Revolutionizing Research: Unleashing Creativity with Yi 34B and Langchain

Raghavan Muthuregunathan

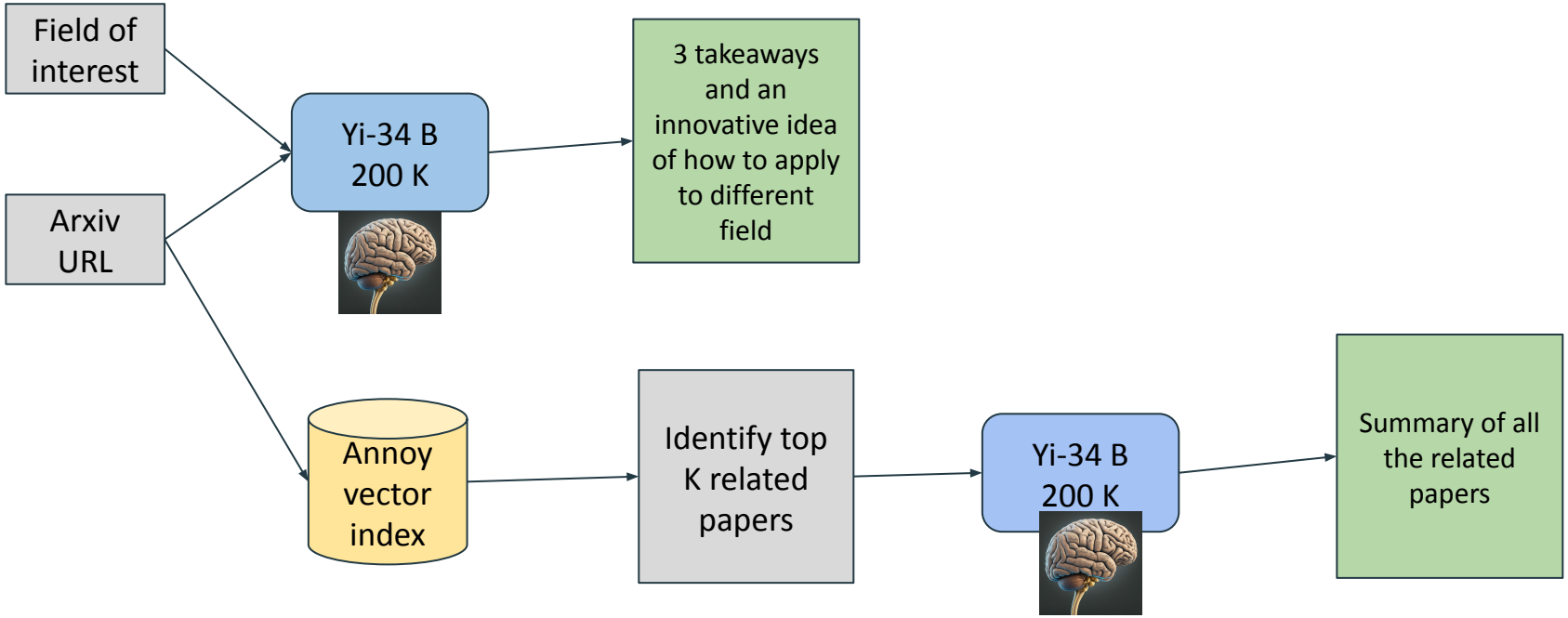
Loom: https://www.loom.com/share/a9d151a59a7c4fe68a8cc94b6a118d9d?sid=90a5d924-d995-4716-8a3b-c4da861e2e1d

Problem

- 100s of new arxiv papers are uploaded everyday
- How do I get new ideas to apply them for my field ?
- How do i get a quick gist of the paper and related papers?

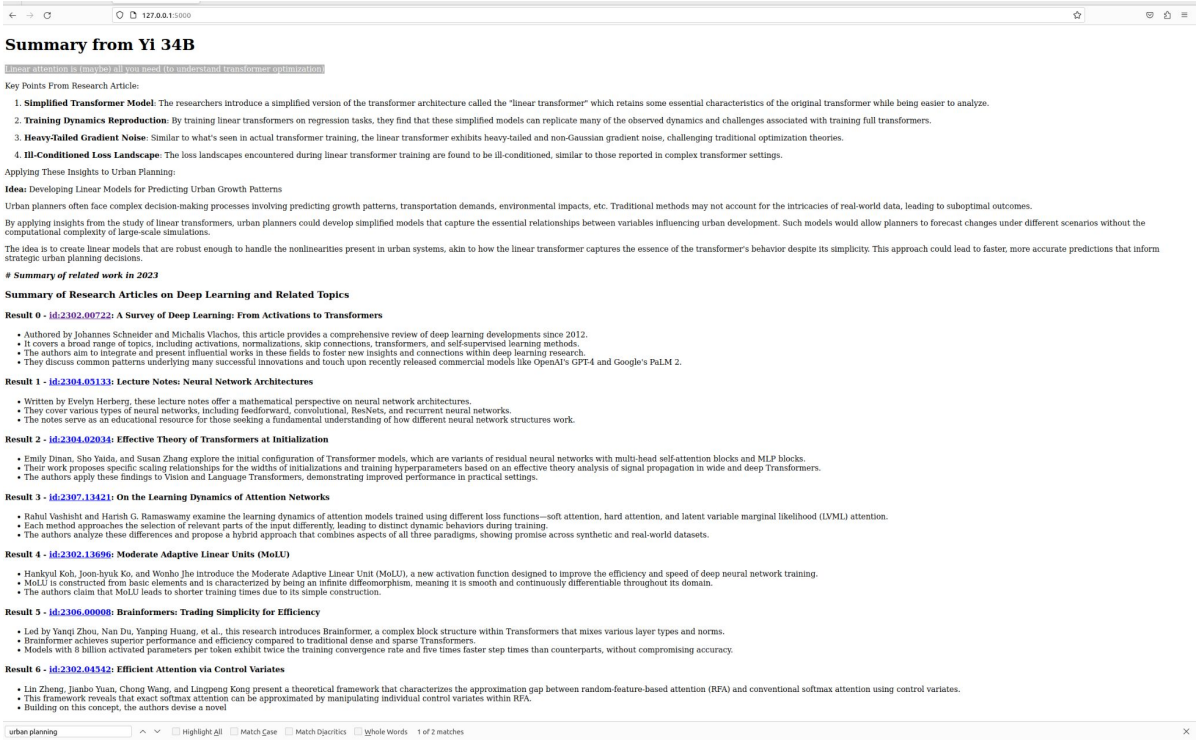
Loom video of Demo:
https://www.loom.com/share/a9d151a59a7c4fe68a8cc94b6a118d9d?sid=90a5d924-d995-4716-8a3b-c4da861e2e1d

Yi with its mighty 200K token to the rescue



Demo:
URL
https://arxiv.org/abs/2310.01082
Paper: linear attention is (may be) all you need
field: **urban planning**

- Yi34B generated 4 key takeaway by reading the entire paper
- Generated an idea for “urban planning”
- Summarized all the papers of 2023



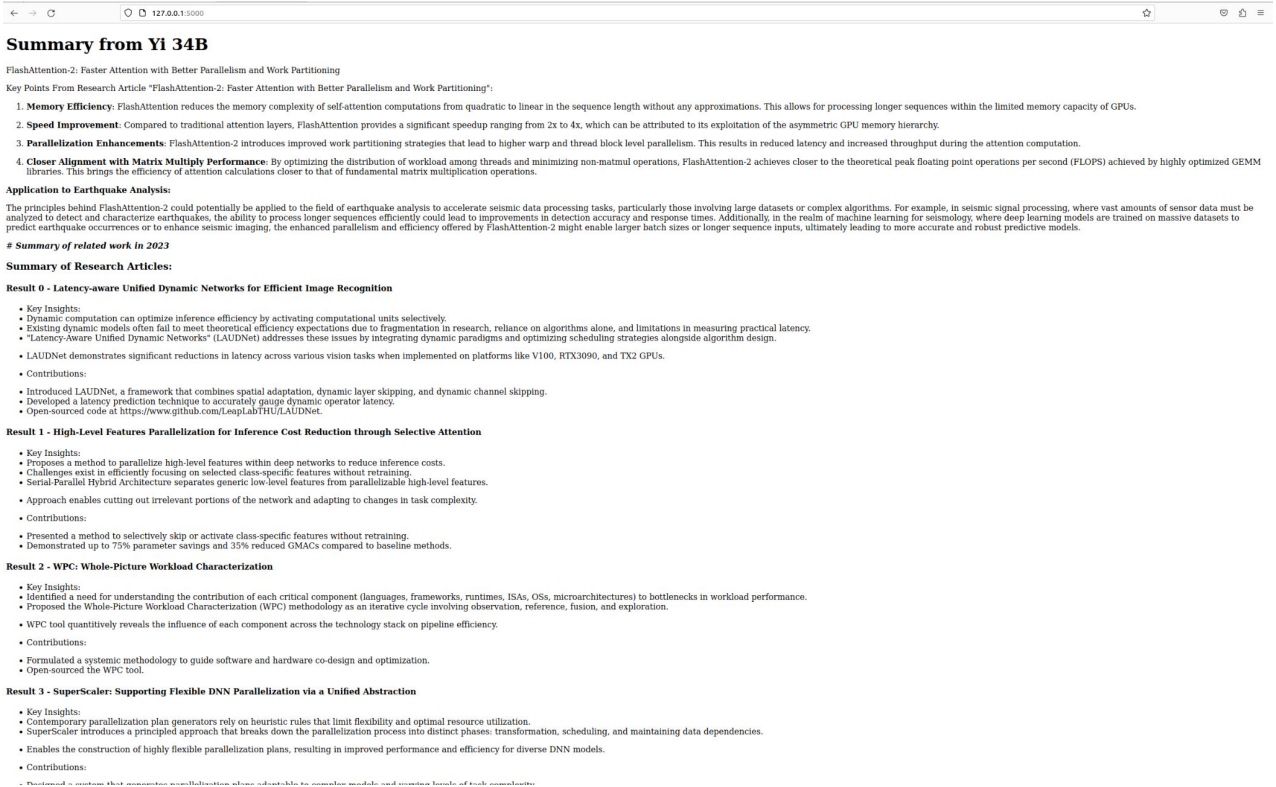
URL:
https://arxiv.org/abs/2307.08691

Paper is about flash attention-2 that is better parallelized

Yi34B summarized the 4 **key points** of the paper by reading the entire paper

Applied to the field of **earthquake data processing** and gave an actionable advice

Summarized all the related papers in 2023.



Applications

1. Researchers can get a quick summary of the paper
2. Researchers can get new ideas by applying the paper to a completely new field
3. Researchers can use the 200K limit and get summary of all the related papers too at the same time

Github

- https://github.com/Raghavan1988/yi34b200k_hackathon
- To run on terminal
  - Pip install -r requirements.txt
  - Python arxiv_summarizer_terminal.py <URL> <field>
- To run flask application
  - Pip install - requirements.txt
  - export FLASK_APP=arxiv_summarizer.py
  - Flask run

Both requires **replicate token**

export REPLICATE_API_TOKEN=<your TOKEN> i can provide the token for judges

Thank you

My linkedin: https://www.linkedin.com/in/raghavanmit/

Reach out to me for any questions.

Loom Video:
https://www.loom.com/share/a9d151a59a7c4fe68a8cc94b6a118d9d?sid=90a5d924-d995-4716-8a3b-c4da861e2e1d