

Paktor Anot?

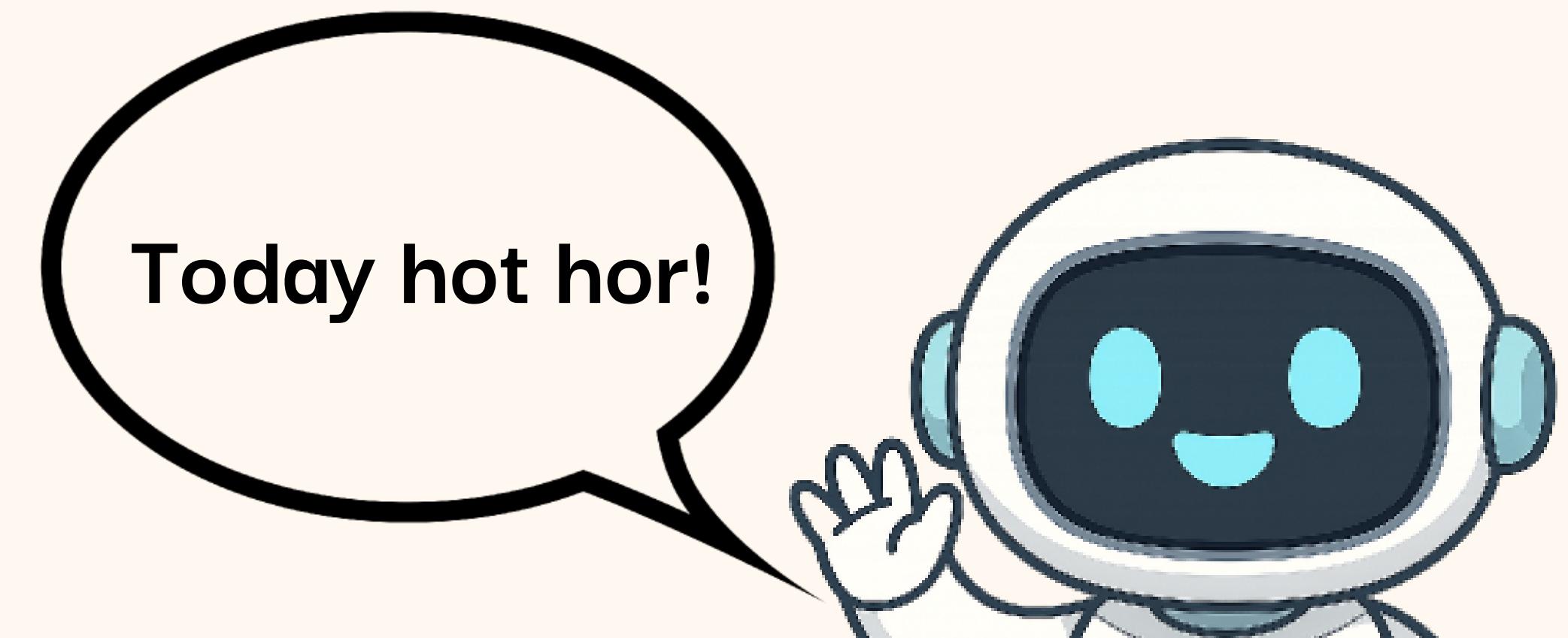
If ChatGPT Grew Up in Singapore.

Modern chatbots and dating
simulators sound generic and
culturally neutral.

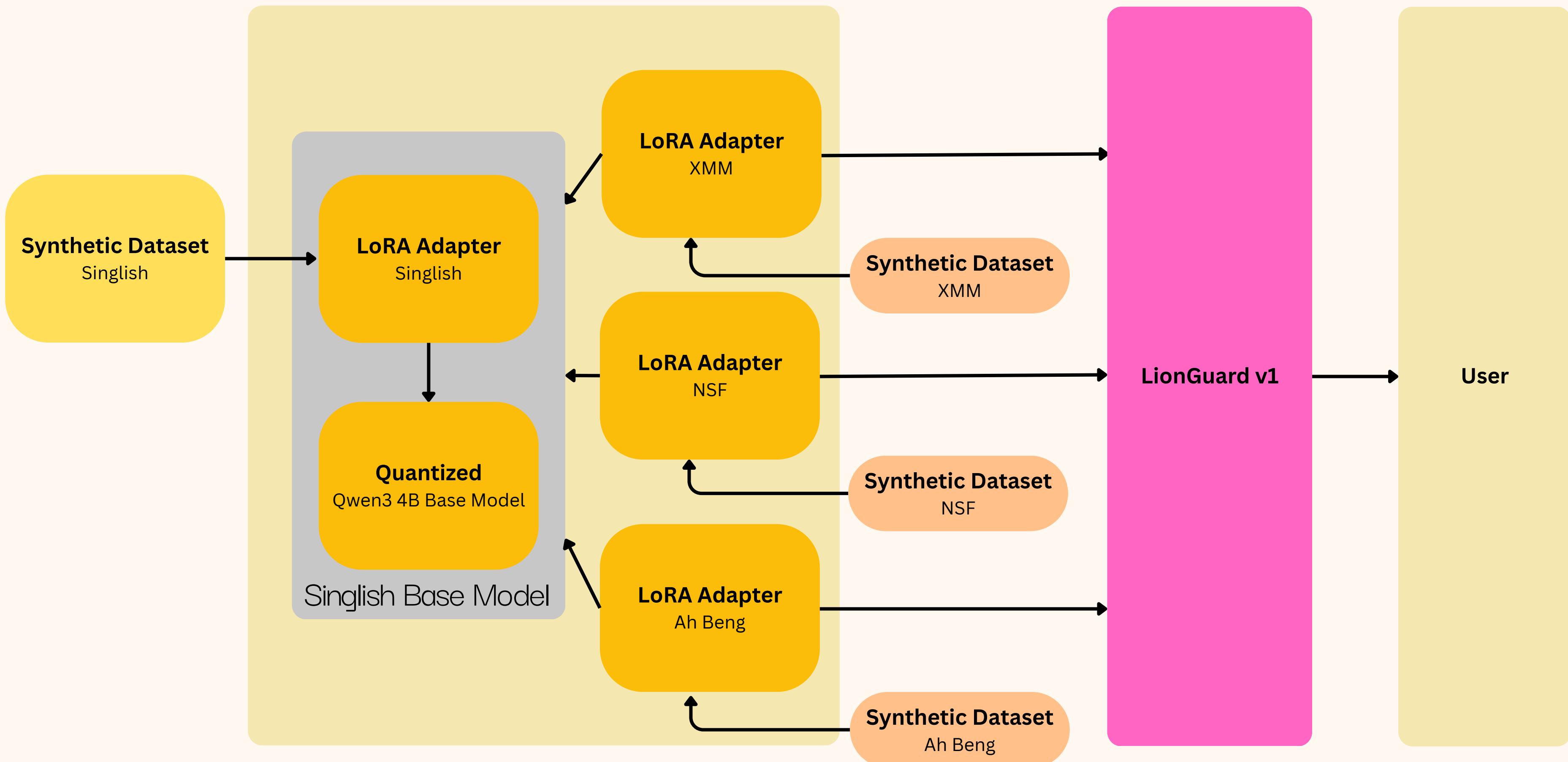


How can we **create** a **conversational AI** that:

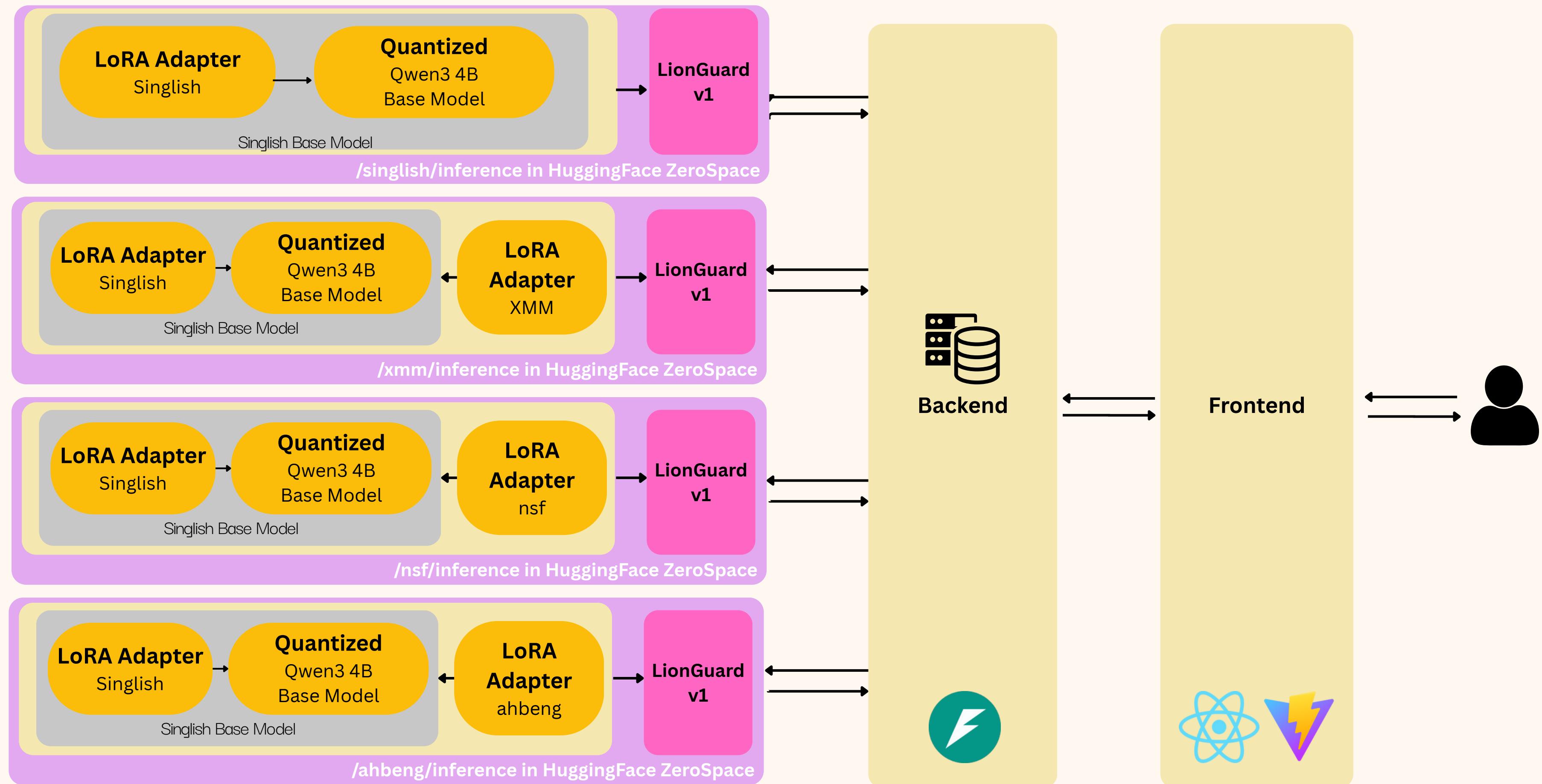
- Speaks **natural Singlish**
- Expresses **distinct local** personas
- Remains **safe and respectful** in **all contexts?**



OUR TRAINING APPROACH



OUR ARCHITECTURE



LoRA (Low-Rank Adaptation)

- Freeze base model weights
- Add small trainable “adapter” layers in a few places (e.g., attention)
- Train only adapters → fast, cheap, low VRAM

Quantization (Q in QLoRA)

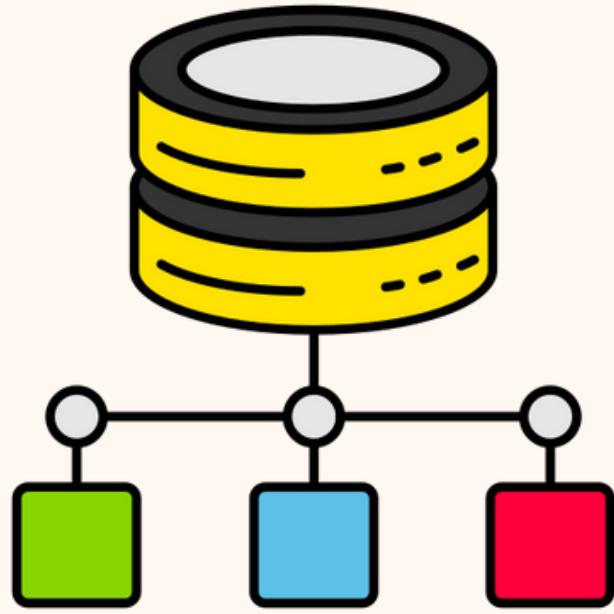
- QLoRA = LoRA + 4-bit base → same adapters, less memory

Supervised Fine-Tuning (SFT)

- Used to teach new behaviour (e.g., XMM, Beng, NSF personas)
- Works together with LoRA/QLoRA → adapters update persona, not full model



Challenges Faced



Creating high quality
persona specific data



Balancing persona tone
with safety rules



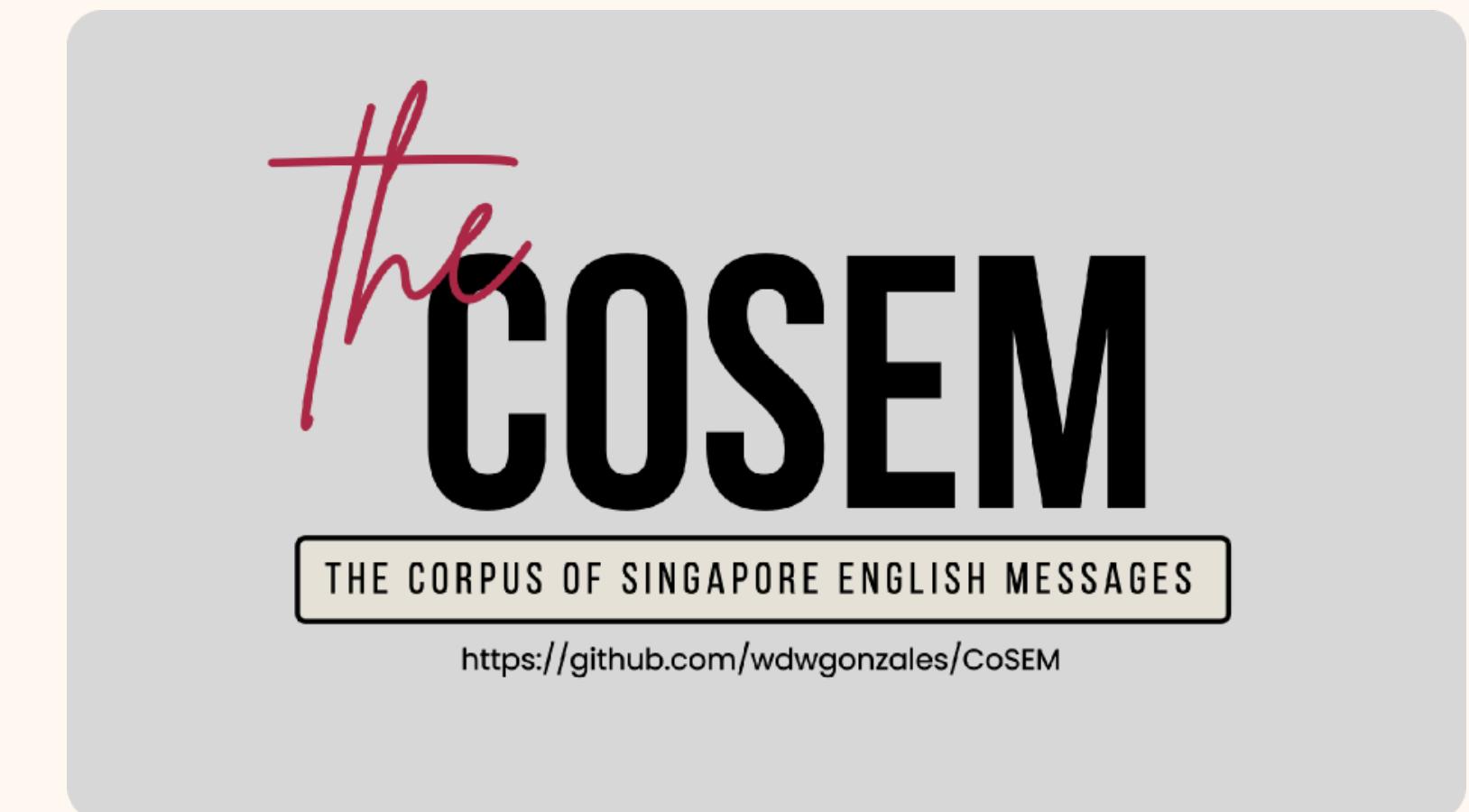
Making it sound more
human

CoSEM

- 3.6M WhatsApp/SMS Message
- Natural Singlish Code-Switching
- Anonymized (No Personal Info)

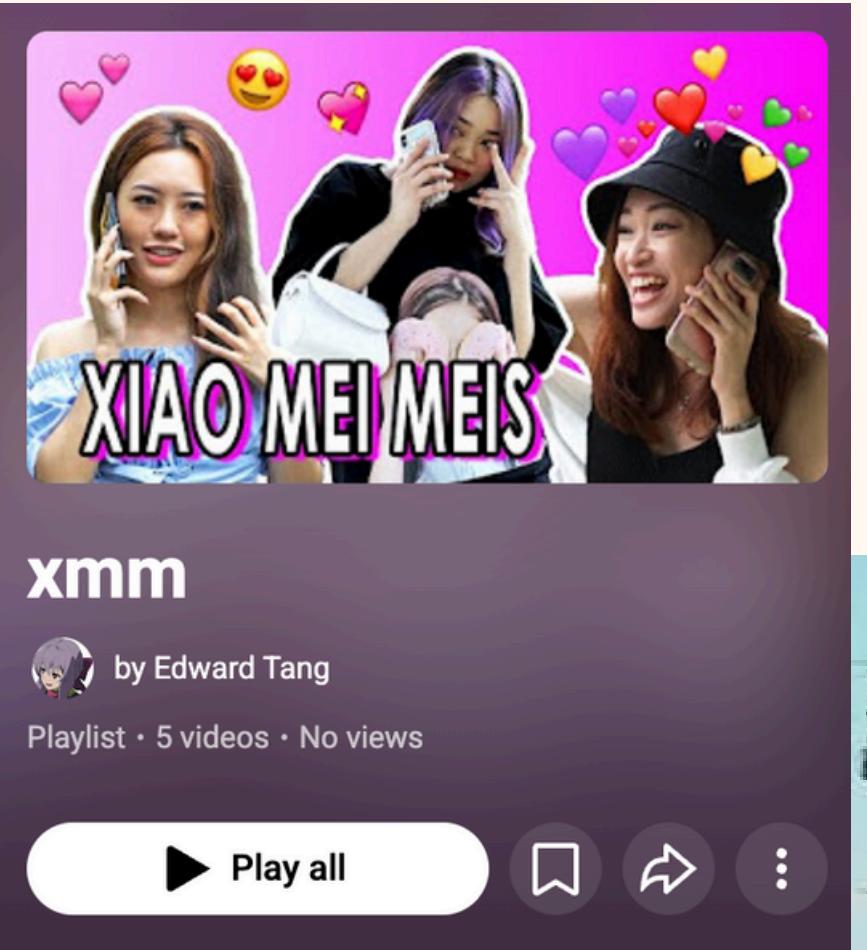
NUS SMS Corpus

Complementary 35,598 Singlish messages from younger demographics

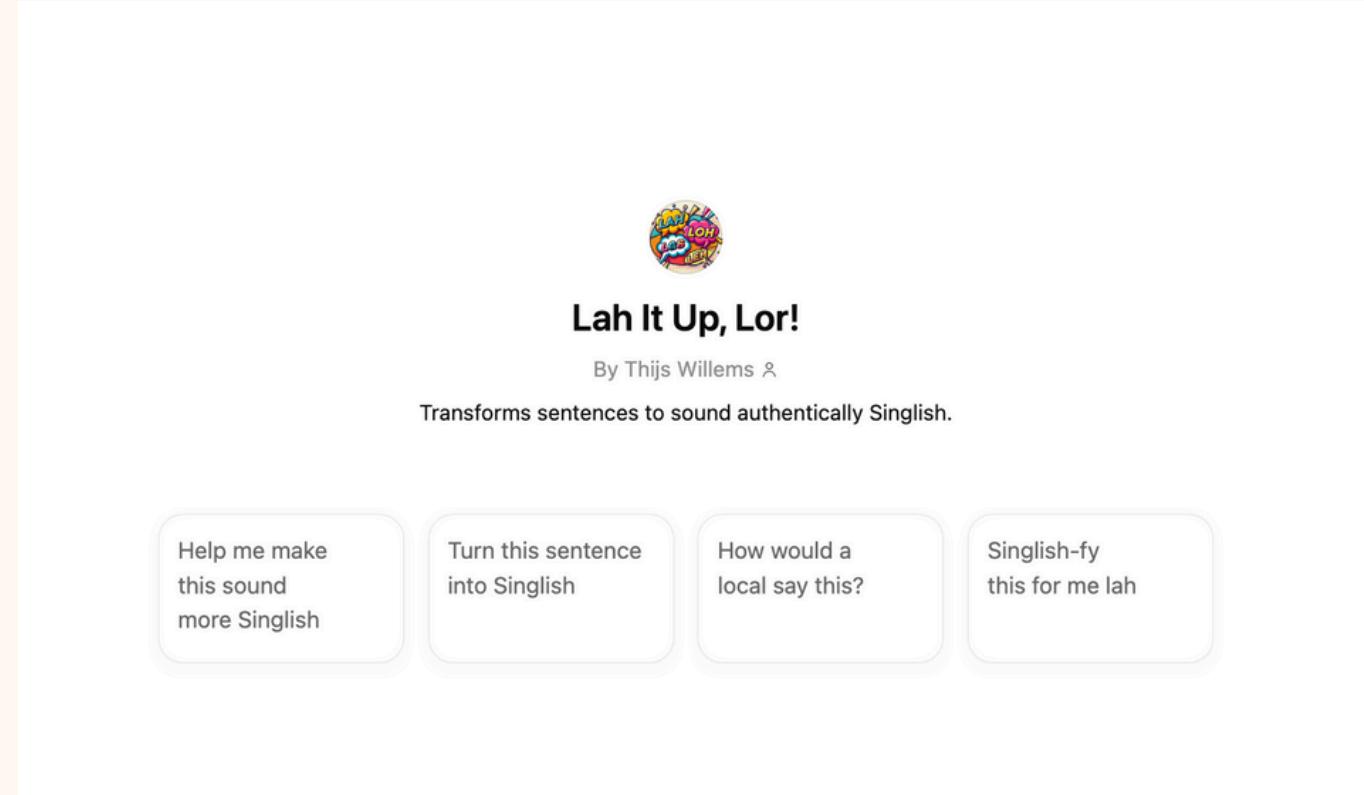


Transcription Approach

- Sample various local sources (e.g. local YouTube channels, local movies, etc.)
- Extract pure conversations (e.g. remove narrator lines).
- Issues:
 - Multi-turn conversations
 - Context-heavy, not suitable for instruction-response tuning.
 - Not all conversations exactly fit persona definition



Synthetic Dataset Creation



```
"messages": [  
    {  
        "role": "user",  
        "content": "Any nice cafes to study at? Prefer somewhere quiet with wifi."  
    },  
    {  
        "role": "assistant",  
        "content": "Go The Book Cafe or Twenty Grammes lah. Got wifi, got plug,  
        and people there don't talk cock loudly."  
    },  
    {  
        "role": "user",  
        "content": "Those got time limit or not?"  
    },  
    {  
        "role": "assistant",  
        "content": "Some got lah. You just buy one drink every few hours lor.  
        Don't be cheap until kena kick out."  
    },  
    {  
        "role": "user",  
        "content": "Thanks bro, I go check them out."  
    },  
    {  
        "role": "assistant",  
        "content": "Go lah, go early. Don't later no seat then you stand there  
        like lost puppy."  
    }  
]
```

[https://www.sutd.edu.sg/future-of-innovation/
use-cases/custom-gpts/](https://www.sutd.edu.sg/future-of-innovation/use-cases/custom-gpts/)

PERSONAS

🔥 Ah Beng 🔥

🥤 XMM 🥤

❗️ NSF ❗️





LionGuard

POWERED BY GOVTECH

Singapore-trained AI
model for moderating
and understanding
local language, slang,
and context.

Base Model Evaluation

Model	Train Time (min)	Train VRAM (GB)	Final Loss	Perplexity
1.7B	2.02	1.80	3.6624	88.14
4B-Instruct	2.22	5.40	2.8593	40.76
8B	2.69	11.66	2.8901	42.05

Model	Semantic Sim	Singlish Score	Latency (s)	Tokens/sec	Infer VRAM (GB)
1.7B	0.4981	1.15	1.965	14.3	12.36
4B-Instruct	0.5754	1.30	1.841	11.0	12.36
8B	0.6187	1.40	2.358	11.0	14.51

Singlish Base Model Evaluation

Model	Perplexity	Semantic Sim	Sem Min	Sem Std	Singlish Score
4B-singlish-base-v1	16.26	0.5373	0.1611	0.1639	1.80
4B-singlish-base-v2	33.43	0.5582	0.2755	0.1546	1.55
4B-singlish-base-v3	32.70	0.5855	0.2243	0.1741	1.25

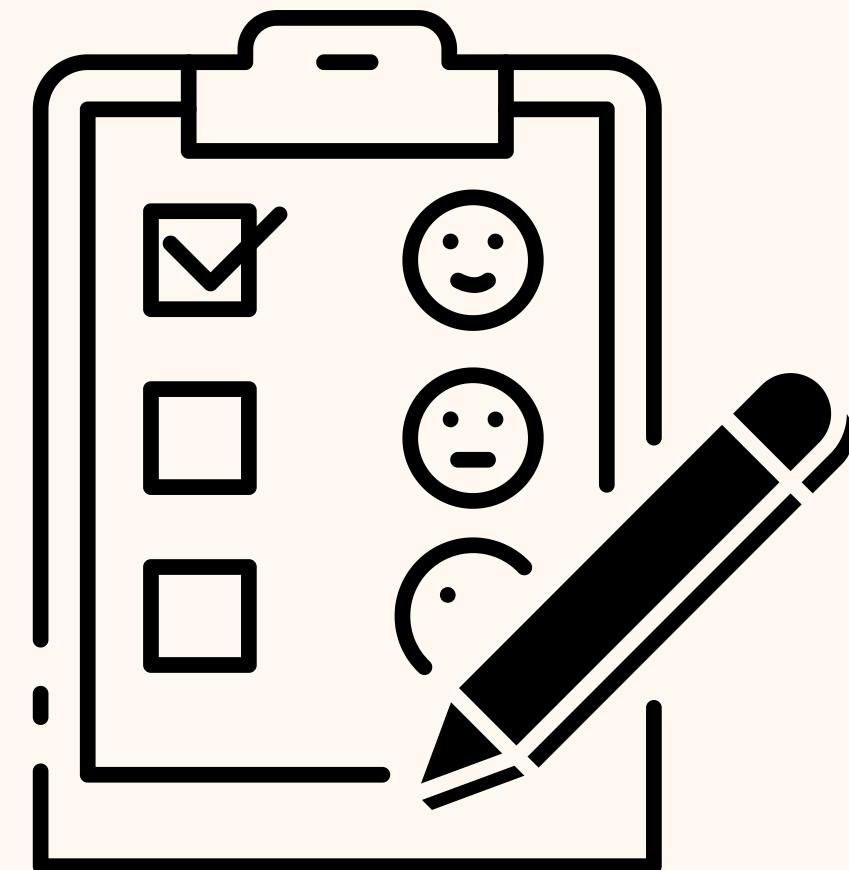
Model	Coverage %	Over-Singlish %	Latency (s)	Tokens/sec	Infer VRAM (GB)
4B-singlish-base-v1	85.0	0.0	1.761	10.2	12.16
4B-singlish-base-v2	80.0	0.0	1.405	17.9	18.14
4B-singlish-base-v3	85.0	0.0	1.326	18.0	18.14

Table 1: Performance Metrics for 4B Singlish Base Model Variants

Comparison of three candidate base models across linguistic performance, Singlish authenticity, efficiency, and inference speed.

Persona Evaluation

“Does this model feel like the intended persona when a real person interacts with it?”



Survey



Human Rating

Example Conversation

Tell them you still not ready lah. Marriage is serious one, don't rush. When you feel right, then talk.

02:51 pm

Ah, that kind of talk can be quite sian lor. Just steer the topic to something lighter, like food or events. If they insist, you can say you need to go and not stay too long.

02:51 pm

Tell them you want to enjoy properly, not just chiong for chiong. If not, you end up sian and regret.

02:51 pm

My family keeps asking when I am getting married. What should I do?

02:51 pm

They bring it up almost every time we meet and it is really tiring.

02:51 pm

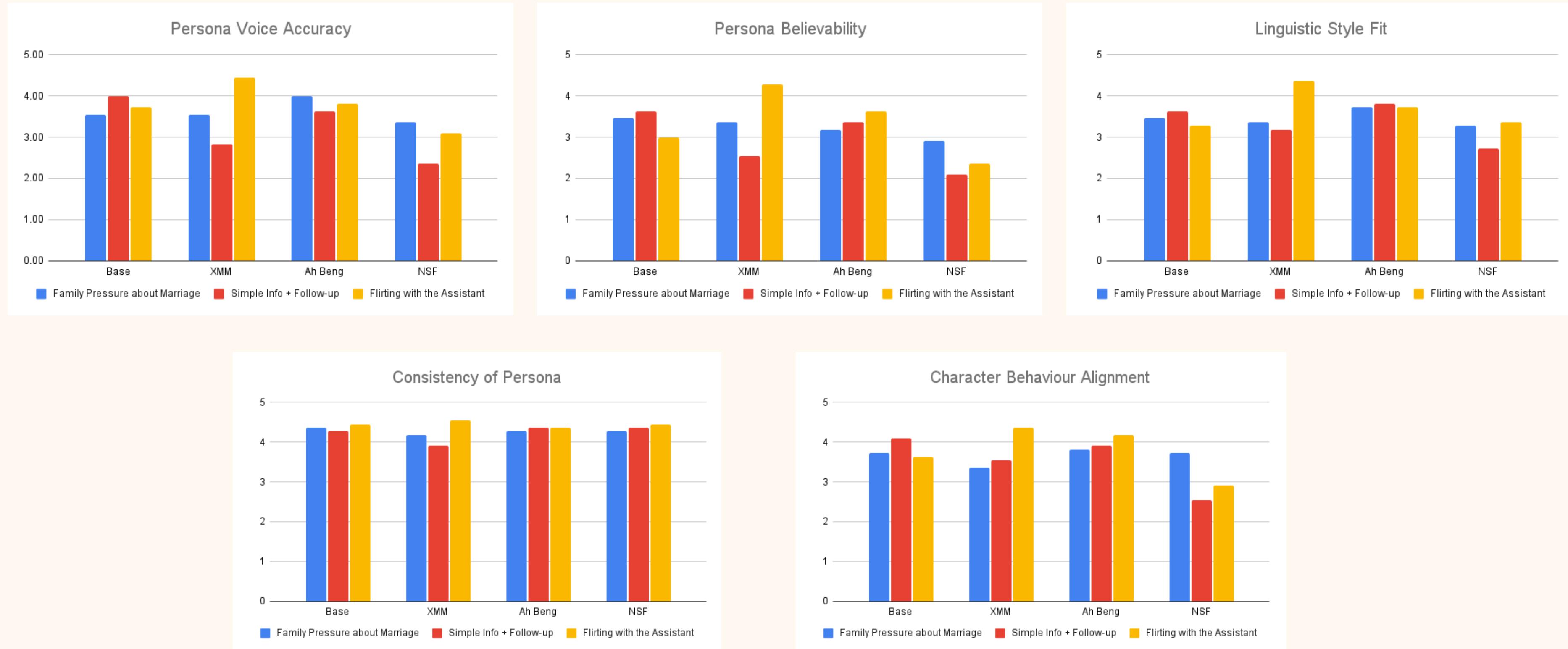
If I tell them I am not ready, they say I am too picky. How should I respond?

02:51 pm

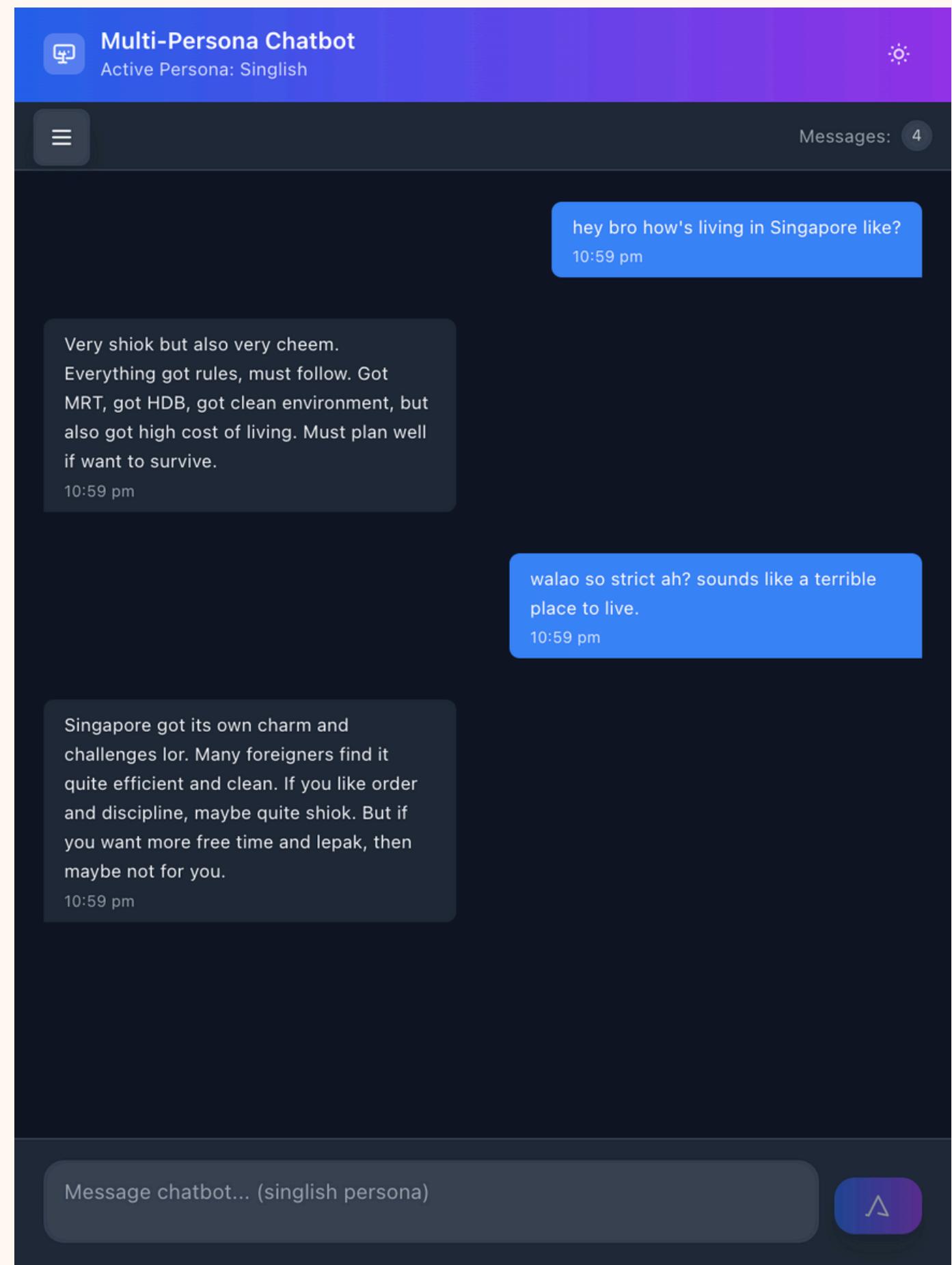
Survey respondents were shown **fixed conversation screenshots** to ensure **equal evaluation** of each persona.

Snippet of Conversation

Survey Results



DEMO



Thank you