

Controllable Clustering with LLM-Driven Embeddings

Kerria Pang-Naylor, Shivani Manivasagan, Amy Zhong, Mehak Garg, Nick Mondello, Blake Buckner, Jonathan P. Chang, Khyati Mahajan, Masoud Hashemi, Fabio Casati

Goal: Discovering patterns with **LLM perspective-driven** clustering on **multidimensional** text data

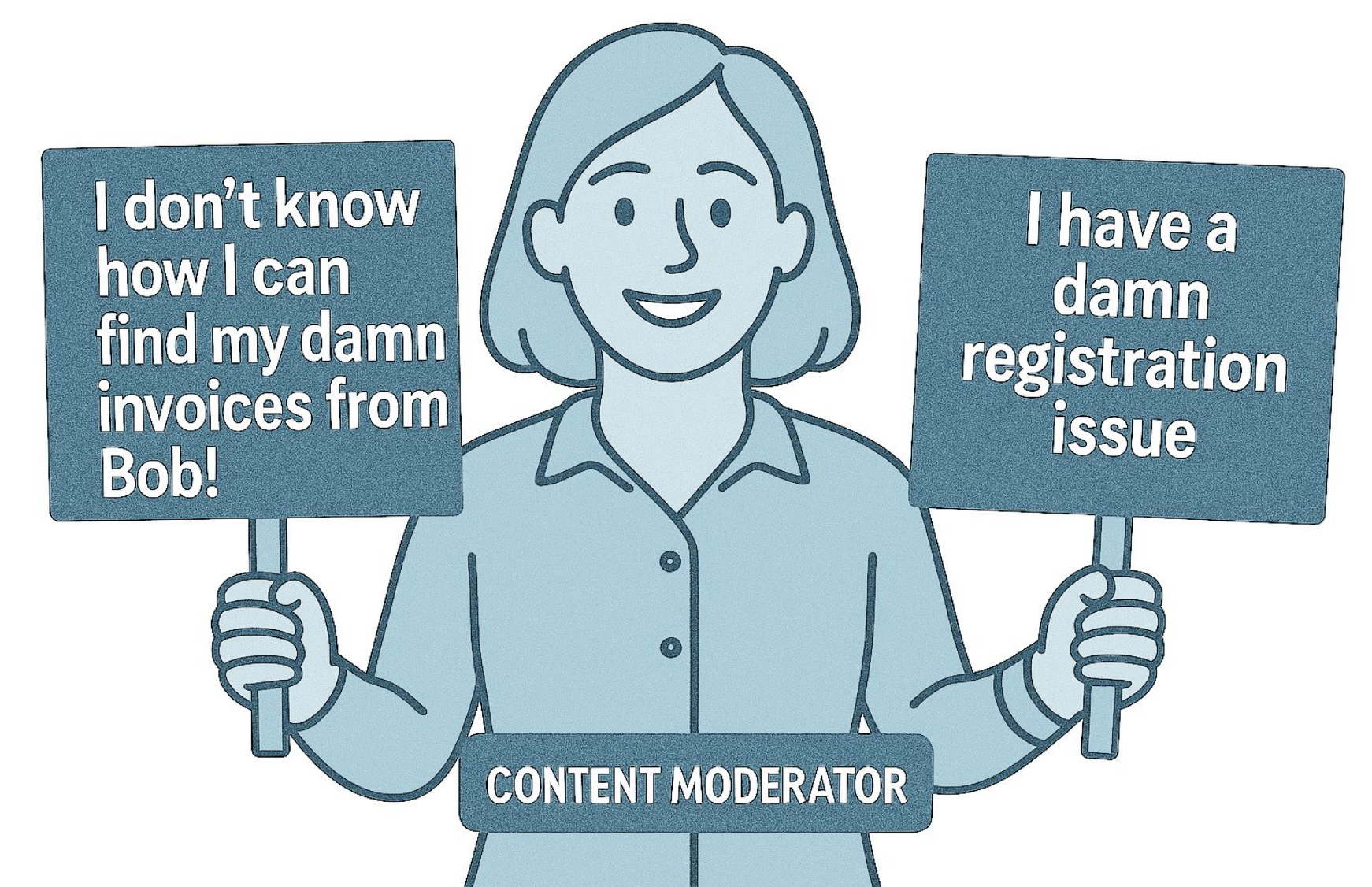


Could you help notify me of **payment** problems?

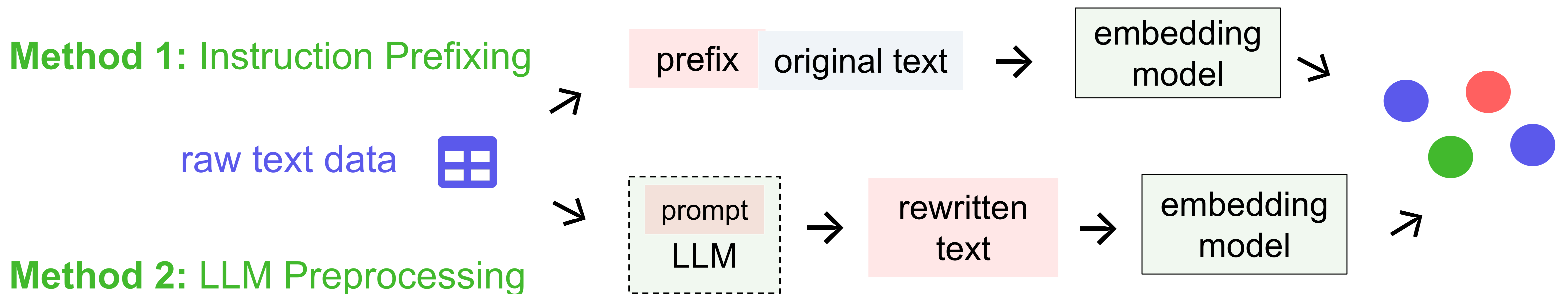
I don't know how I can find my **damn** invoices from Bob!

Could you help me edit my user information?

I have a **damn** registration issue.



Enhanced Clustering Pipeline - Achieving Perspective with LLMs



Results & Takeaways

Customer Service Dataset Clustering Results



Generic embeddings will favor one perspective while completely ignoring the alternative.

Exclusion preprocessing to ignore the dominant perspective improves the alternative perspective.

Prefixing can dramatically improve clustering performance, particularly for the non-dominant perspective.

