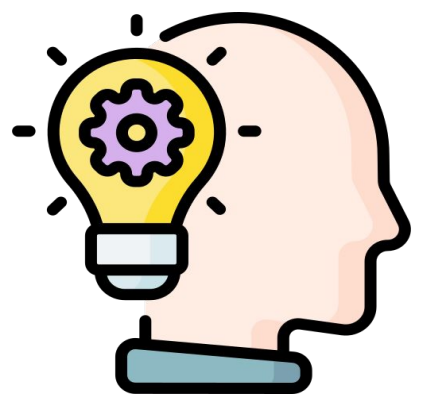


Measuring User's Mental Models of Speech Translation in Human-AI Collaboration



HyoJung Han

Nishant Balepur

Jordan Boyd-Graber

Marine Carpuat

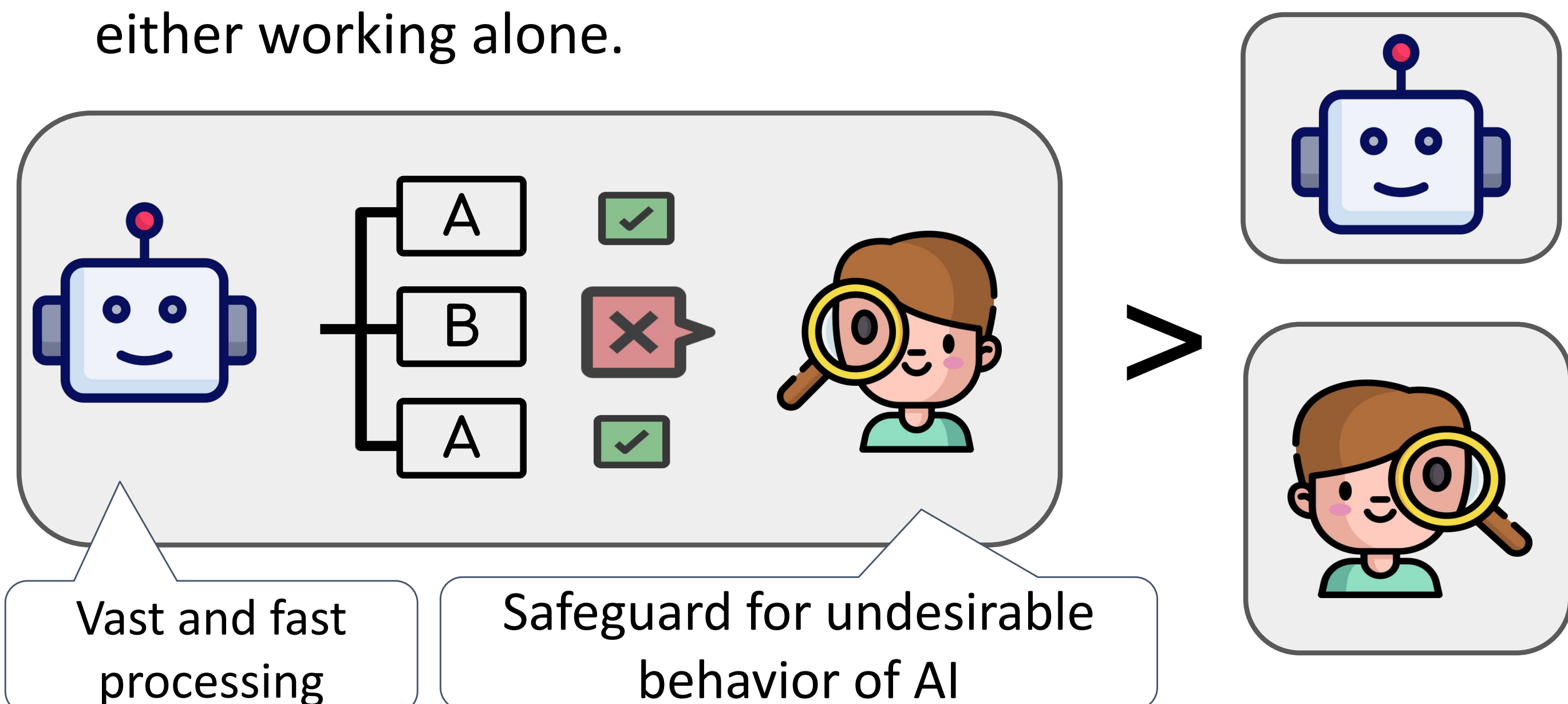
University of Maryland, College Park, USA



Optimal Human-AI Collaboration...

Motivation

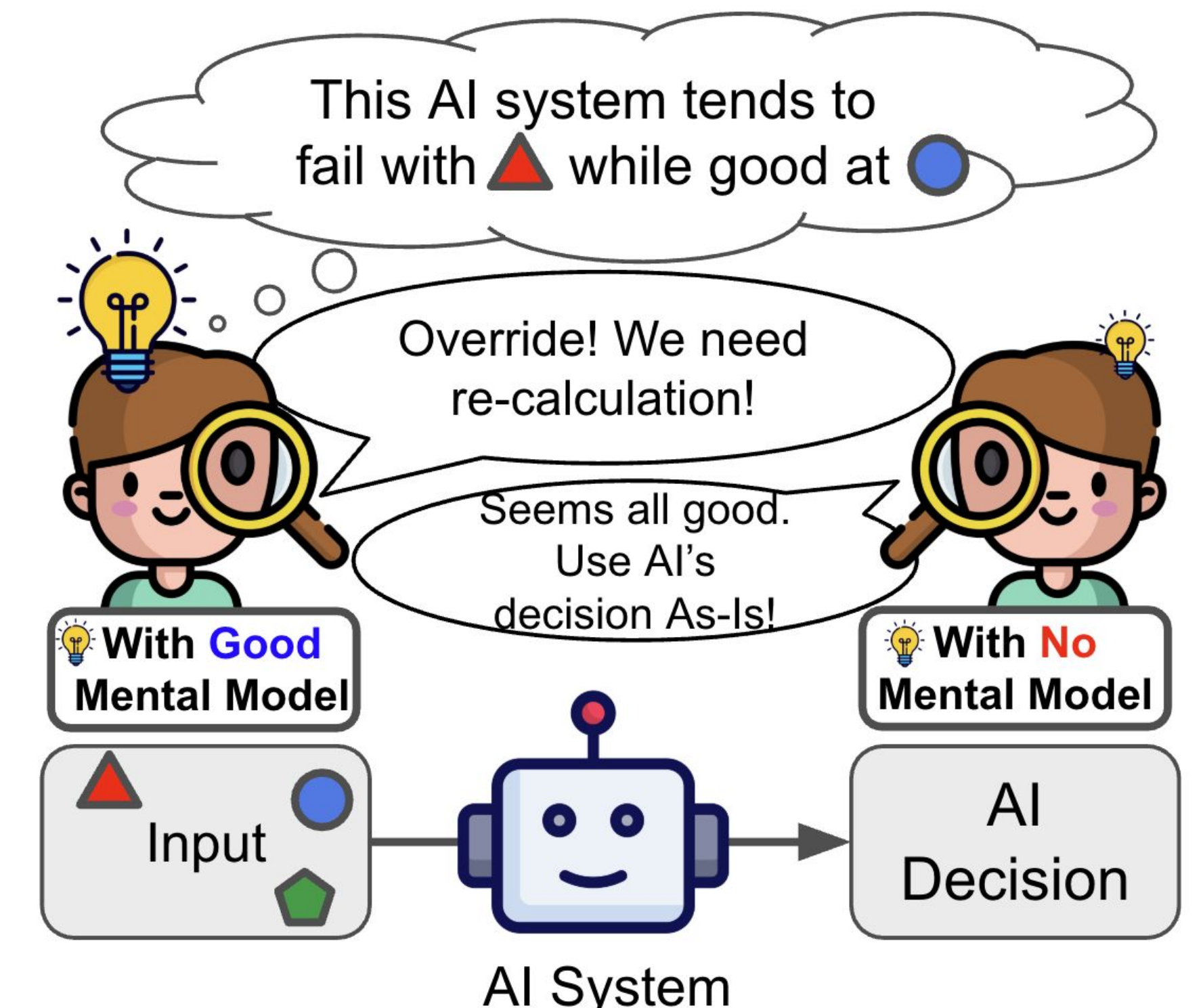
- Human-AI's complementary relationship often outperforms either working alone.



...Requires a Good Mental Model!

Mental Model: Users' understanding of an AI's capabilities and limitations

- Developing **accurate mental models** of AI systems is crucial for effectively integrating AI into **decision-making** [1]



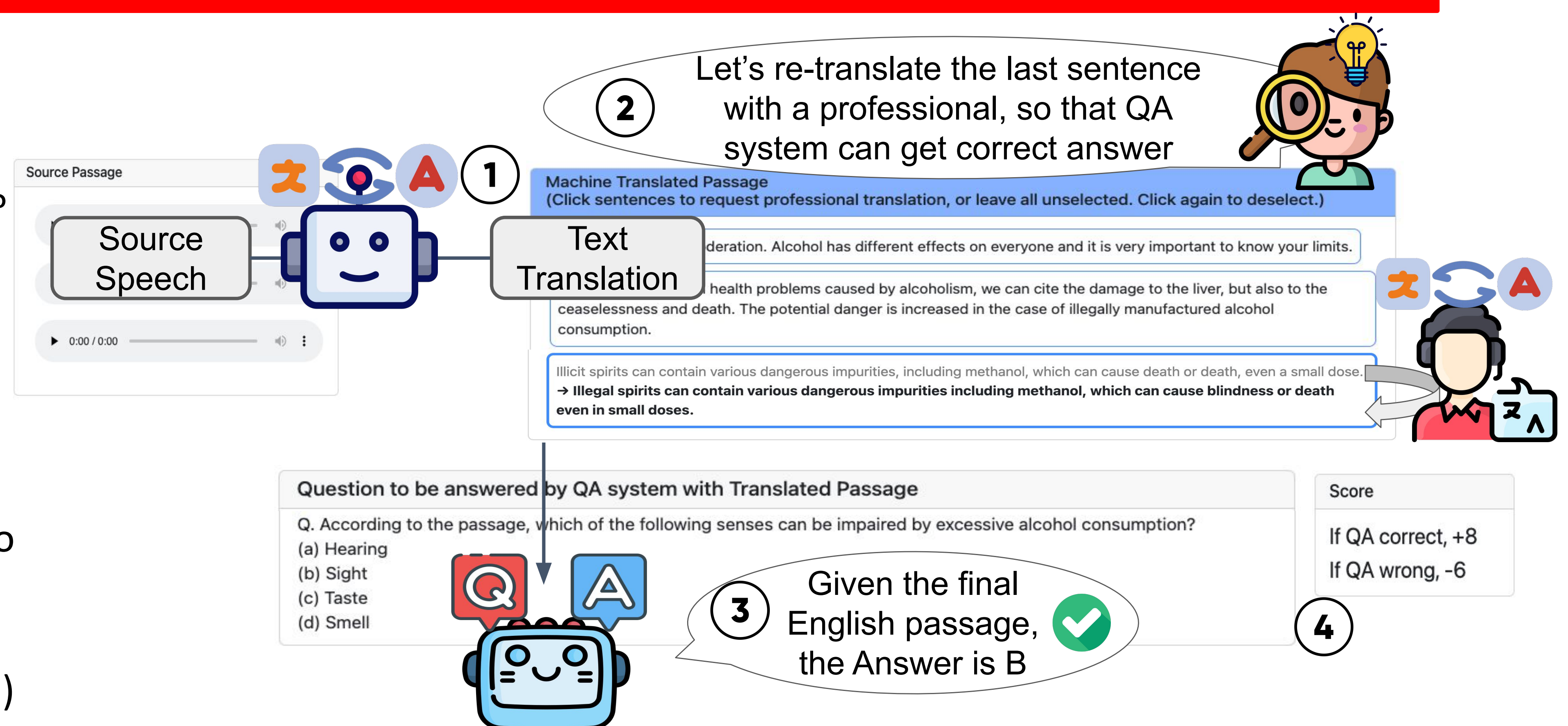
Goal: to understand and improve user's Mental Model of AI System (Speech Translation)

Research Question

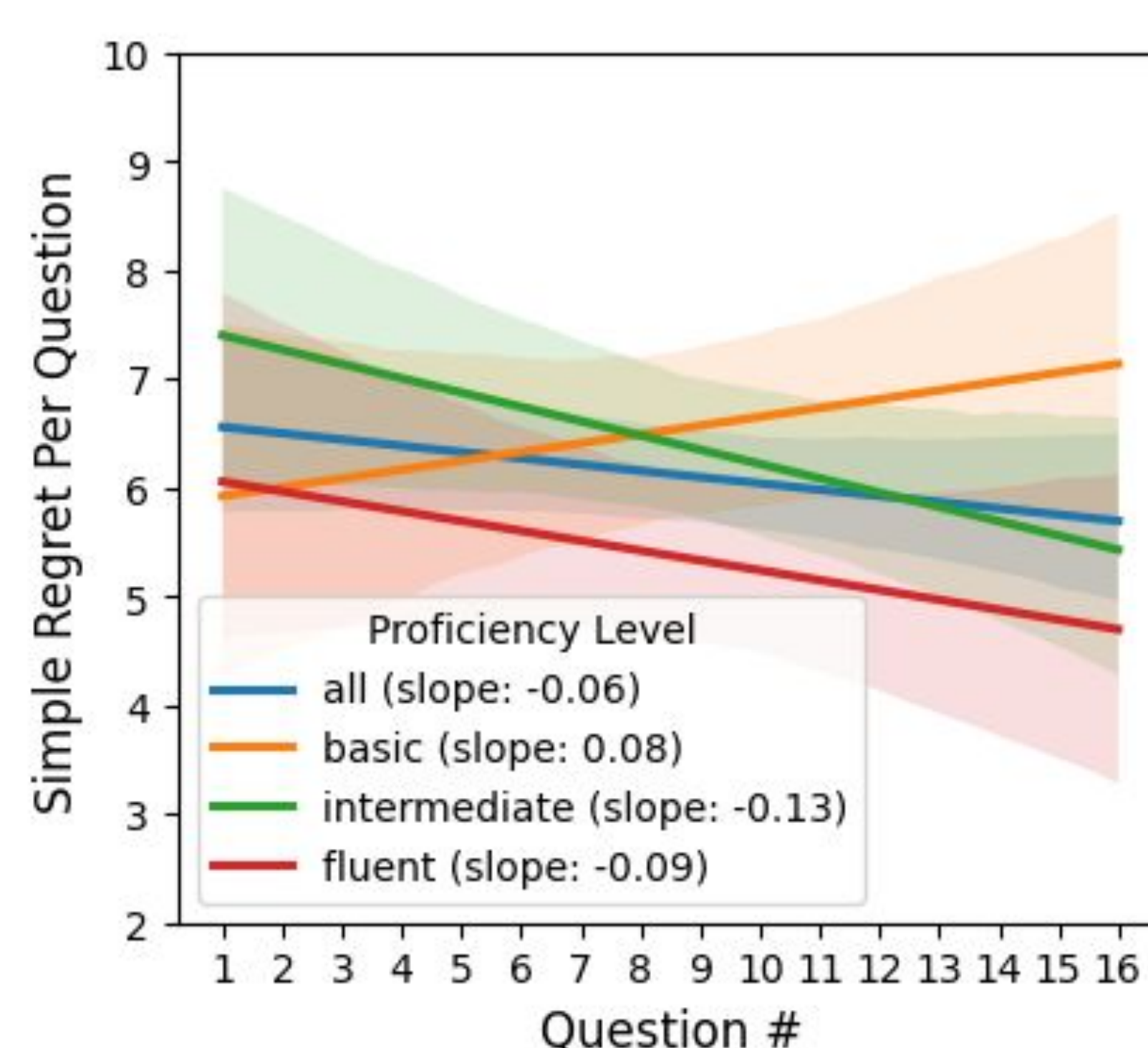
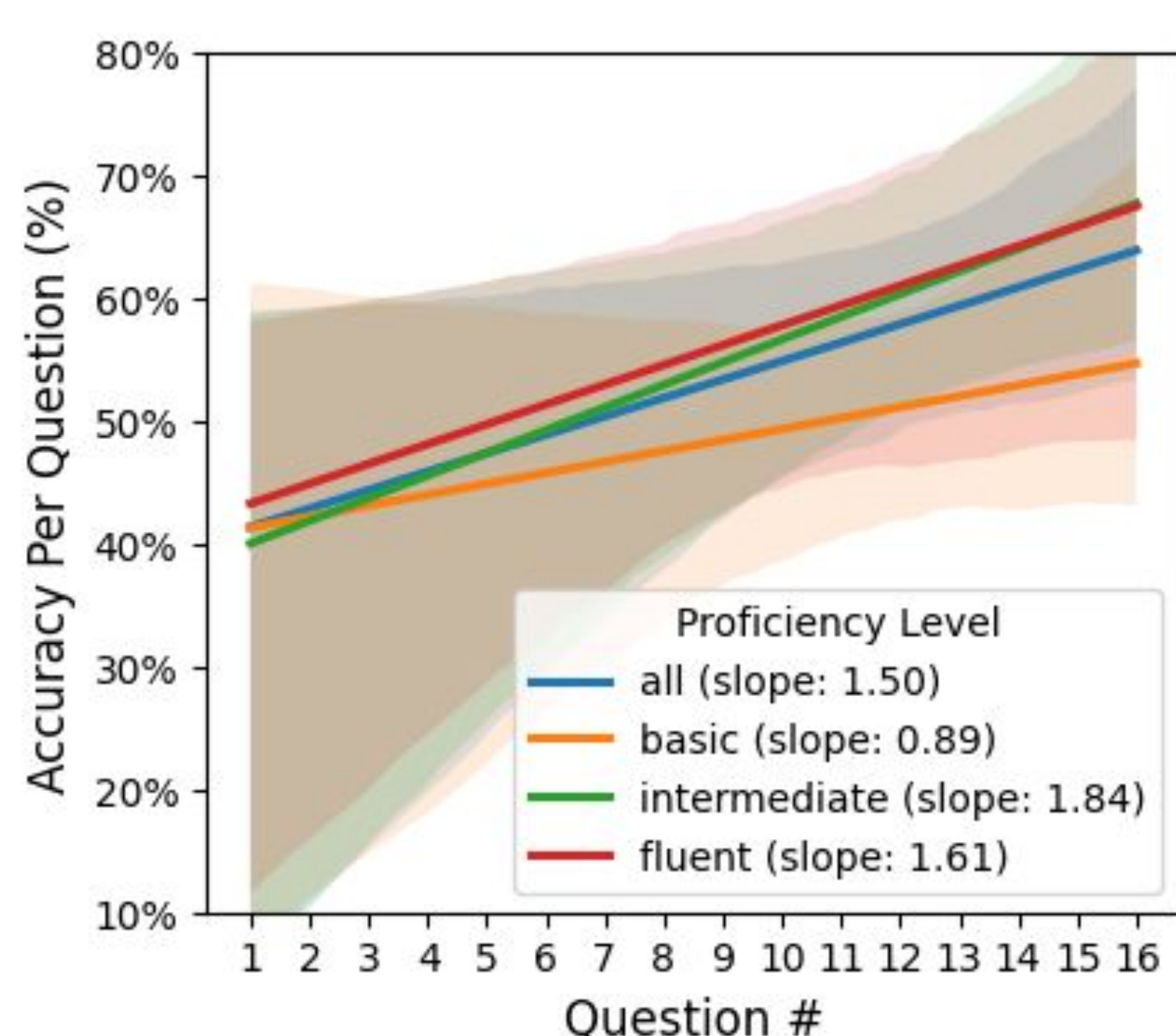
- How users **make sense of** what kinds of things that the **AI system is good/bad at?**
 - How this understanding varies with their language fluency?
- How can we **assist** users in building and updating their mental model of an MT?

Question Answering Framework

- MT system translate source passage input.
- Among each segment, user decide which one to re-translate by professional translator, and which one to use as-is. → forming the final passage
- QA system do the QA consuming the passage.
- Get rewarded by the QA correctness (- # of re-trans req)



Participants Gradually Update their Mental Models



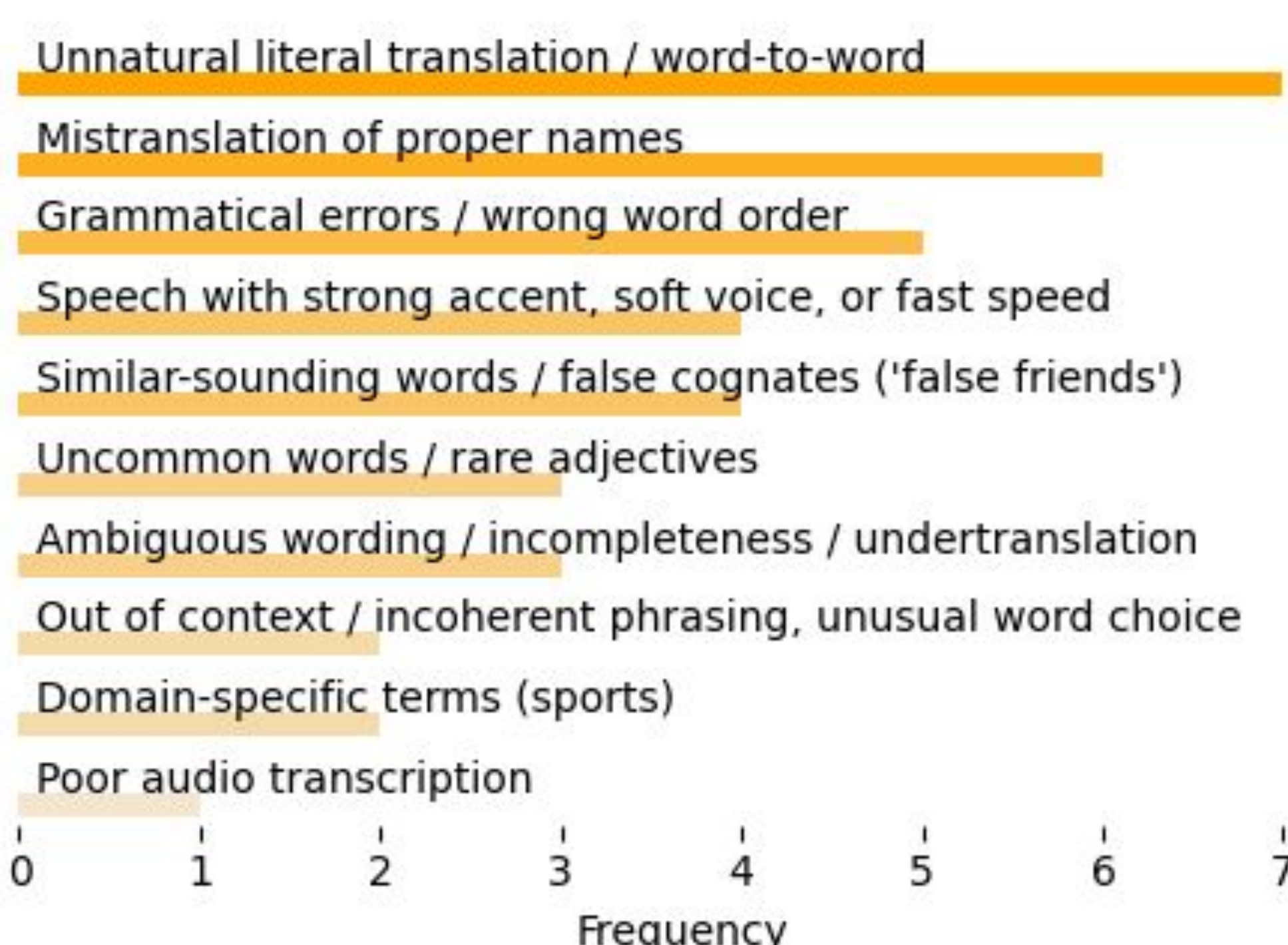
Useful Explanation for Better Mental Model

Final Score				
Condition (↓)	Basic	Intermediate	Fluent	All
Default	39.60	30.00	71.00	46.87
Transcription	34.00	56.67	54.20	49.80
Error Span	41.75	34.14	43.00	38.53
All	38.54	40.50	57.00	45.07

Overall Accuracy (%)				
Condition (↓)	Basic	Intermediate	Fluent	All
Default	63.75	60.00	70.00	64.58
Transcription	60.94	71.02	70.00	67.99
Error Span	68.75	71.43	68.75	70.00
All	64.42	68.12	69.64	67.52

User's Reflection

- Participants were asked to reflect on the types of errors the MT system is likely to make and the characteristics they learned to look for when identifying incorrect translations.



- [Beyond Accuracy: The Role of Mental Models in Human-AI Team Performance](#) (Bansal et al., AAAI HCOMP-19)
- [2M-BELEBELE: Highly Multilingual Speech and American Sign Language Comprehension Dataset](#) (Costa-jussà et al., 2024)
- [Robust Speech Recognition via Large-Scale Weak Supervision](#). Radford et al, 2022
- [xcomet: Transparent Machine Translation Evaluation through Fine-grained Error Detection](#) (Guerreiro et al., TACL 2024)
- [Looking for a Needle in a Haystack: A Comprehensive Study of Hallucinations in Neural Machine Translation](#) (Guerreiro et al., EACL 2023)