

Software Engineering and Foundation Models: Insights from Industry Blogs Using a Jury of Foundation Models

Hao Li

hao.li@queensu.ca



This work was made possible with the support of my collaborators



Cor-Paul Bezemer



UNIVERSITY
OF ALBERTA



Ahmed E. Hassan



Queen's
UNIVERSITY



How to cite this session?

```
@misc{Li2024AlwareFMJury,  
author = {Hao Li and Cor-Paul Bezemer and Ahmed E. Hassan},  
title = {Foundation Models and Software Engineering: Insights from Industry Blogs},  
howpublished = {Tutorial presented at the Alware Leadership Bootcamp 2024},  
month = {November},  
year = {2024},  
address = {Toronto, Canada},  
note = {Part of the Alware Leadership Bootcamp series.},  
url =  
{https://aiwarebootcamp.io/slides/2024\_aiwarebootcamp\_li\_foundation\_models\_and\_software\_engineering.pdf}}
```



Check this paper for more information about this session

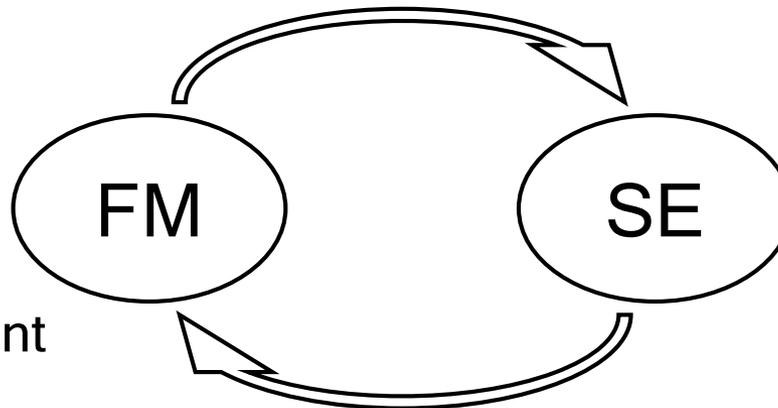
```
@article{li2024fmjury,  
title={Software Engineering and Foundation Models: Insights from Industry Blogs Using a Jury of Foundation Models},  
author={Hao Li and Cor-Paul Bezemer and Ahmed E. Hassan},  
journal={arXiv preprint arXiv:2410.09012},  
year={2024}  
}
```



Why study foundation models (FMs) and software engineering (SE)?

- FM4SE: FMs are transforming SE by automating various SE tasks and workflows
- SE4FM: SE practices are being adapted to support the development and deployment of FM(ware)

- Model serving & scaling
- Workflow orchestration
- Data management
- ...



- Code generation
- Code assistant
- Vulnerability detection
- ...



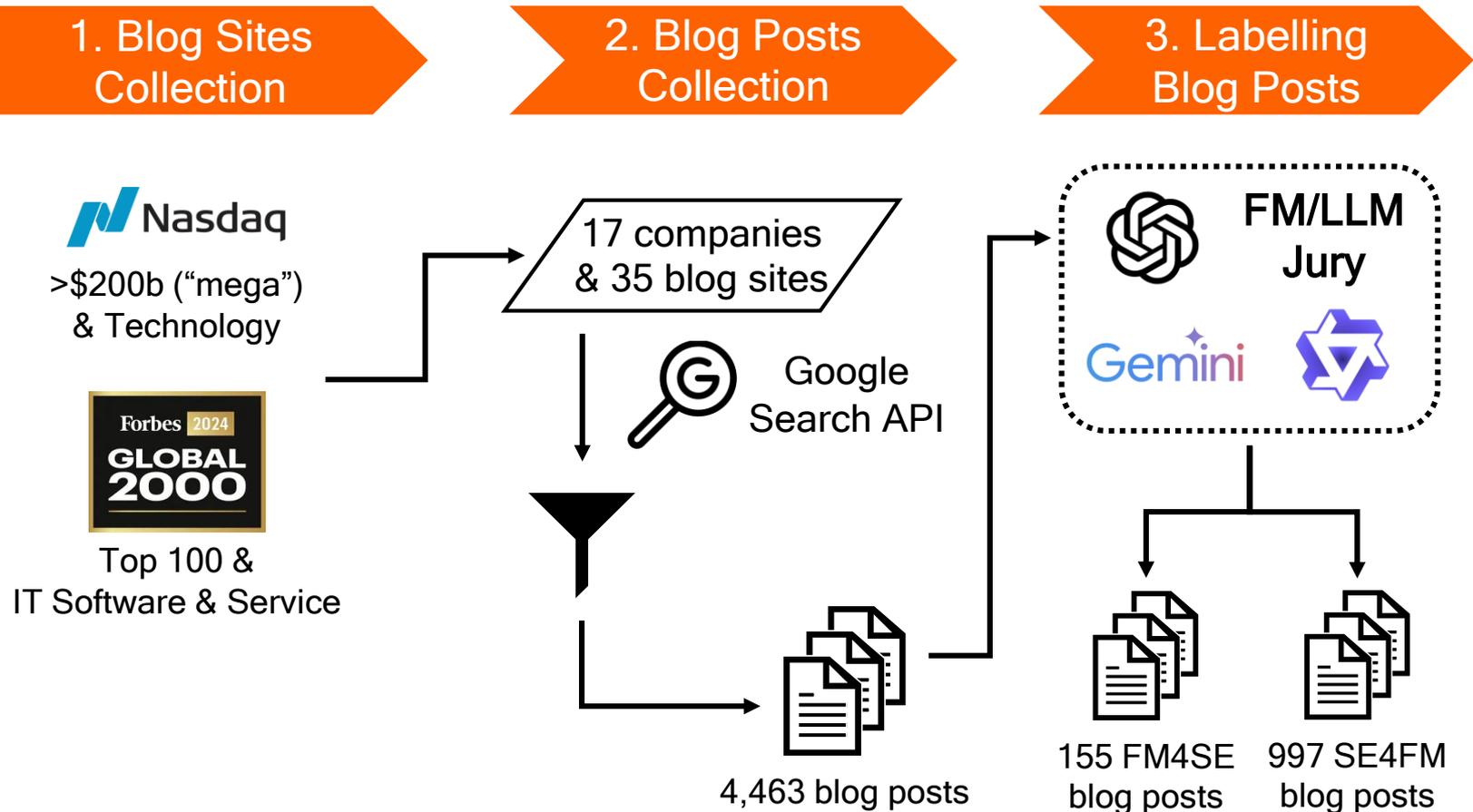
While academic research is growing, industry perspectives have been missing

Tech companies often publish blog posts to establish leadership in innovation

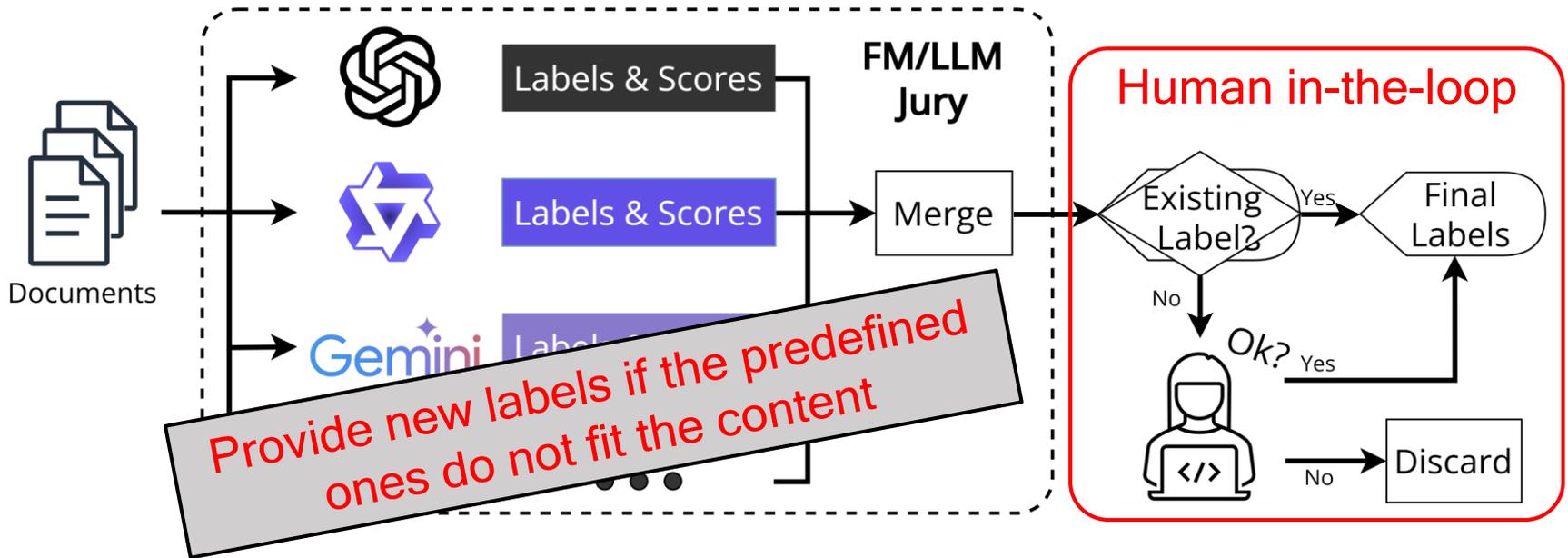
The image is a collage of three tech company blog pages. On the left is the AWS Blog, featuring a 'Featured posts' section with two articles: 'Amazon Aurora PostgreSQL and Amazon DynamoDB zero-ETL integrations with Amazon Redshift now generally available' and 'New courses and certification from AWS Training and Certification October 2024'. On the right is the NVIDIA Developer Technical Blog, showing a post titled 'IBM's New AI Models Are Smarter, More Accurate and Efficient'. A third blog post is partially visible at the bottom right, titled 'Developer'. Three grey callout boxes with red text are overlaid on the collage: 'Real-world insights' is positioned over the NVIDIA post, 'Practical use cases' is over the AWS posts, and 'Fast-moving trends' is over the bottom right post.



We collected blog posts from leading companies and employed the FM Jury to label them



FM Jury: an efficient approach for surveying grey literature



- Mitigate model bias and enhance accuracy
- Reduce human effort
- Scalable for surveying fast-evolving fields



Leveraging FM Jury for labelling SE-FM area, FM4SE activities, and SE4FM activities

Model	SE-FM Area	FM4SE Activities	SE4FM Activities
Gemini-1.5-Flash-002	0.65	0.66	0.81
GPT-4o-mini-2024-07-18	0.70	0.74	0.81
Qwen2-72B-Instruct	0.85	0.79	0.76
FM/LLM Jury	0.95	0.87	0.91

- **SE-FM area:** FM4SE, SE4FM, and unrelated
- **FM4SE activities:** Software development, quality assurance, maintenance, ...
- **SE4FM activities:** Model deployment, system architecture, data management, ...
- FM Jury achieved **excellent agreement (Cohen's $\kappa > 0.78$)** with human labels



In FM4SE blog posts, software development tasks are the most frequently discussed

Activity	Most popular task	%Posts	%Company
Software development	Code generation	78.1	100.0
Software quality assurance	Vulnerability detection	11.0	36.3
Software maintenance	Code refactoring	9.7	36.3
Software management	Tool configuration	0.6	9.1
Requirement engineering	Requirements analysis	0.6	9.1
Software design	-	0.0	0.0

Align with research interests



Limited focus on software management, requirements engineering, and software design

Activity	Most popular task	%Posts	%Company
Software development	Code generation	78.1	100.0
Software quality assurance	Vulnerability detection	11.0	36.3
Software maintenance	Code refactoring	9.7	36.3
Software management	Tool configuration	0.6	9.1
Requirement engineering	Requirements analysis	0.6	9.1
Software design	-	0.0	0.0

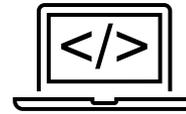
Research direction 1: Real-world validation of FMs in software management, requirements engineering, and software design



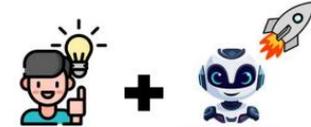
Code assistants provide multifunctional support beyond code generation



idea



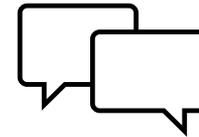
code



Software
Engineering 3.0



debug



chat



review



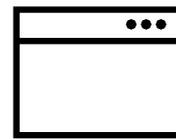
config

Research direction 2: Evaluating code assistants in software development workflows for tasks beyond code generation



Software maintenance tasks focus on legacy system refactoring and modernization

- FMs are widely used to refactor, translate, and transform entire systems



Application



Cloud

COBOL



Research direction 3: Leveraging FMs for legacy system modernization and transformation beyond code translation



Practitioners discussed SE4FM(ware) much more often than FM4SE



997 SE4FM
blog posts

VS.



155 FM4SE
blog posts

Research direction 4: Expanding research on SE4FM(ware), especially as FMs support modalities beyond text (e.g., image)



In SE4FM blog posts, model deployment & operation is the most frequently discussed

- Cloud deployment dominates due to the resource-intensive nature of FMs
- Model serving & scaling techniques are widely adopted
 - Speculative decoding, workload balance, automatic scaling
- Model compression reduces resource consumption
 - Pruning, quantization, distillation
- Growing interest in local deployment
 - Mobile, edge, PC

Research direction 5: Performance engineering (e.g., load testing and latency optimization) for FM(ware)

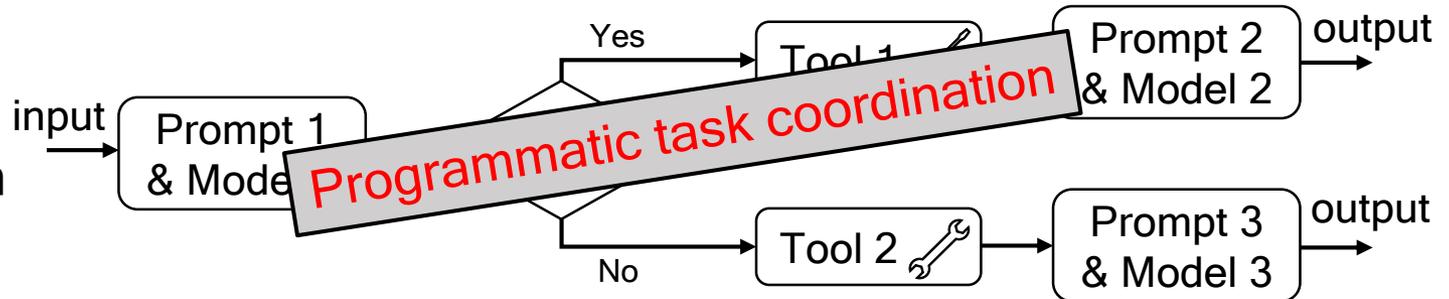


System architecture & orchestration tasks have become a popular topic in SE4FM blog posts

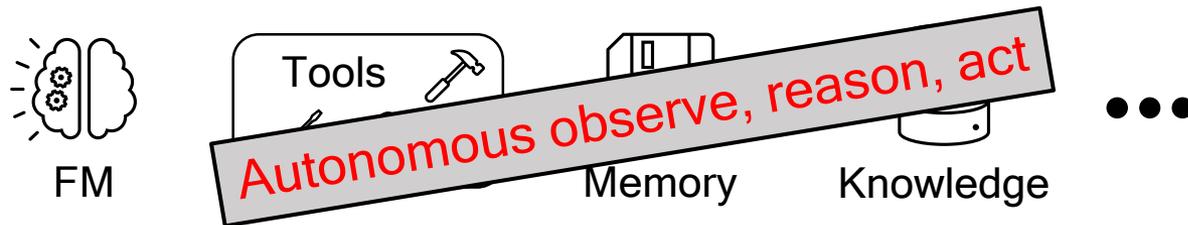
Model & prompt chaining



Workflow orchestration



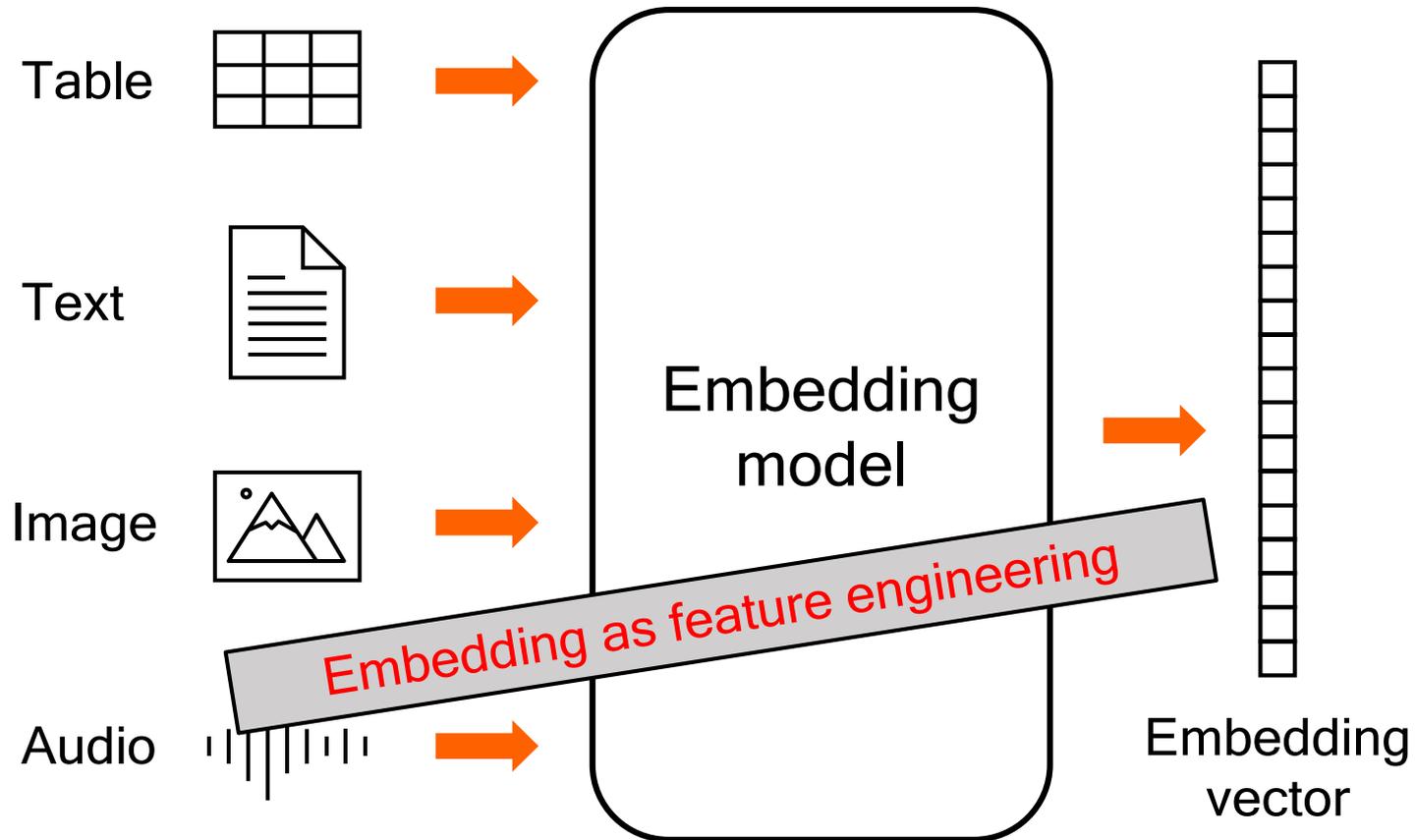
AI agent(s)



Research direction 6: Supporting FM chaining, workflow orchestration, and AI agents via SE activities (e.g., design patterns)



In FM era, data management has evolved to support both structured and unstructured data



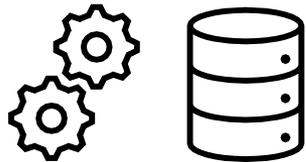
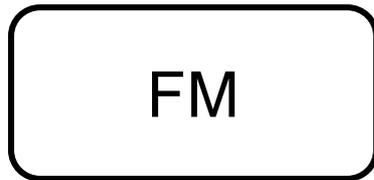
Specialized databases (e.g., VectorDBs) are designed for storing and retrieving embeddings

- VectorDBs are optimized for storing and retrieving embedding vectors efficiently
- Provide advanced features such as multimodal search, where users search image/video using text queries
- Enabling retrieval-augmented generation (RAG) which combines FMs with data retrieval from databases

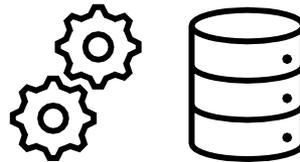
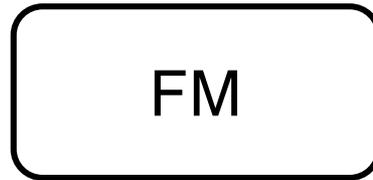
Research direction 7: Investigating how these evolving data management strategies affect the dev lifecycle of FMware.



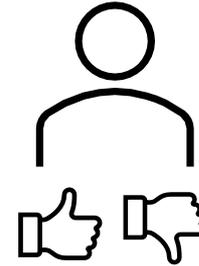
Model customization relies on fine-tuning methods such as LoRA and RLHF



Fine-tune on domain-specific dataset



Efficient Fine-Tuning with LoRA



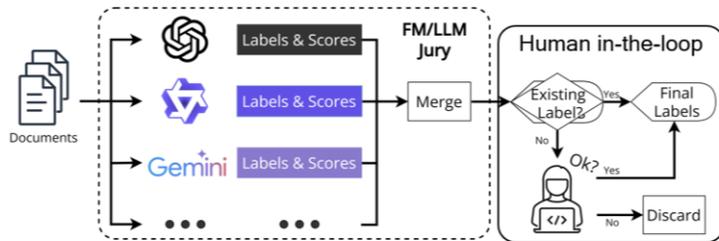
RLHF



Research direction 8: Investigating the impact of the shift from training models from scratch to fine-tuning on SE activities



FM Jury: an efficient approach for surveying grey literature



- Mitigate model bias and enhance accuracy
- Reduce human effort
- Scalable for fast-evolving fields

Limited focus on software management, requirements engineering, and software design

Activity	Most popular task	%Posts	%Company
Software development	Code generation	78.1	100.0
Software quality assurance	Vulnerability detection	11.0	36.3
Software maintenance	Code refactoring	9.7	36.3
Software management	Tool configuration	0.6	9.1
Requirement engineering	Requirements analysis	0.6	9.1
Software design	-	0.0	0.0

Research direction 1: Real-world validation of FMs in software management, requirements engineering, and software design

Practitioners discussed SE4FM(ware) much more often than FM4SE

155 FM4SE
blog posts

VS.

997 SE4FM
blog posts

Research direction 4: Expanding research on SE4FM(ware), which support modalities beyond text (image, audio, etc.)



Hao Li
hao.li@queensu.ca

