European Commercial Real Estate Data Alliance (E-CREDA)

2026 E-CREDA ANNUAL CONFERENCE

Has AI changed our view on real estate markets, theories and research approaches?

Thies Lindenthal | University of Cambridge

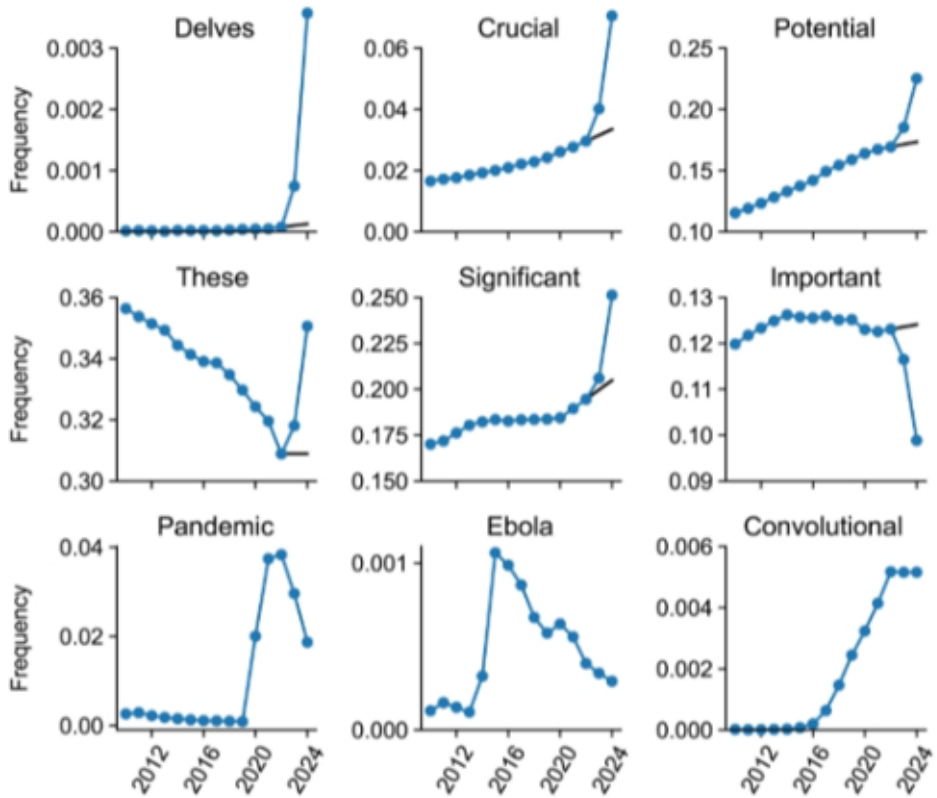
HAS AI CHANGED OUR VIEW ON REAL ESTATE MARKETS, THEORIES AND RESEARCH APPROACHES?

Thies Lindenthal · University of Cambridge · 2026

CHATGPT SHOCK IN ACADEMIA

How much do researchers use LLMs for writing? A lot!!

- Kobak et al. (2025, *Science Advances*): At least 13.5% of 2024 biomedical abstracts processed with LLMs — reaching 40% in some subcorpora



… AND REAL ESTATE RESEARCH?

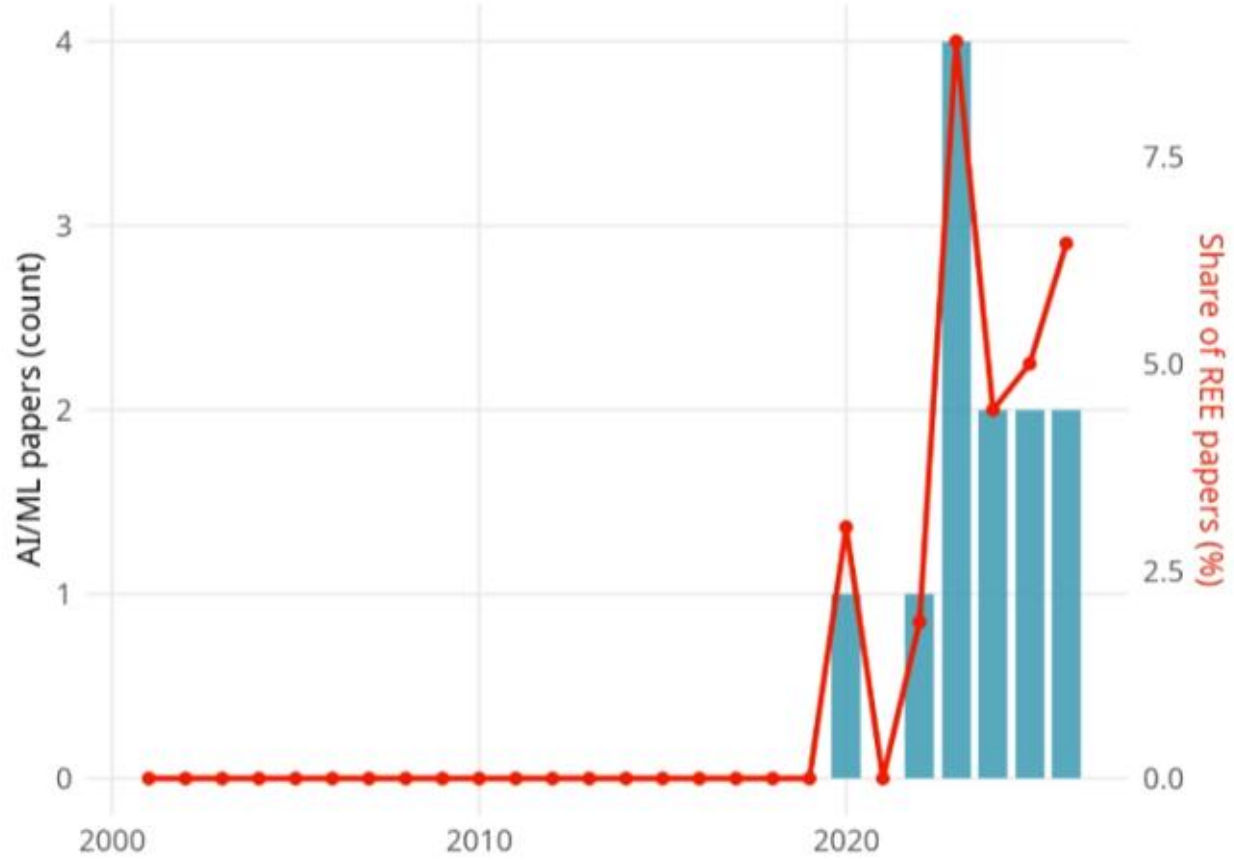
Same picture in real estate academic literature: LLMs are everywhere!

- Replication of Kobak et al. (2025, *Science Advances*) using real estate papers
  - ~101k OpenAlex papers tagged real estate / housing / urban economics / REITs, 2010–2025



MORE THAN JUST A CLEVER TYPEWRITER

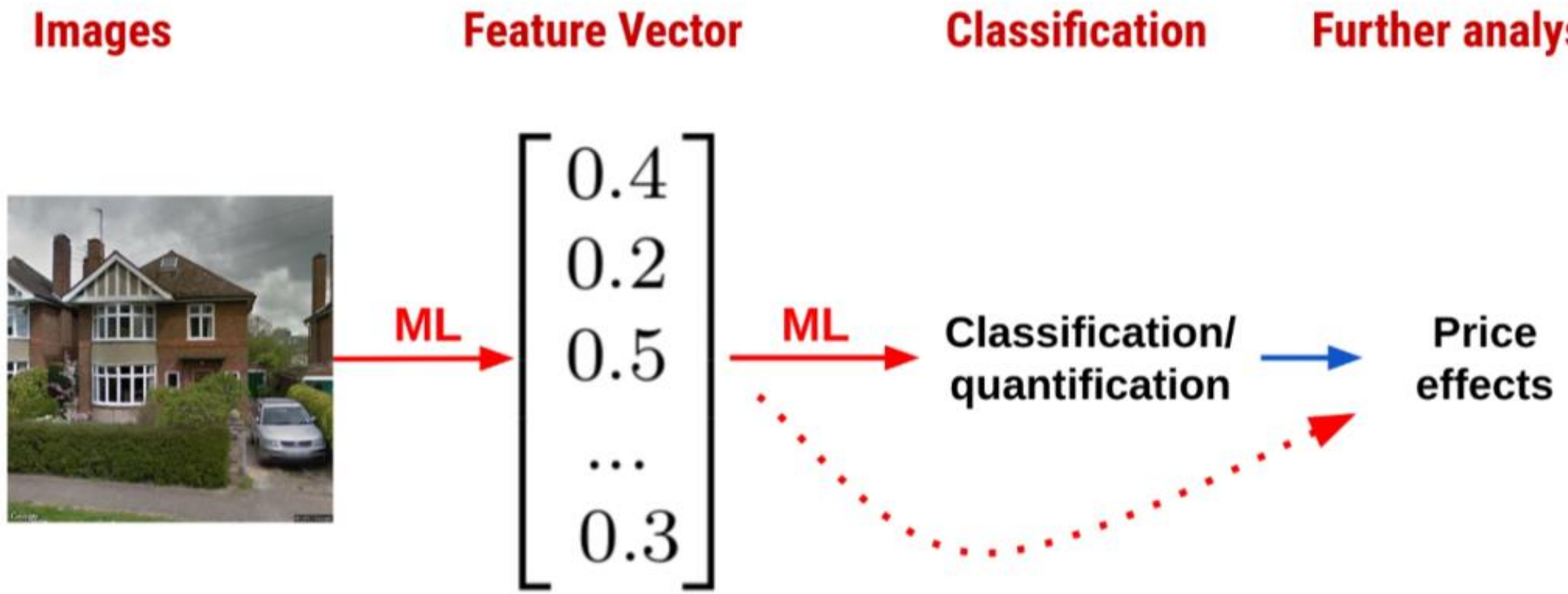
Publications in REE that became doable* due to advances in ML/AI



- *) Based on my subjective classification

NEW DATA SOURCES

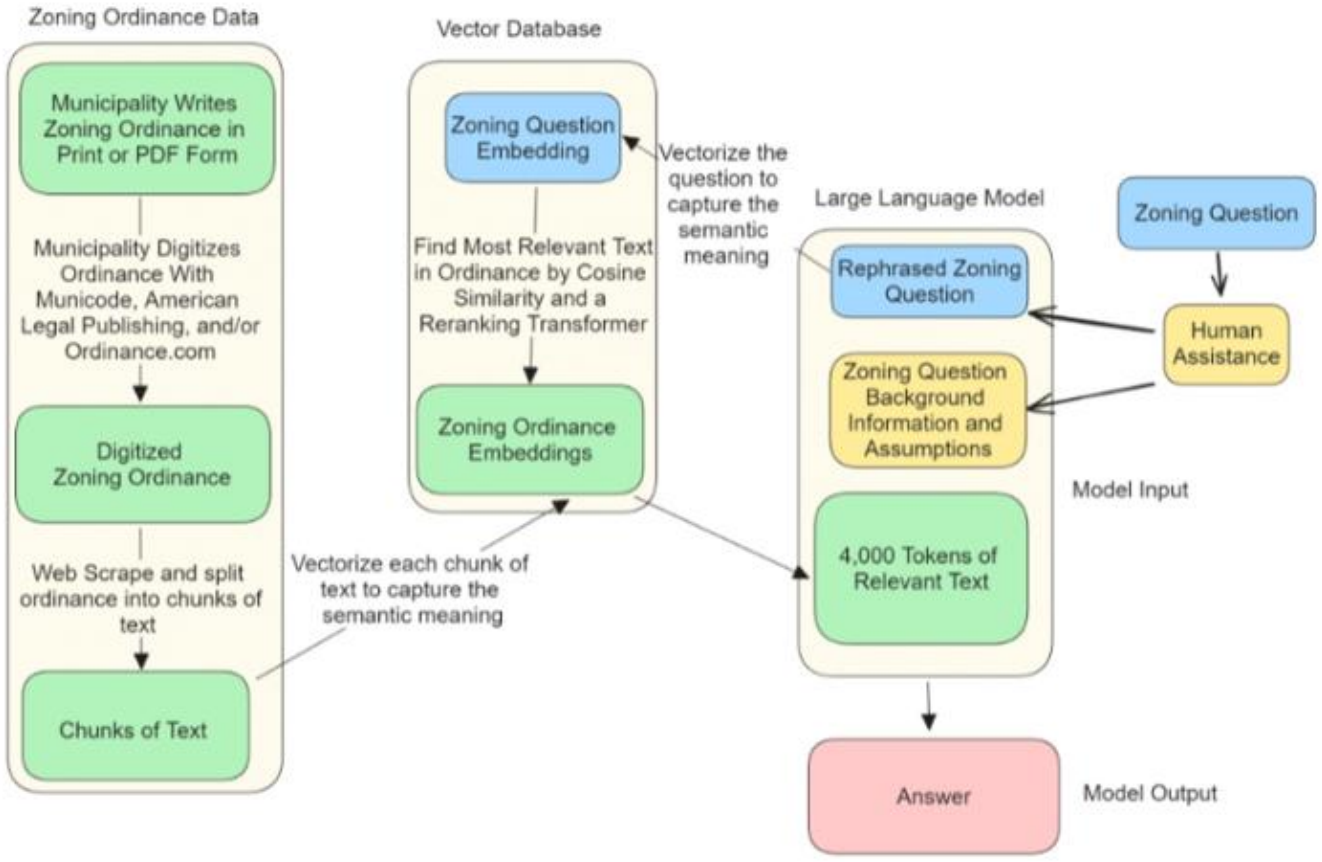
Lindenthal & Johnson (2021): Google Street View as a source of property-level data



READING THE REGULATORY UNIVERSE

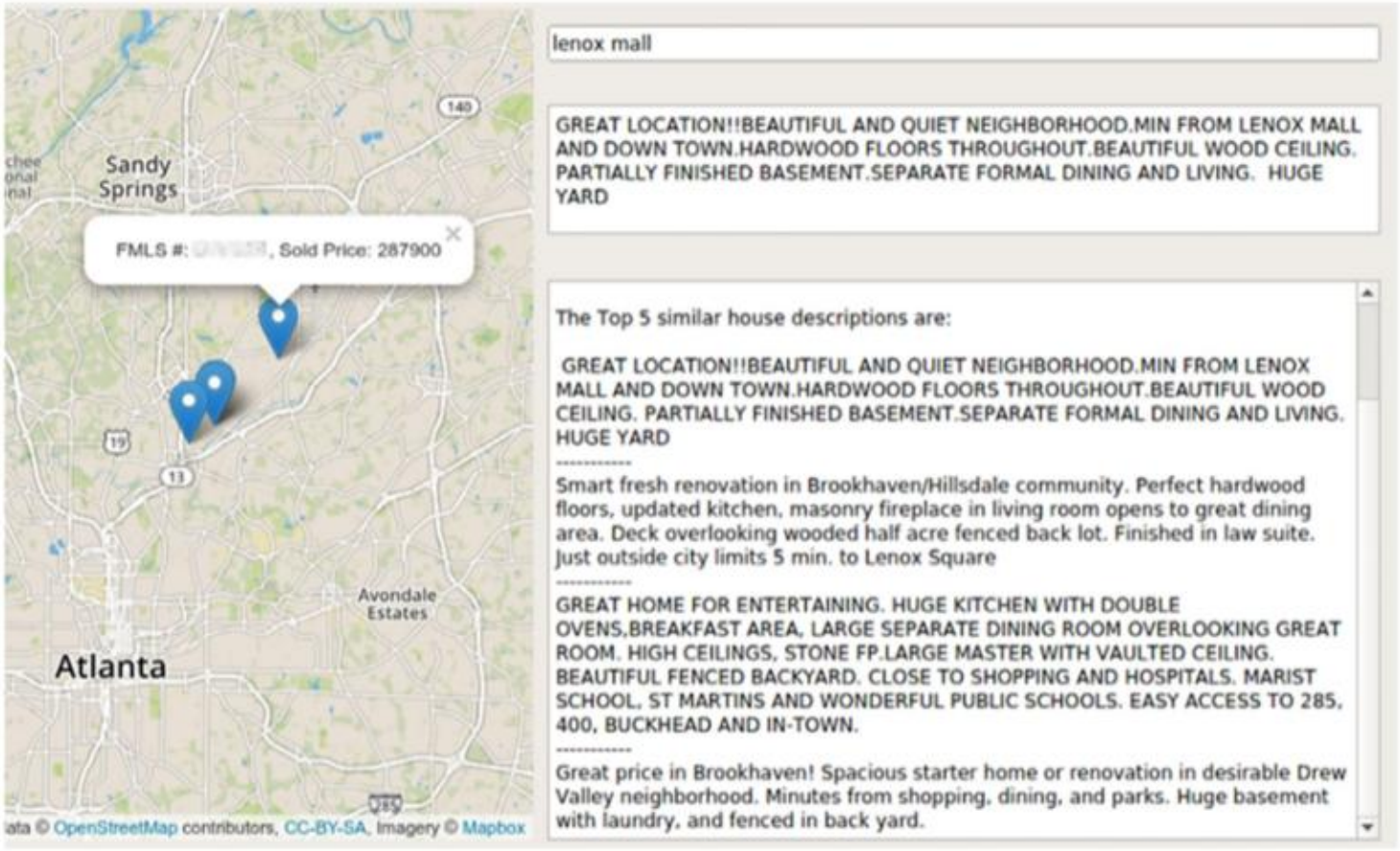
Bartik, Gupta & Milo (2025): NLP on the full corpus of US planning documents

- Better measures of planning restrictiveness



BETTER MEASUREMENT FROM TEXT

Shen & Ross (2021): NLP on MLS listing descriptions to capture property quality beyond observable attributes



SQUEEZING SIGNAL FROM NOISE

Leow & Lindenthal (2025): ML to tease out predictability in REIT returns

- ML models high-dimensional feature sets better! We can outperform traditional approaches.



NEW CHALLENGES: TEST AI SYSTEMS

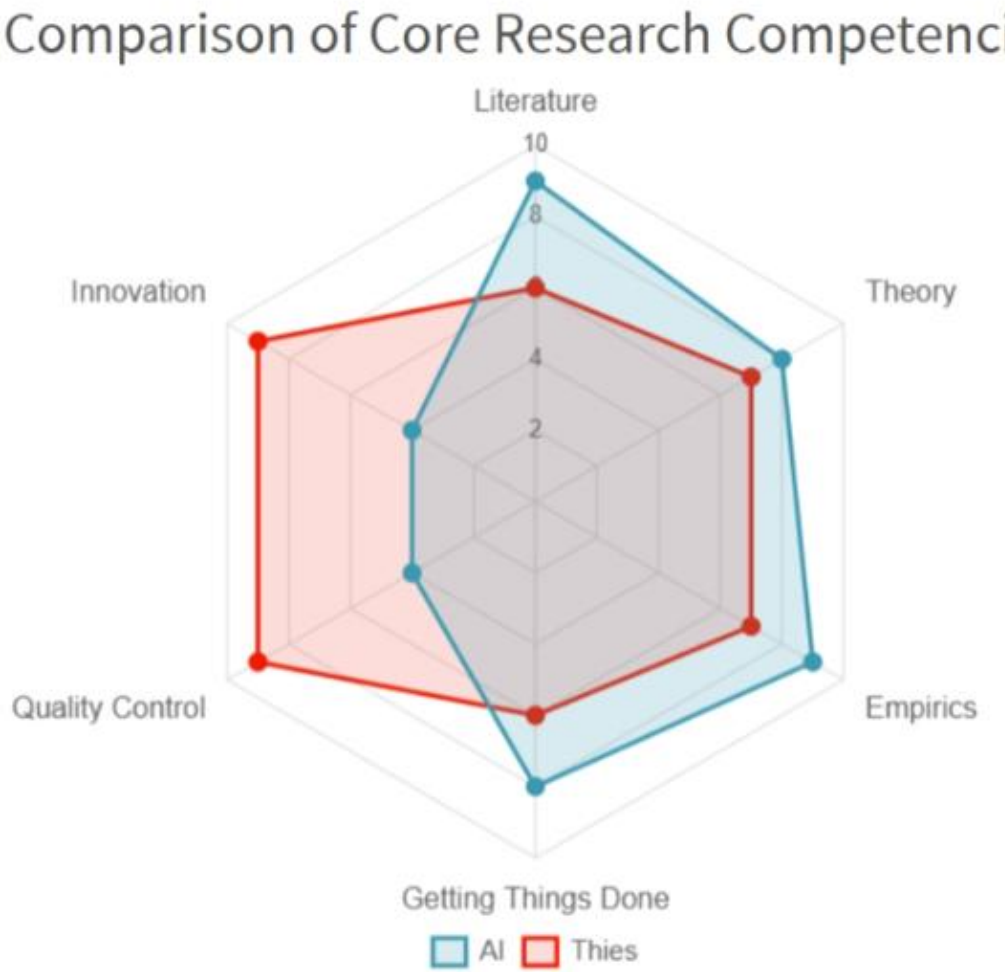
Do the AI models perform the way we need them to?

- Wan & Lindenthal (2023): any deployed AI system must be stress-tested before used in practice (you might be discriminating and/or breaking the law/do other stupid things)



MOVING UP THE VALUE CHAIN...

Am I still useful in the research process? Does a human competitive edge remain?



AI AS CO-INVESTIGATOR

In the hard sciences, AI is already generating and testing hypotheses autonomously

- **AlphaFold2** (Jumper et al., *Nature* 2021): predicted structures for ~200 million proteins — a problem unsolved for 50 years.
- **GNoME** (Merchant et al., *Nature* 2023): discovered 2.2 million stable crystal structures; 736 synthesised and experimentally validated.
- **Coscientist** (Boiko et al., *Nature* 2023): autonomous agent that searched the literature, wrote code, and ran wet-lab experiments without human intervention.
- **Insilico Medicine** (2025, *Nature Medicine*): AI-designed drug candidate completed Phase II clinical trial — from target identification to molecule in under 4 years.

CAN AI GENERATE GOOD RESEARCH IDEAS?

Eight conditions vary what context the LLM sees — from nothing to full corpus to methods borrowed from other fields

- **A — Naïve:** no context, LLM draws on training only
- **B — Full REE corpus:** all 1,676 abstracts provided as context
- **C — Cluster-seeded:** one idea per research cluster; 51 ideas
- **D/F — Method/paradigm transfer from Econ/Finance**
- **E/G — Method/paradigm transfer from Psychology**
- **H — Citation-guided:** LLM steered toward high-impact territory

Each idea embedded, projected into the research space, and scored on **novelty**, **atypicality**, and **feasibility**.

Example (Condition A — Naïve):

Idea: How do changes in remote work accessibility affect the capitalization of neighbourhood human capital density and local labour market externalities into property values, and do these effects persist?

Method: Staggered diff-in-diff exploiting differential remote work adoption across firms; hedonic model with interaction between remote work share and neighbourhood skill density.

Data: Zillow transactions, BLS occupational employment, Burning Glass work location data.

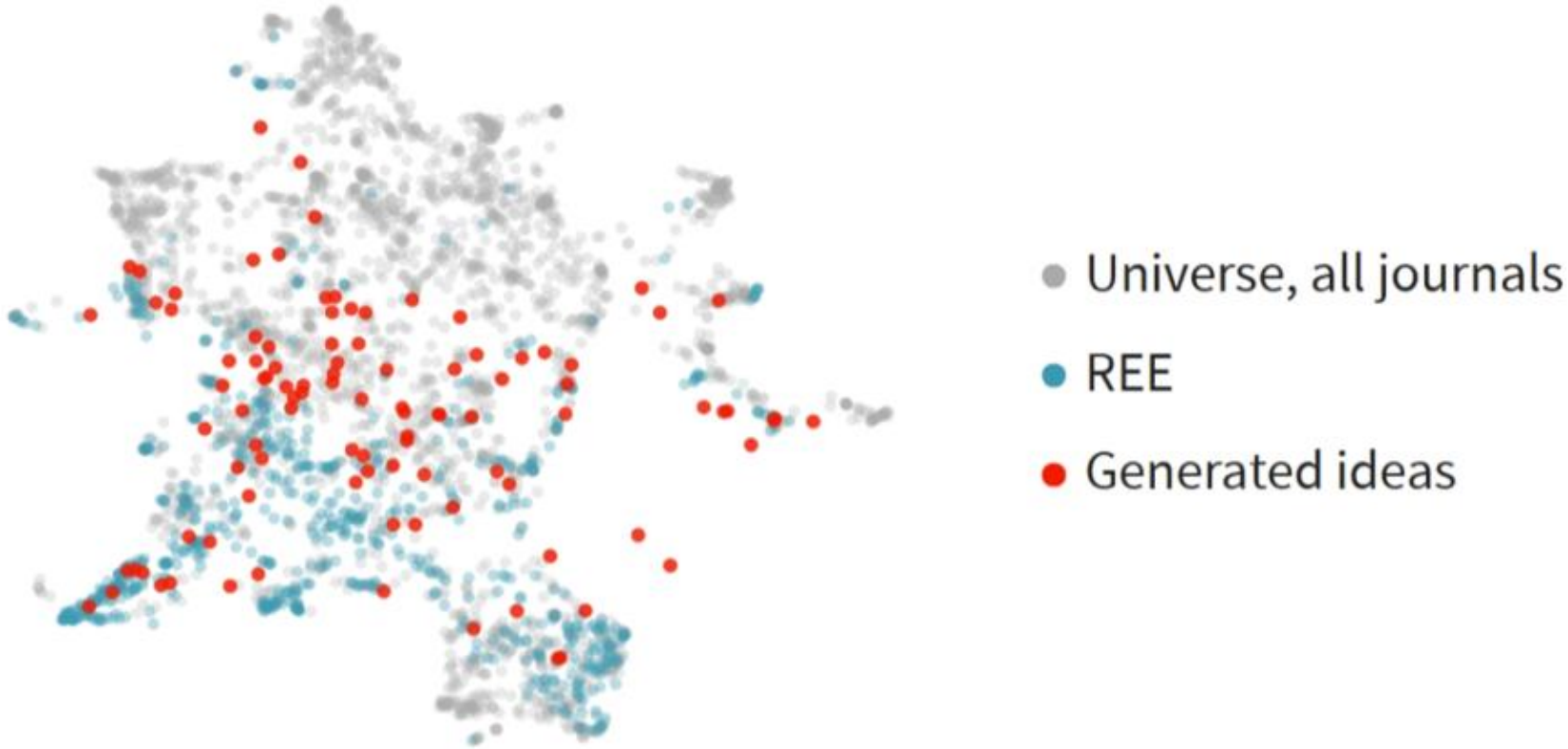
50 YEARS OF REAL ESTATE RESEARCH

1,676 REE articles + top-journal subsets embedded and clustered into a research space



WHERE DO AI-GENERATED IDEAS LAND?

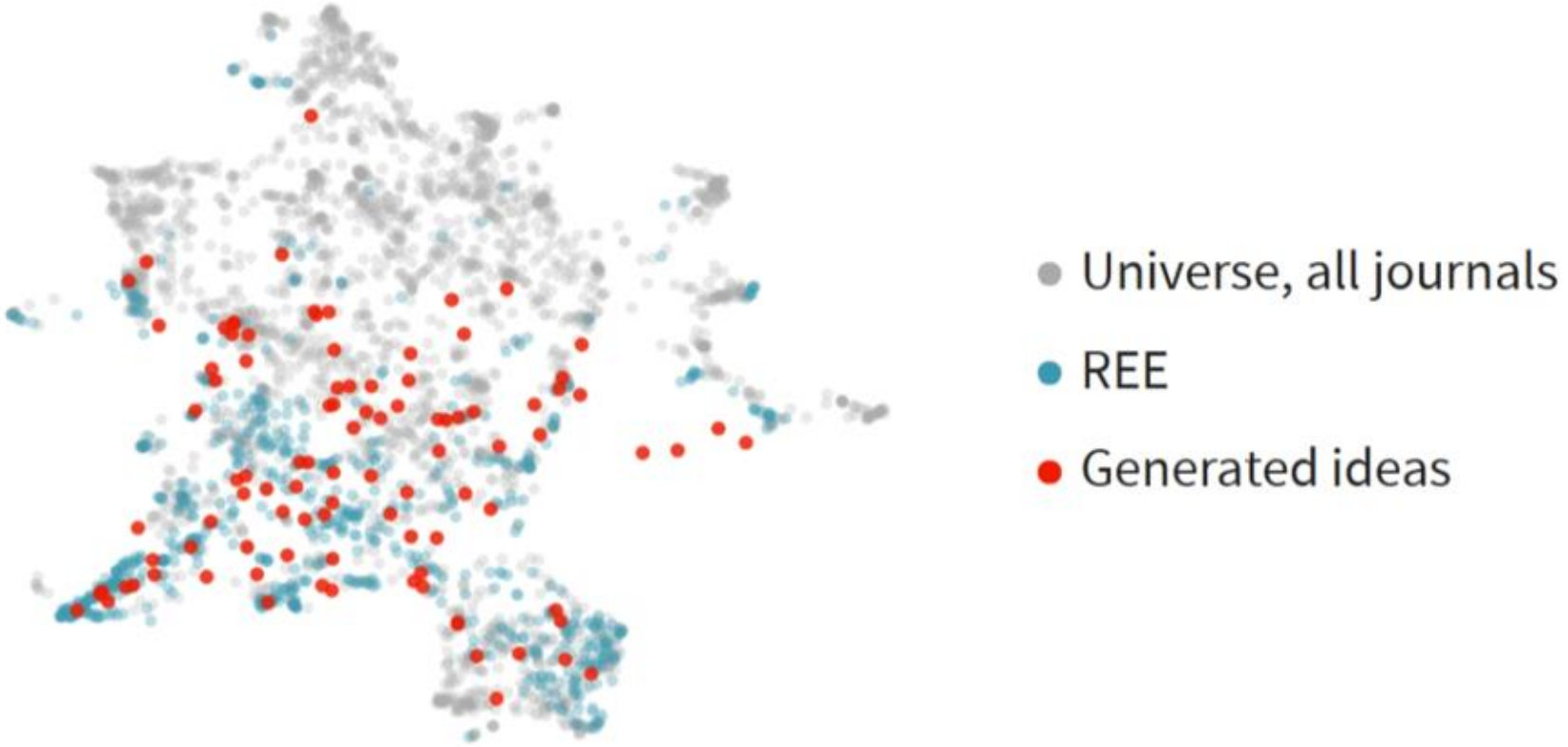
Each dot is a generated idea projected into the research space



A: Naive generation, based on LLM training only

WHERE DO AI-GENERATED IDEAS LAND?

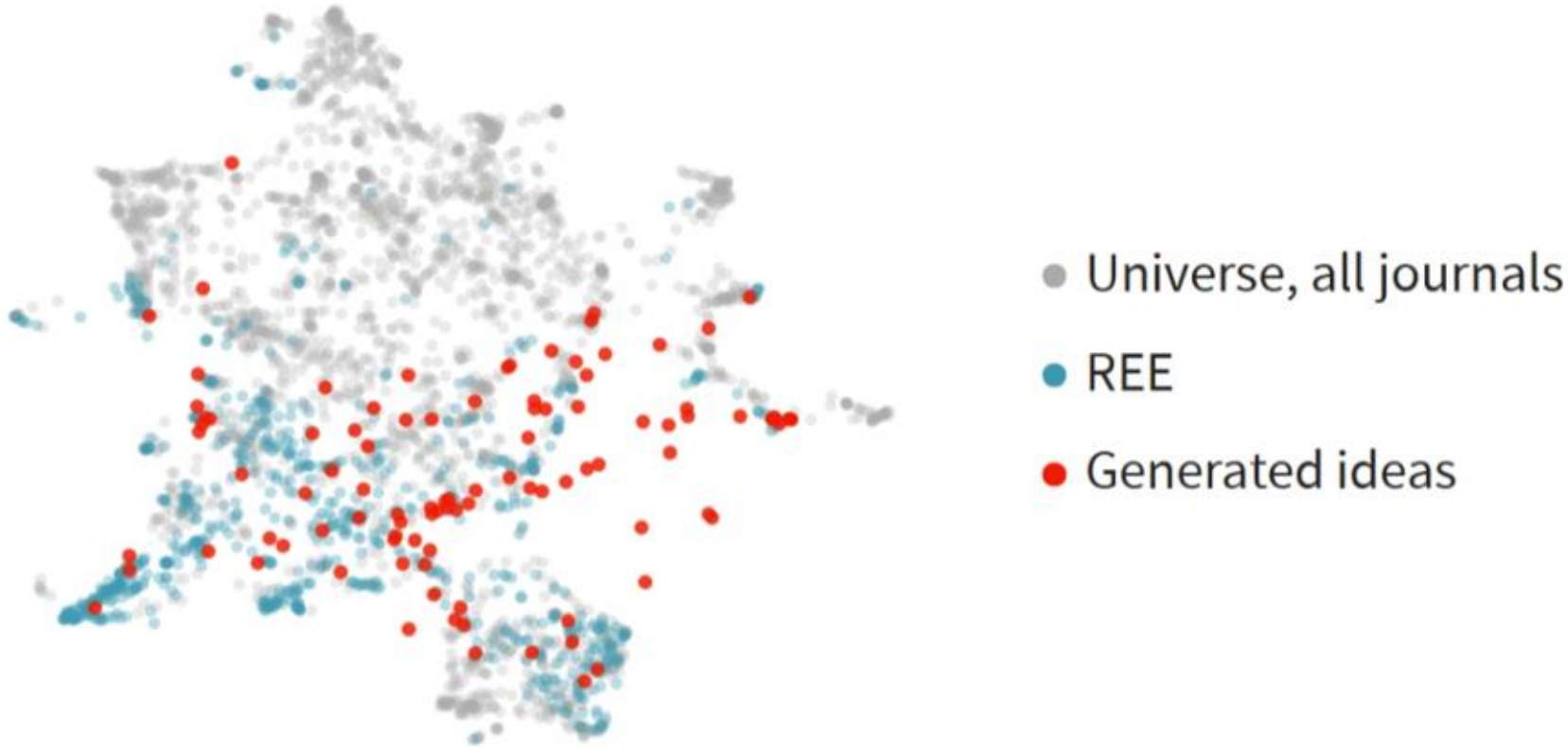
Each dot is a generated idea projected into the research space



B: Full REE corpus provided as context

WHERE DO AI-GENERATED IDEAS LAND?

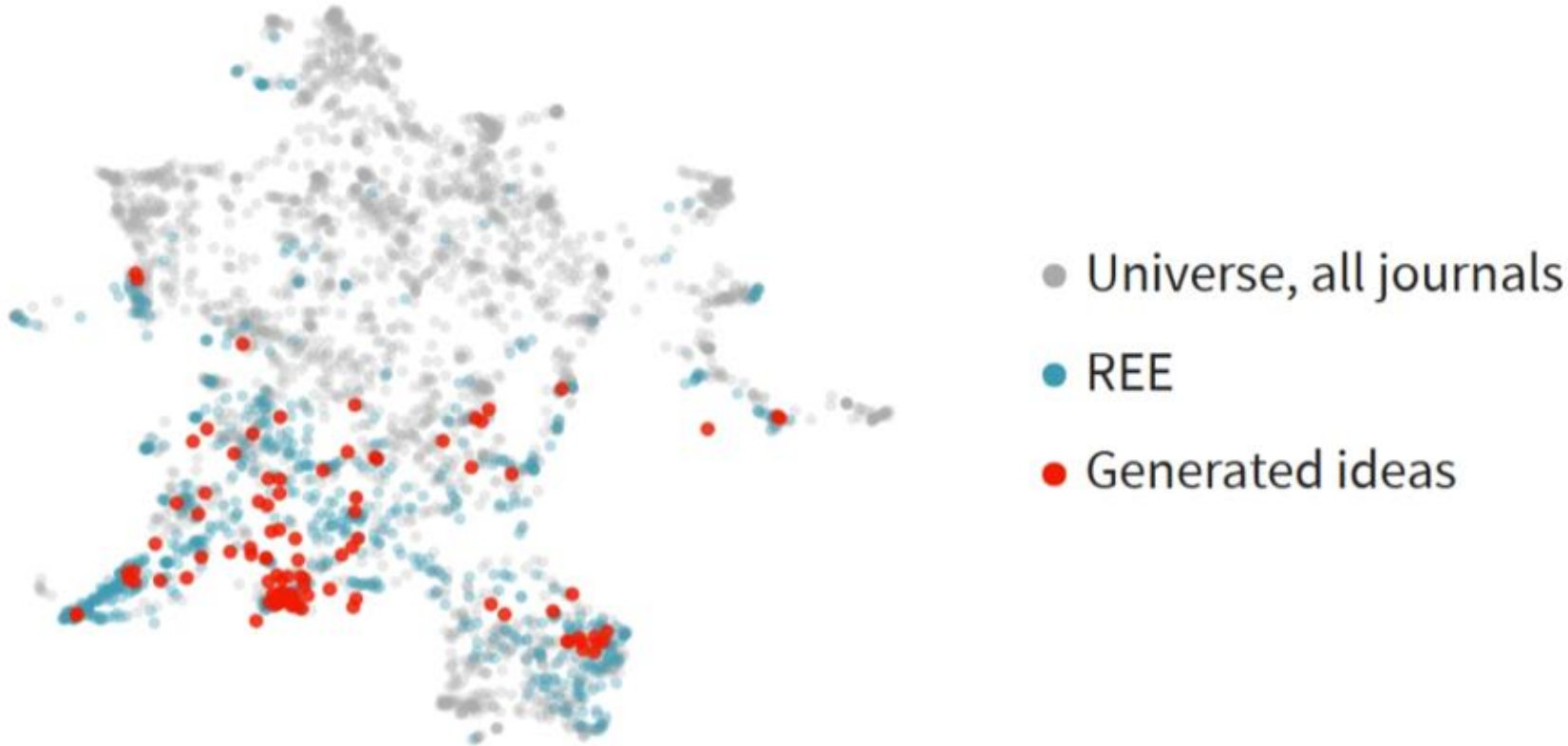
Each dot is a generated idea projected into the research space



C: Transfer of research methods and ideas from Econ/Finance

WHERE DO AI-GENERATED IDEAS LAND?

Each dot is a generated idea projected into the research space



D: Transfer of research methods and ideas from Psychology

SCORING GENERATED IDEAS

Three dimensions, each computed from sentence embeddings or LLM judgement

- **Novelty** — nearest-neighbour distance to all existing papers in the corpus. High novelty means the research question and approach are unlike anything published before.
- **Atypicality** — how unusual the *combination* of topic cluster and method is, relative to what typically appears together. Atypical work pairs a standard question with a rare method, or vice versa.
- **Feasibility** — can this idea plausibly be executed? Scored by the LLM against data availability, methodological maturity, and scope.

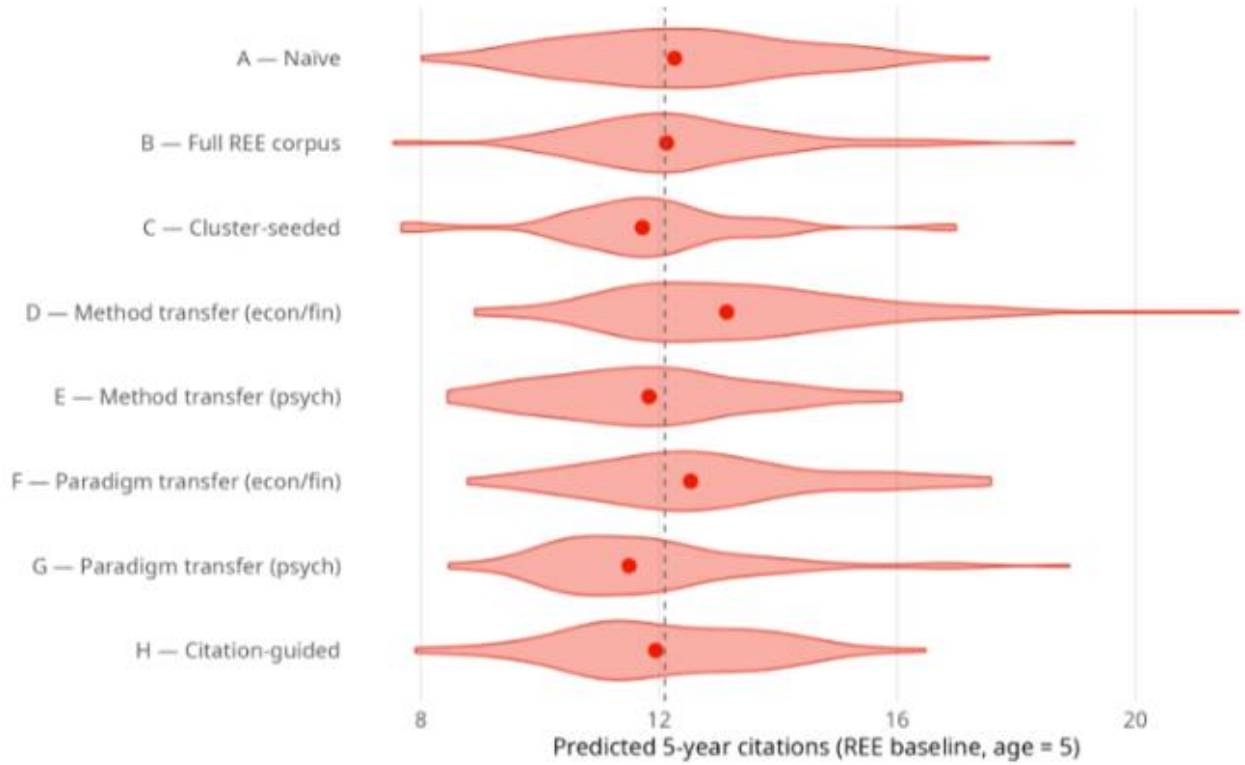
COMPARING GENERATED RESEARCH IDEAS

Unconstrained variant (a); 100 ideas per condition except C (51, one per cluster)

Condition	Clusters	HHI	Atypicality	Periph./REE	NN dist./REE	Pred. cit.
A — Naïve	21 / 51	0.070	0.726 ± 0.180	3.16	2.57	11.9
B — Full REE corpus	23 / 51	0.074	0.707 ± 0.248	2.75	2.48	11.9
C — Cluster-seeded	31 / 51	0.043	0.720 ± 0.206	2.89	2.48	11.5
D — Method transfer (econ/fin)	25 / 51	0.089	0.735 ± 0.147	3.70	2.19	**13.4**
E — Method transfer (psych)	21 / 51	0.112	0.693 ± 0.210	1.85	2.70	11.2
F — Paradigm transfer (econ/fin)	23 / 51	0.080	0.735 ± 0.158	4.36	2.21	**13.1**
G — Paradigm transfer (psych)	15 / 51	0.161	0.632 ± 0.251	1.74	2.67	11.2
H — Citation-guided	27 / 51	0.066	0.729 ± 0.179	2.80	2.36	12.5

PREDICTING IMPACT OF IDEAS

Predicted 5-year citations by generation condition — constrained variant



$$\log(1 + cit_i) = \alpha + \beta_1 \log(age) + \beta_2\, atypicality + \beta_3\, nn\ dist + \beta_4 \log(frontier\ dist) + \beta_5\, cluster\ mean\ cit + \text{journal FE}$$

R^2 = 0.182; n = 1,074; evaluated at REE baseline, age = 5

THE CHANGE IS HERE

The question is not IF AI will change research—only when we will realise it.

- *Progressive:* "Go on, strike away — quickly! How long am I to go on kneeling?"
  Moderates: "Why, you are already beheaded, if you would only be so kind as to nod."



Source: "Spanische Geschichte", Fliegende Blätter (1845)