systematic, we focused on three questions:

- **Defining "New"**: What separates a conceptually new idea from a repackaged one?
- **Measuring Distance**: How can semantic distance be mapped against an existing patent?
- **Prioritizing Value**: How can novelty be separated from business feasibility and relevance?

## MEASURING DISTANCE

Three Dimensions Of Novelty

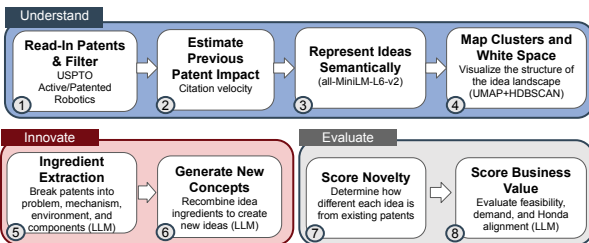


PEMPSA Schema

Field	Description
<b>Problem</b>	The core challenge the concept aims to resolve
<b>Environment</b>	The context in which the idea operates
<b>Mechanism</b>	How it addresses the problem
<b>Physical Ingredients</b>	The physical components involved
<b>Source Domain</b>	The technical domain of each component
<b>Applicability</b>	Other domains that it can be applied to

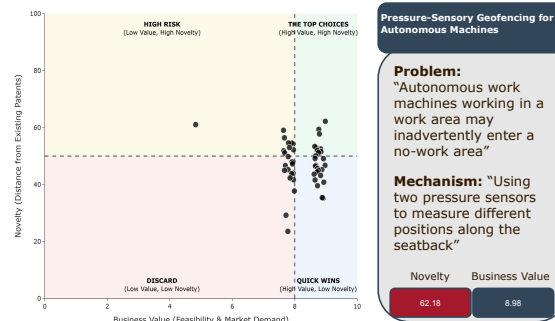
## METHODS AND MATERIALS

To operationalize this framework, we built a pipeline that reads and clusters HRI's patent corpus, extracts structural idea ingredients, generates cross-domain combinations, and scores each concept on both originality and business value.



## RESULTS

Concept Prioritization Matrix



## CONCLUSIONS

This research provides a structured method for measuring novelty at the **concept level** rather than surface wording. By comparing ideas through **meaning**, not phrasing, the framework shows that novelty can be quantified in a more rigorous and useful way. **White space mapping** and **value ranking** can help teams **prioritize** where to explore next, but **high novelty alone does not guarantee business value**.

- **Key finding:** The framework shifts the focus from idea generation alone to the representation, analysis, and evaluation of ideas.
- **Current limitation:** Business value scoring remains heuristic and still requires expert validation for real-world use.
- **Next step:** Future work will expand the dataset, improve ingredient extraction and performance measures, and test citation velocity as an additional scoring signal.

CONTACT



David Fabyonic



Henry Kim



Carlos Ramirez



Daniel Yu

REFERENCES

- Giordano, V., Puccetti, G., Chiarello, F., Pavanello, T., & Fantoni, G. (2022). Unveiling the inventive process from patents by extracting problems, solutions and advantages with natural language processing. SSRN Electronic Journal. <https://doi.org/10.2139/ssrn.4223458>
- Girotra, K., Meincke, L., Terwiesch, C., & Ulrich, K. (2023). Ideas are dime a dozen: Large language models for idea generation in innovation. SSRN Electronic Journal. <https://doi.org/10.2139/ssrn.4526071>
- Youn, H., Strumsky, D., Bettencourt, L. M. A., & Lobo, J. (2015). Invention as a combinatorial process: Evidence from US patents. Journal of The Royal Society Interface, 12(106), Article 20150272. <https://doi.org/10.1098/rsif.2015.0272>