Paper Craft AI
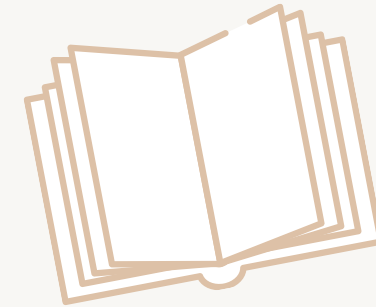
An ML Paper Recommendation System

For People who love researching and reading

Vishnu Priya Thanda
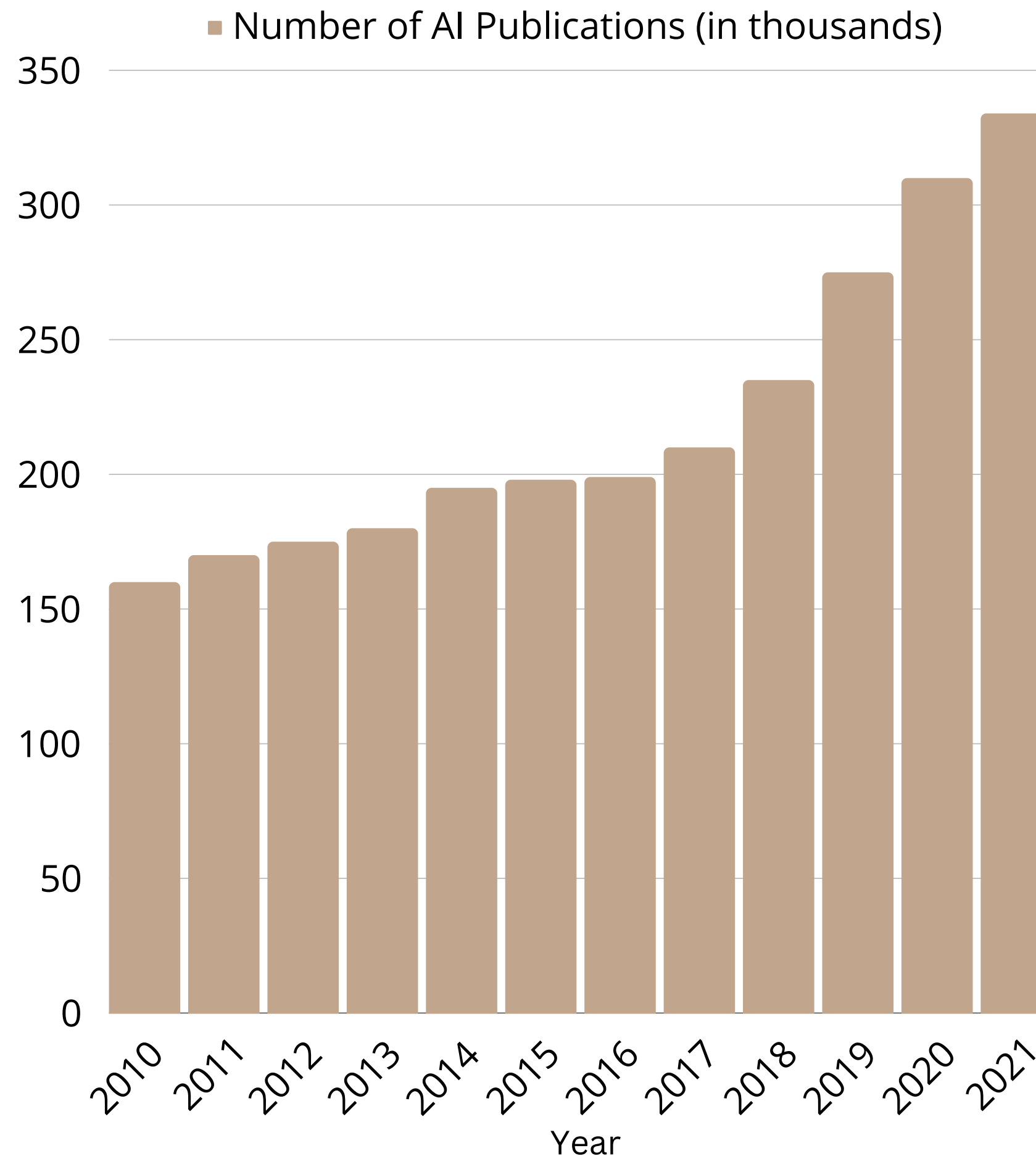Under the guidance of: Dr. Chaojie (Jay) Wang
UMBC

Outline

1. Background
2. Data Collection
3. EDA
4. Implementation
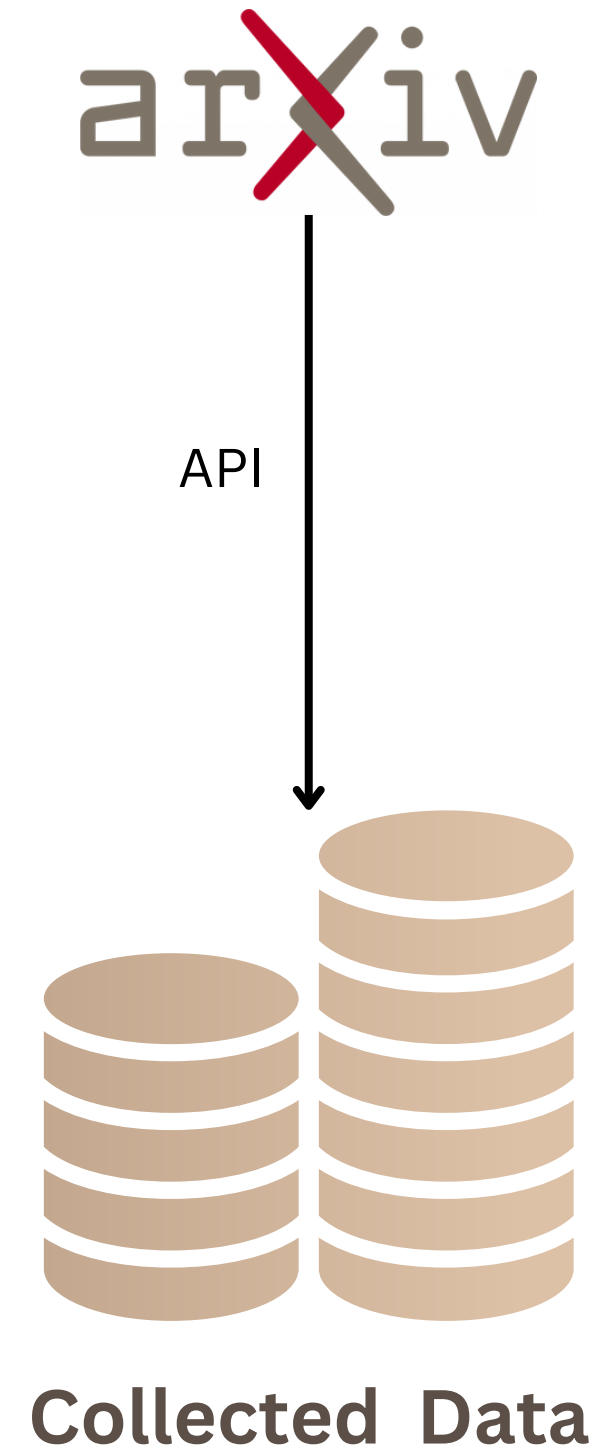5. Embedding Models
6. Demo
7. Conclusion

Number of AI Publications (in thousands)

Data Collection

- Extracted data from arXiv API

- 9500 Machine Learning Papers

- Each paper has data about summary, title, authors and link to the paper.

arXiv

API

Collected Data

Sample Data

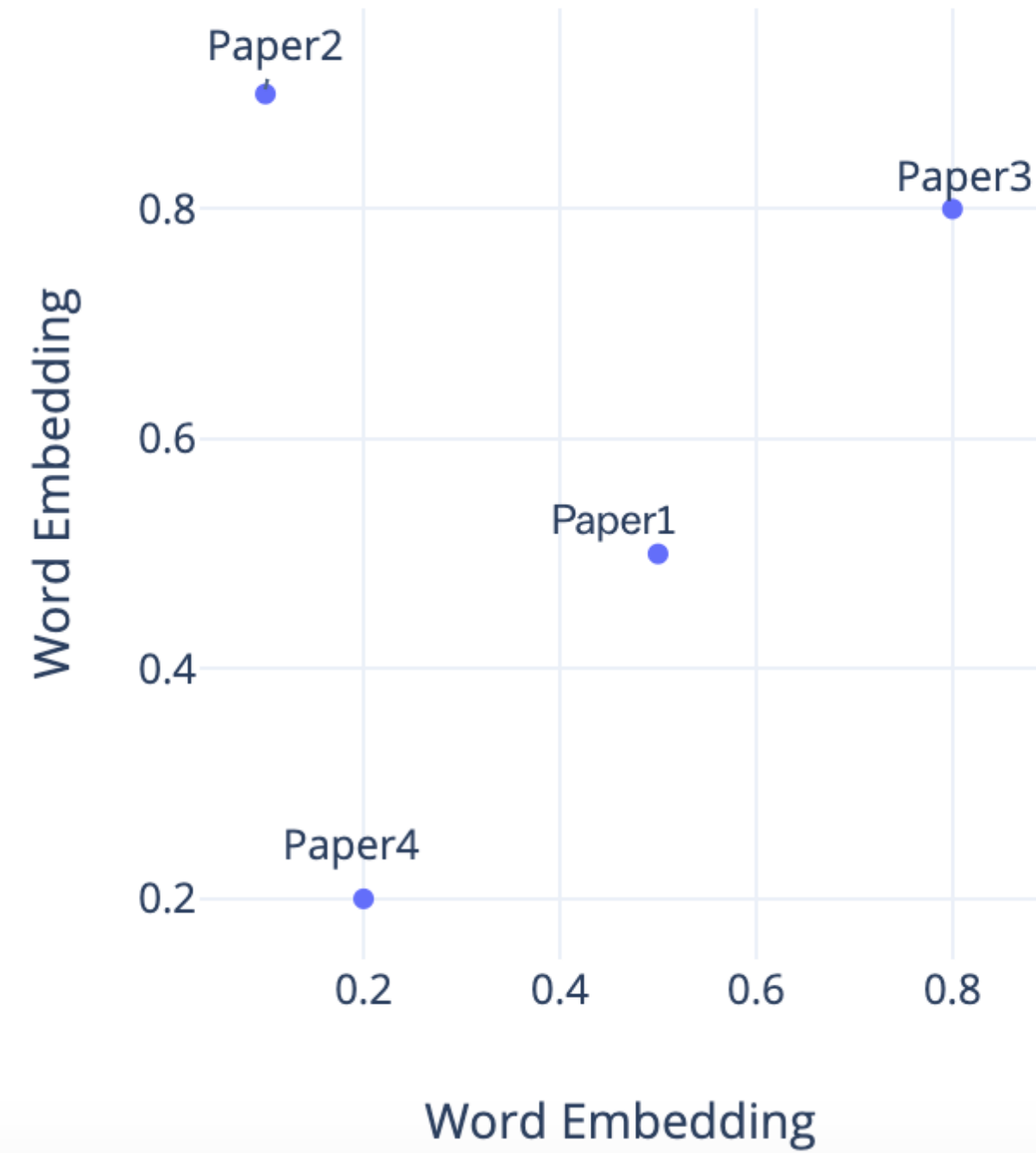| Id	Title	Summary	Authors	Published	Link
1	A Market-Oriented Programming Environment and	Market price systems constitute a well-under...	M. P. Wellman	2022-08-08	http://arxiv.org/abs/cs/9308102v
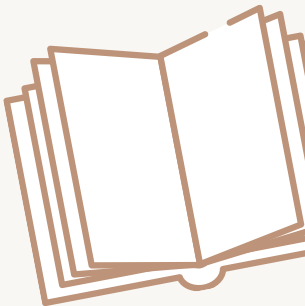
Top Bi-Grams of the Summary
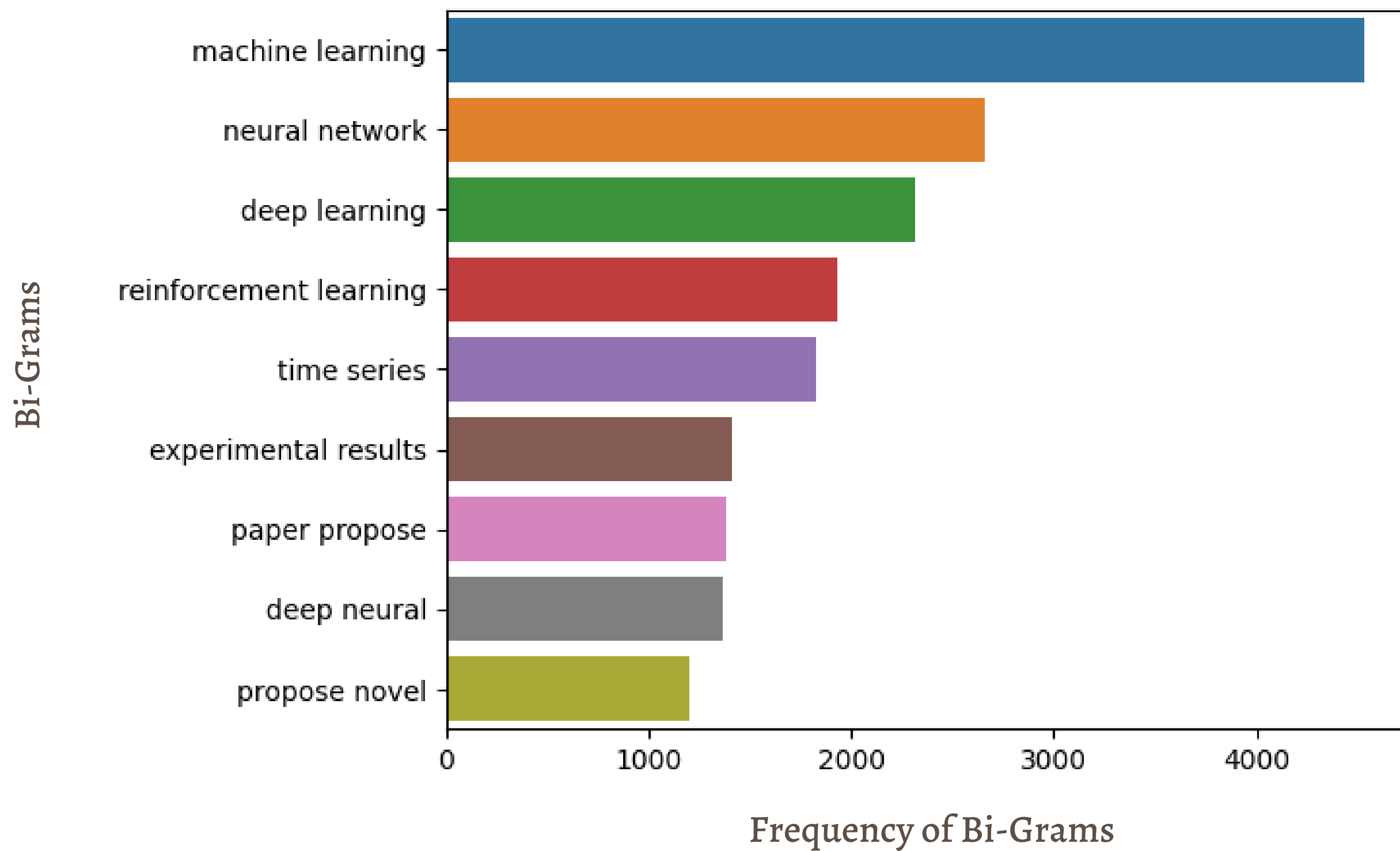
Word Embeddings Models

- Tf-IDF

- Sentence Transformers

- Specter 2.0 (Bert Based)

System

```
┌─────────────────┐       ┌─────────────────┐       ┌─────────────────┐
│                 │       │                 │       │  • Paper1       │
│  Select a paper/│──────▶│  Paper Details  │──────▶│  • Paper2       │
│  Search for a   │       │                 │       │  • Paper 3      │
│  Paper          │       │                 │       │  • Paper 4      │
│                 │       │                 │       │  • Paper5       │
└─────────────────┘       └─────────────────┘       └─────────────────┘
```

Demo

Conclusion

- Recommendation tool helpful for research.

- Collaborative Filtering based on the user interactions.

- Try multiple new embedding models like ada from OpenAI.

THANK YOU