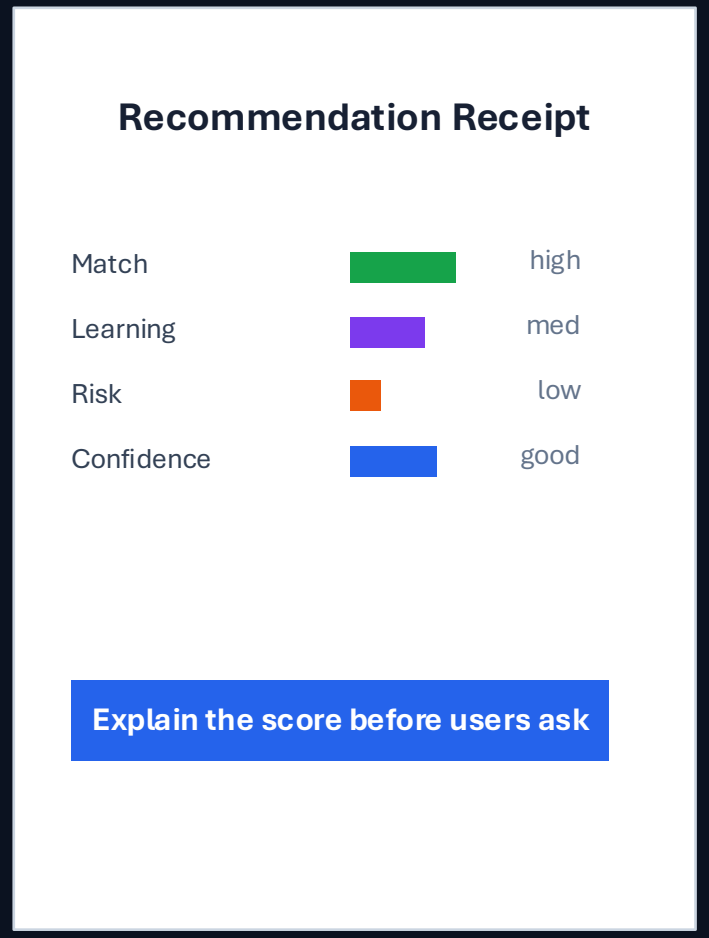
Active Inference as an Intrinsically Explainable Recommender Architecture

Epistemic–Pragmatic Score Decomposition and Experimental Repository

- Explanation is generated inside the ranking process, not added afterward
- Every score splits into match, learning value, risk, and confidence
- Results are reproducible through an open experimental repository



Bhagyeshkumar Chokhawala • Dr. Atif Farid Mohammad
Capitol Technology University

The everyday problem

People often ask: “Why did the system recommend this to me?”

?

Was it recommended because it matches the user, because it is popular, because the system wants to learn, or because it is simply uncertain?

Accuracy is not enough
A system can pick relevant items but still hide why it picked them.

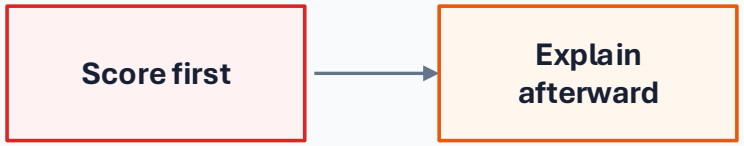
Trust needs reasons
Users and reviewers need explanations that are tied to the real decision.

Trade-offs matter
Good recommendations balance relevance, novelty, diversity, risk, and uncertainty.

The research question in plain language

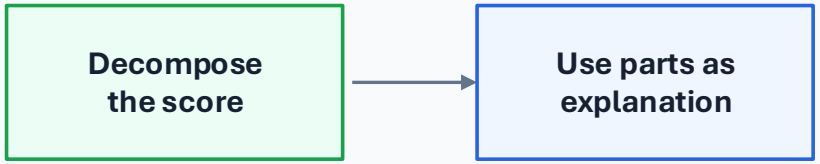
Can a recommender explain itself using the same logic it uses to rank items?

Typical approach



Post-hoc explanation

Proposed approach



Intrinsic explanation

Goal: explanation by design, not explanation after the fact

The central idea: a recommendation receipt

Instead of one mysterious score, the system keeps four visible reasons.

1. Match
How well does this item fit what the user seems to like?

2. Learning
Would showing this item help the system learn more about the user?

3. Exposure risk
Is the system over-promoting the same popular items?

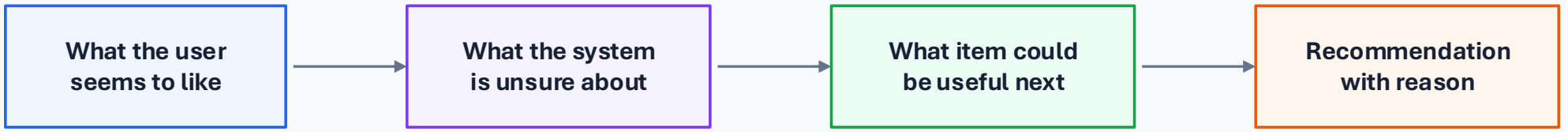
4. Confidence
How reliable is the prediction for this item?

Final recommendation = visible combination of these four reasons

The paper calls this decomposition Expected Free Energy, or EFE.

Active Inference in plain English

Active Inference is a way for a system to choose actions while managing uncertainty.



Simple metaphor

A good guide does not only repeat what you already know. Sometimes the guide suggests something new because your reaction will reveal what you truly prefer.

Why it helps recommenders

The model can balance familiar choices with useful exploration instead of using fixed rules for everyone.

What the EFE score looks like

The technical equation is simple to explain: good recommendations lower bad costs and raise useful learning.

$$G(i \mid b) = \lambda p \cdot \text{Match Cost} - \lambda e \cdot \text{Learning Value} + \lambda r \cdot \text{Risk} + \lambda a \cdot \text{Uncertainty}$$

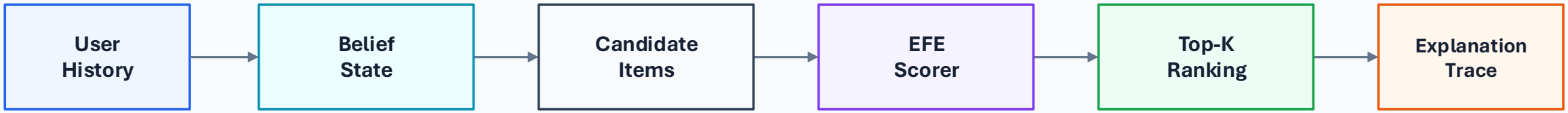
Lower is better
The system ranks items by the lowest final EFE score.

Learning is rewarded
The epistemic term is subtracted because useful information is valuable.

Risk is visible
Popularity or exposure concerns appear as explicit score components.

For a non-technical audience, this is simply a transparent scoring recipe.

End-to-end architecture



Feedback updates the belief state over time

Modular deployment
The EFE scorer can sit on top of existing recommender infrastructure rather than replacing every model.

Audit-ready output
Every displayed item can store the exact reasons used to rank it.

How recommendations change with uncertainty

The system should behave differently for a new user than for a familiar user.

New or uncertain user
The model explores more because it still needs to learn what the user likes.

Exploration when learning matters

Known or stable user
The model exploits more because preferences are clearer.

Preference matching when confidence is high

Experimental repository

The research code turns the paper claim into reproducible experiments.

Datasets
MovieLens 1M and Amazon Beauty support standard public evaluation.

Baselines
BPR, LightGCN, and SASRec represent common recommender families.

Outputs
The repository generates paper-ready CSV, LaTeX tables, figures, and traces.

How success is measured

The evaluation asks two questions: Does it stay accurate, and does it improve transparency and discovery?

Accuracy
Recall@10 and NDCG@10 check whether relevant items appear near the top.

Discovery quality
Diversity and novelty check whether results avoid repetitive, obvious recommendations.

Catalog health
Coverage and average popularity check whether exposure is overly concentrated.

Ablation study: remove one component at a time

Full model

No learning

No risk

No ambiguity

This shows which part of the score creates which behavior.

Main findings from the paper

The reported pattern is: competitive accuracy with stronger beyond-accuracy behavior.

Why this matters

The model does not need to hide explanation behind a separate tool. The explanation is generated from the same score that produces the ranking.

Research integrity note

The repository is designed to regenerate and replace result tables from benchmark runs rather than bundling fixed result numbers.

Why this matters outside the lab

Transparent recommendations are useful wherever ranking decisions affect people.

Consumer platforms
Explain why a movie, product, article, or lesson is recommended.

Enterprise knowledge
Help employees trust recommendations for documents, actions, or next steps.

Responsible AI
Support audit, governance, and human control over ranking behavior.

From black-box ranking to auditable decision-making

The same concept can support user-facing explanations, administrator controls, fairness reviews, and operational monitoring.

Takeaway and discussion

A recommender can be accurate, adaptive, and explainable by construction.

Core idea
Keep the reasons inside the score instead of adding explanations afterward.

Technical contribution
Use Active Inference / EFE to decompose ranking into match, learning, risk, and uncertainty.

Practical contribution
Generate reproducible experiments, paper tables, figures, and explanation traces from code.

Discussion prompt: What should a recommender be required to explain before we trust it?

Thank You

Research Profile QR Code

Scan the QR code to open my research profile, publications, Google Scholar, ORCID, and professional links.