



The Learning Ideas Conference

TIM SCHLIPPE & JÖRG SAWATZKI

AI-BASED MULTILINGUAL

INTERACTIVE EXAM PREPARATION

New York, USA

June 16, 2021

AGENDA

Motivation

1

Related Work

2

AI-based Interactive Exam Preparation

3

Experiments and Results

4

Conclusion and Future Work

5

MOTIVATION

MOTIVATION



MOTIVATION



Sources: United Nations: Sustainable Development Goals: 17 Goals to Transform our World (2021); OpenClipart-Vectors/154119/Pixabay; Statista: The Most Spoken Languages Worldwide in 2019 (2020).

MOTIVATION

A screenshot of a WhatsApp chat conversation. The background is a green gradient with a faint image of a person on a bicycle. The chat contains the following messages:

- Message 1:** A question mark icon, "Wat is het belangrijkste voordeel van functieargumenten die via verwijzing worden doorgegeven?" (22:07)
- Message 2:** "Het voordeel is dat variabelen die als argument doorgegeven worden direkt aangepast worden." (22:08, with two green checkmarks)
- Message 3:** "Bedankt voor je antwoord!"
- Message 4:** A question mark icon, "De vraag: Wat is het belangrijkste voordeel van functieargumenten die via verwijzing worden doorgegeven?"
- Message 5:** A pencil icon, "Uw antwoord: Het voordeel is dat variabelen die als argument doorgegeven worden direkt aangepast worden."
- Message 6:** A book icon, "Model antwoord: Het vermijdt het maken van kopieën van grote datastructuren bij het aanroepen van functies."
- Message 7:** A graduation cap icon, "Je cijfer: 3 van 5 punten" followed by three yellow stars.
- Message 8:** "Geweldig, dat geeft ons brandstof voor 6 miljoen extra kilometers!"
- Message 9:** "Slechts 44 miljoen km verwijderd van Mars!" followed by a globe icon, a rocket icon, and a planet icon.

The time "22:08" is shown at the bottom right of the final message.

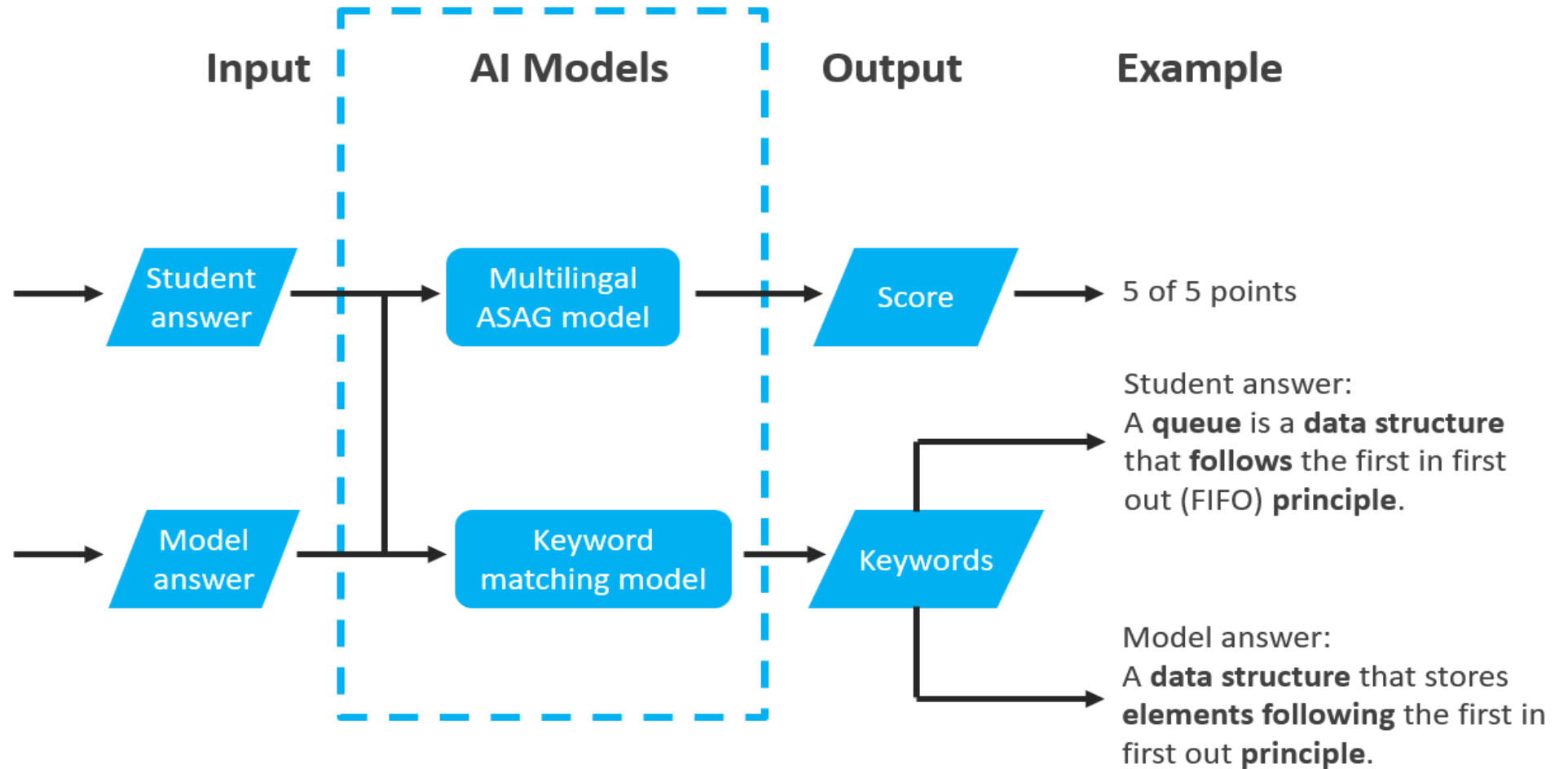
MOTIVATION

Example

What is a queue?

A queue is a data structure that follows the first in first out (FIFO) principle.

A data structure that stores elements following the first in first out principle.



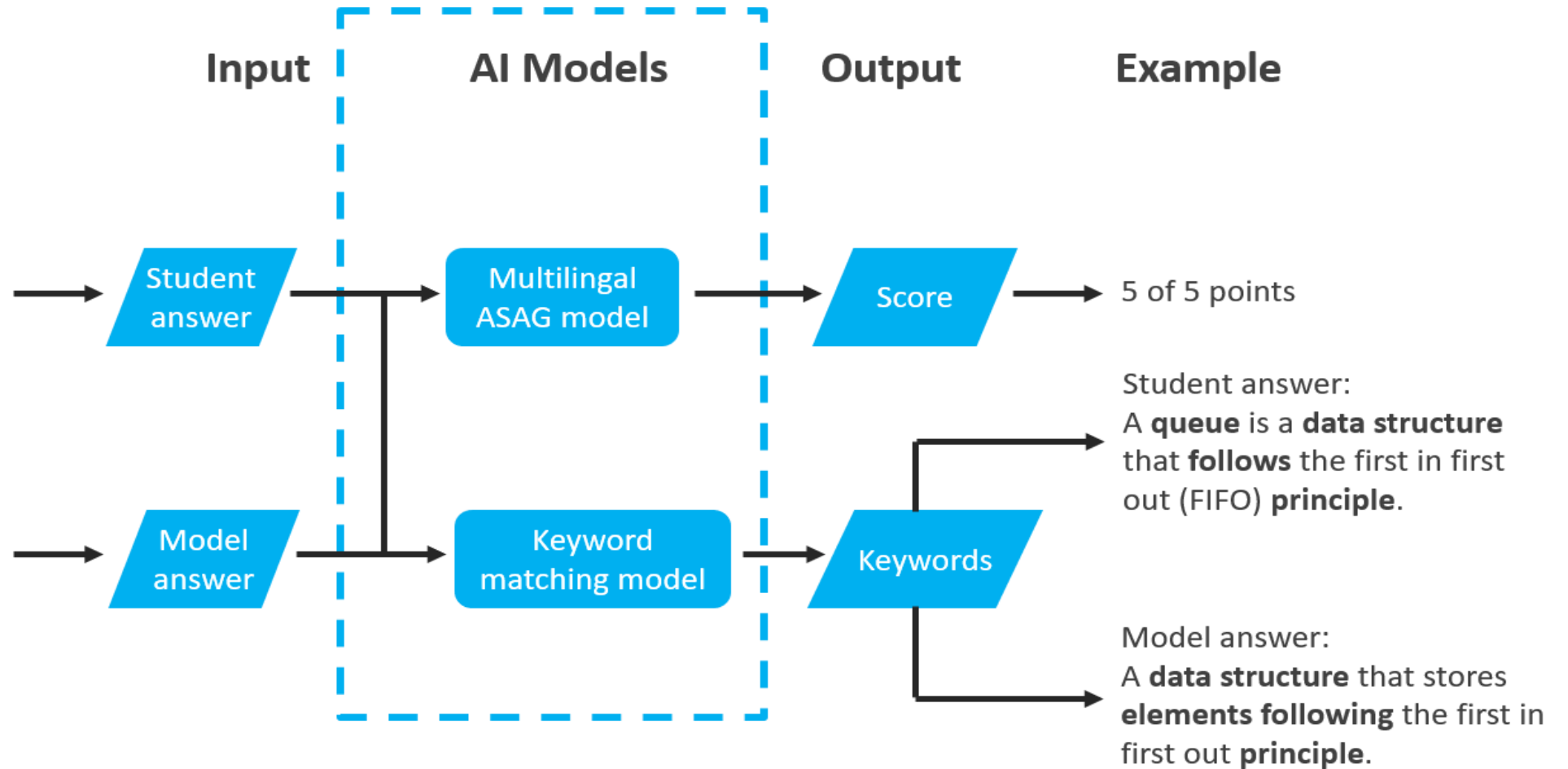
MOTIVATION

Example

What is a queue?

A queue is a data structure that follows the first in first out (FIFO) principle.

A data structure that stores elements following the first in first out principle.



MOTIVATION

Example

What is a queue?

A queue is a data structure that follows the first in first out (FIFO) principle.



A data structure that stores elements following the first in first out principle.



Input

AI Models

Output

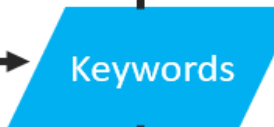
Example

Multilingual ASAG model



5 of 5 points

Keyword matching model



Student answer:
A **queue** is a **data structure** that **follows** the first in first out (FIFO) **principle**.

Model answer:
A **data structure** that stores **elements following** the first in first out **principle**.

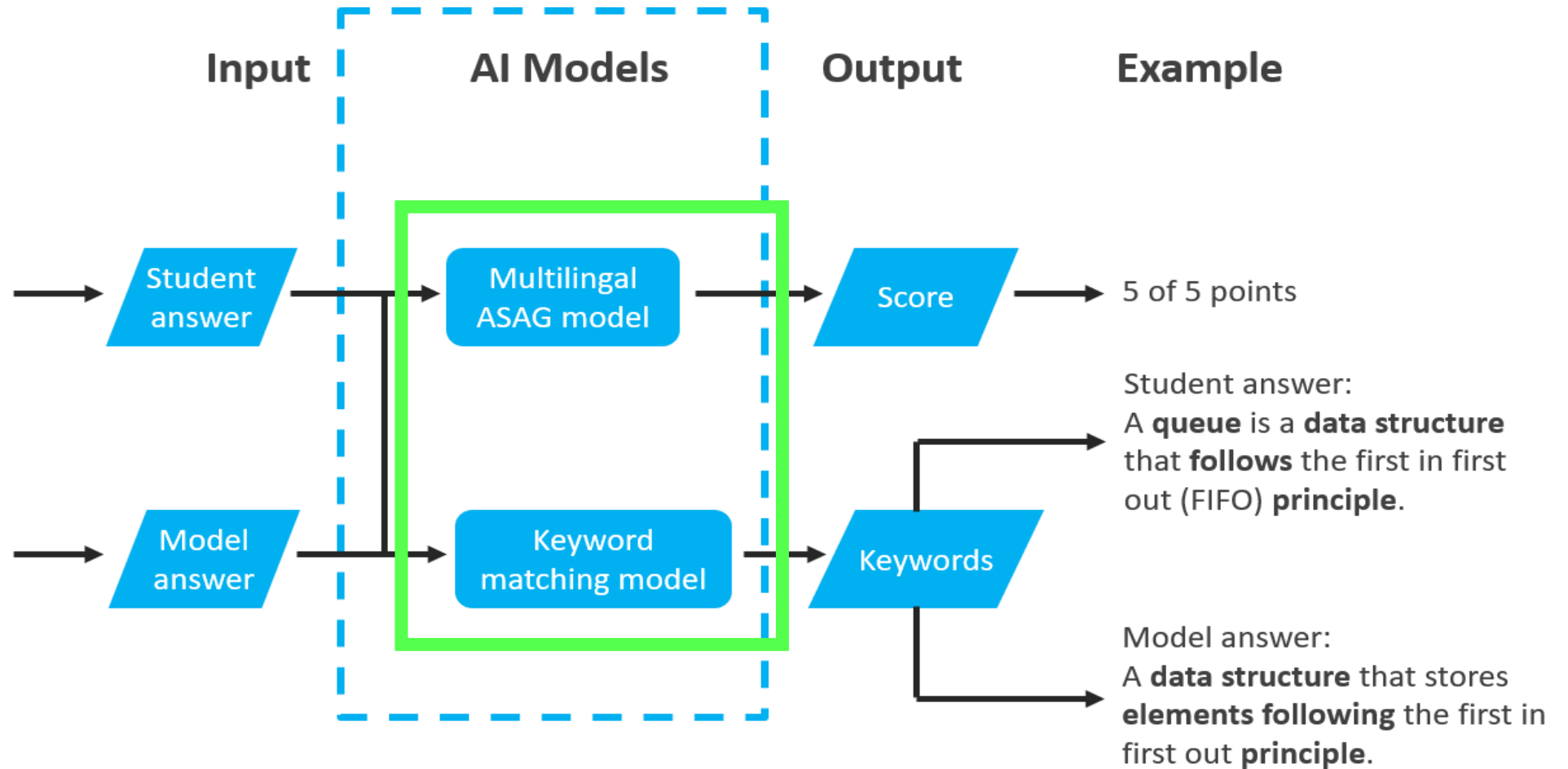
MOTIVATION

Example

What is a queue?

A queue is a data structure that follows the first in first out (FIFO) principle.

A data structure that stores elements following the first in first out principle.



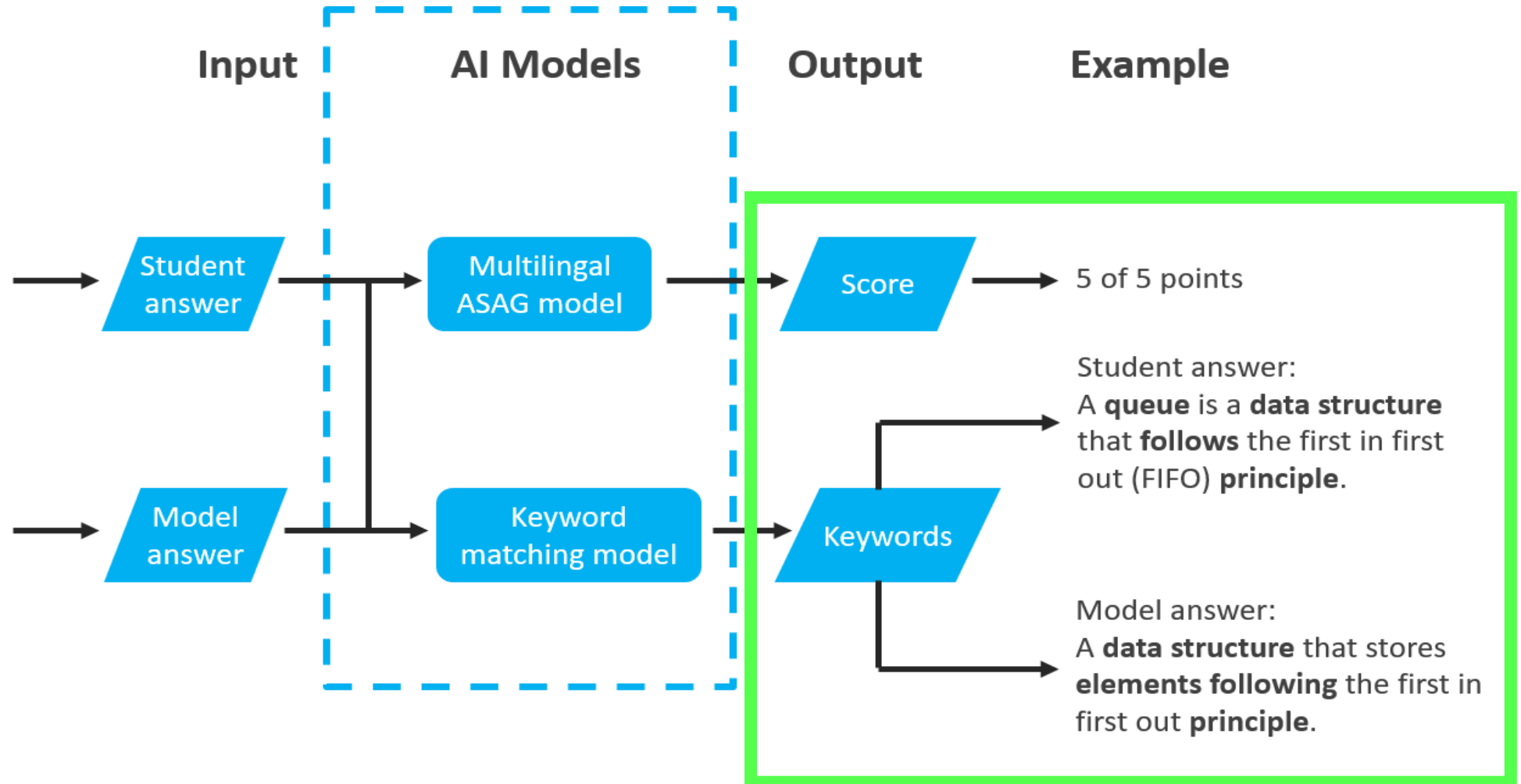
MOTIVATION

Example

What is a queue?

A queue is a data structure that follows the first in first out (FIFO) principle.

A data structure that stores elements following the first in first out principle.



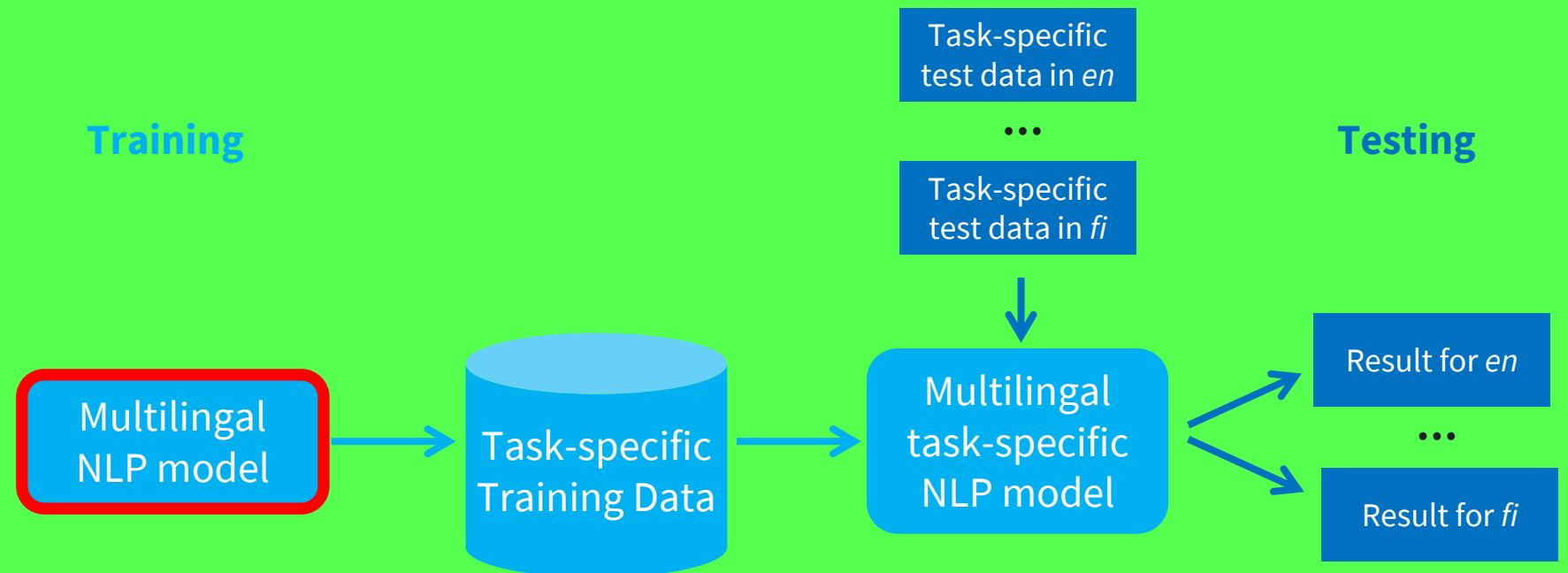
02

RELATED WORK

MULTILINGUAL NLP MODELS

Transfer learning
Cross-lingual transfer

e.g., Multilingual BERT
(Devlin et al., 2019;
Pires et al., 2019),
RoBERTa (Liu et al., 2019),
XLM-R (Conneau, 2018)



RELATED WORK

MULTILINGUAL NLP MODELS

Transfer learning
Cross-lingual transfer

e.g., Multilingual BERT

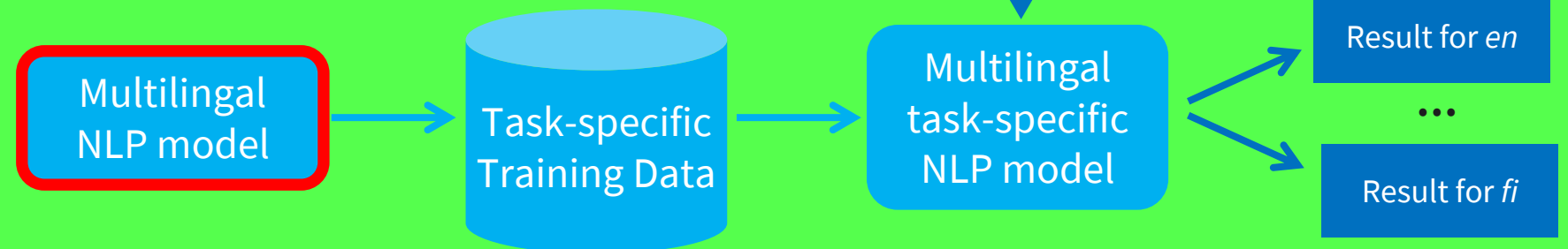
Multilingual BERT

(Liu et al., 2019),

RoBERTa (Liu et al., 2019),

XLM-R (Conneau, 2018)

Training

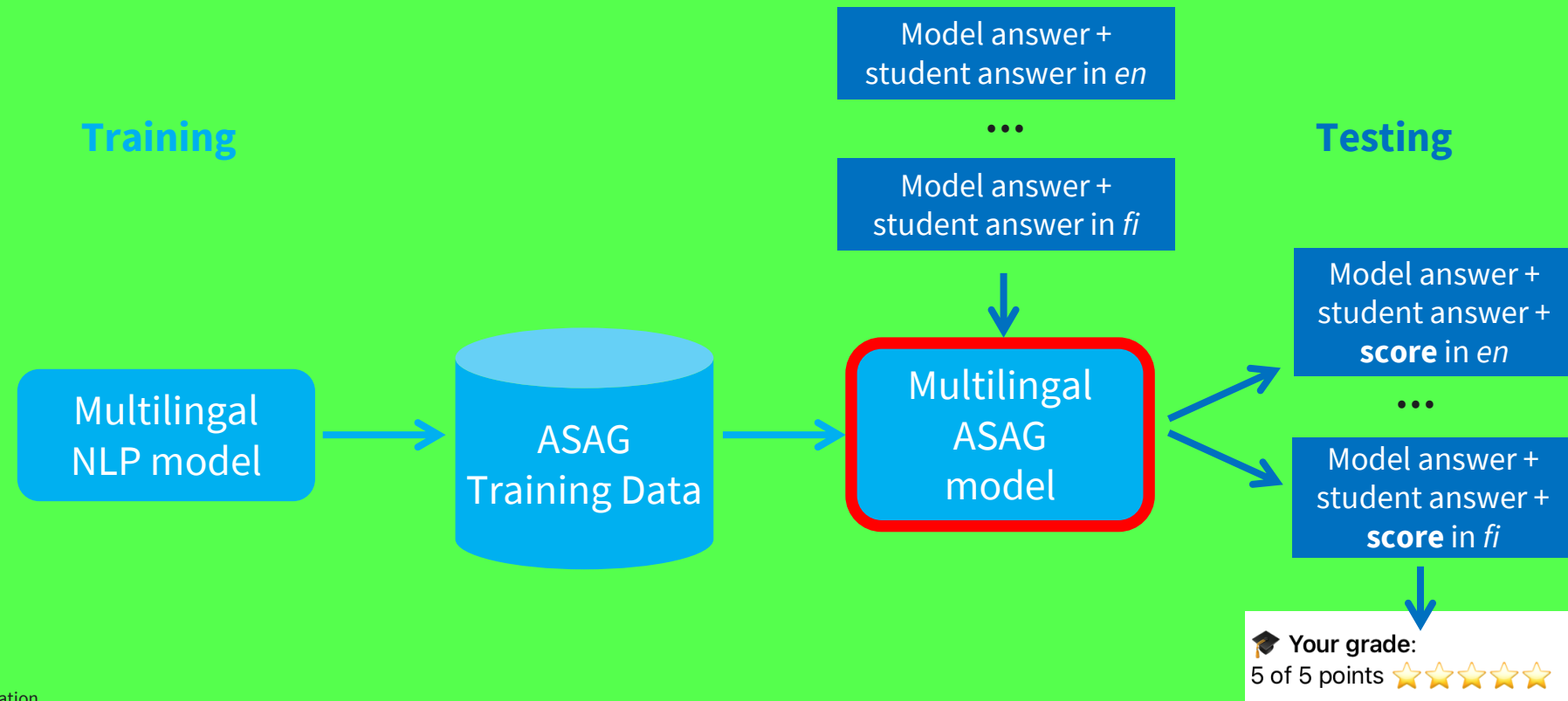


RELATED WORK

AUTOMATIC SHORT ANSWER GRADING

Deep learning

e.g., (Burrows et al., 2014;
Camus & Filighera, 2020;
Sawatzki et al., 2021;
Schlippe & Sawatzki, 2021b)

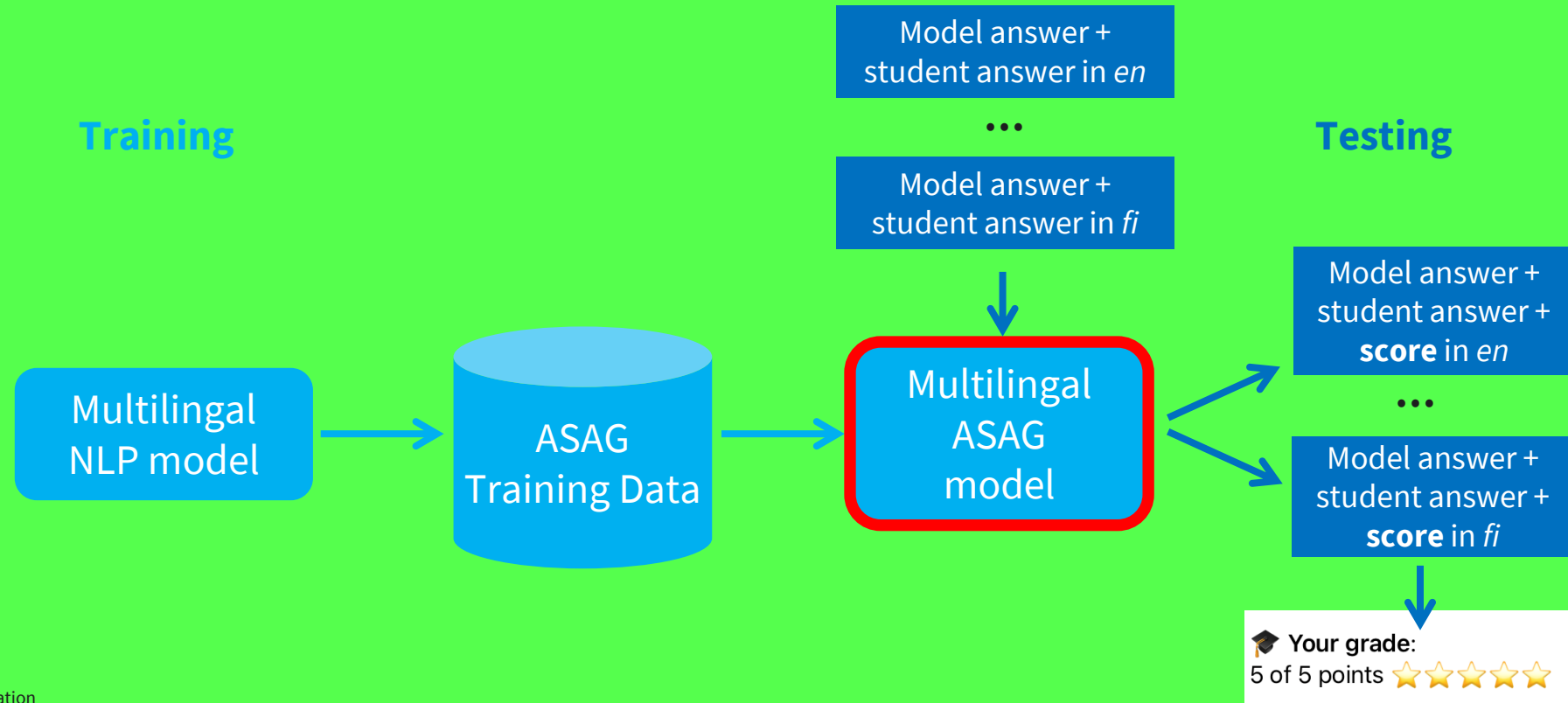


RELATED WORK

**AUTOMATIC
SHORT ANSWER
GRADING**

Deep learning

**Cross-lingual
Automatic Short
Answer Grading**



CONVERSATIONAL

AI

Pedagogical Conversational Agents

e.g. (Wölfel, 2021)

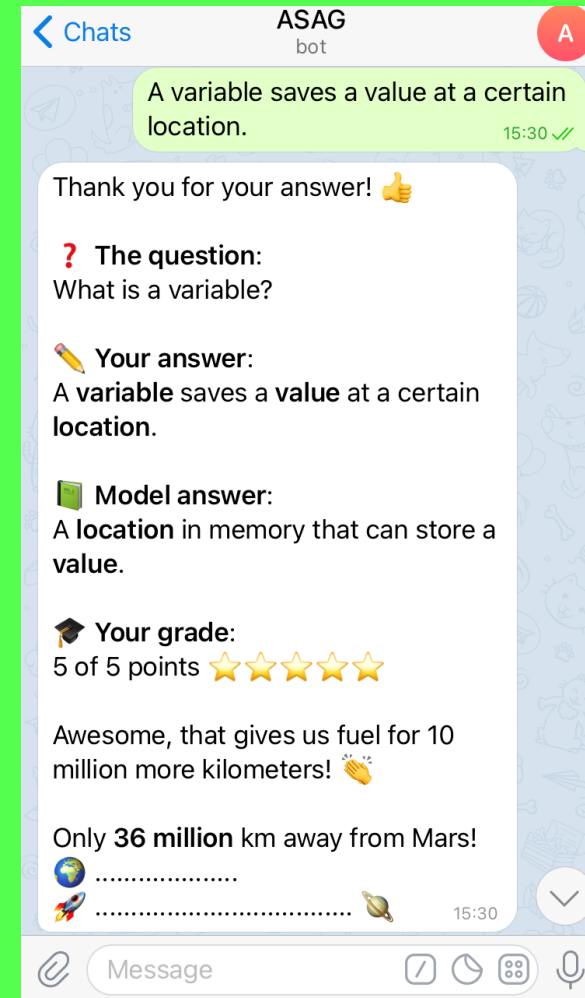
Frameworks

RASA (Bocklisch et al., 2019),

Dialogflow (Reyes et al., 2019),

Telegram (Setiaji & Paputungan, 2017),

Graphic Source: Custom Depiction.



CONVERSATIONAL

AI

Pedagogical Conversational Agents

e.g. (Wölfel, 2021)

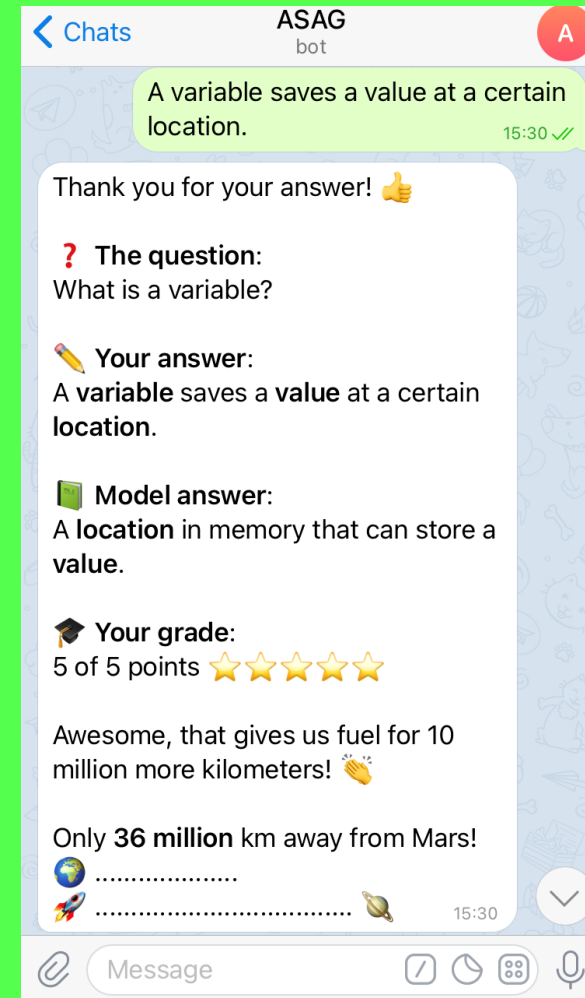
Frameworks

RASA (Bocklisch et al., 2019),

Dialog **Telegram** (, 2019),

Telegram (Setiaji & Paputungan, 2017),

Graphic Source: Custom Depiction.



RELATED WORK

KEYWORD EXTRACTION


SEMANTIC SIMILARITY


Keyword extraction



e.g., (Hasan, 2014;
Merrouni et al., 2020)

Semantic similarity

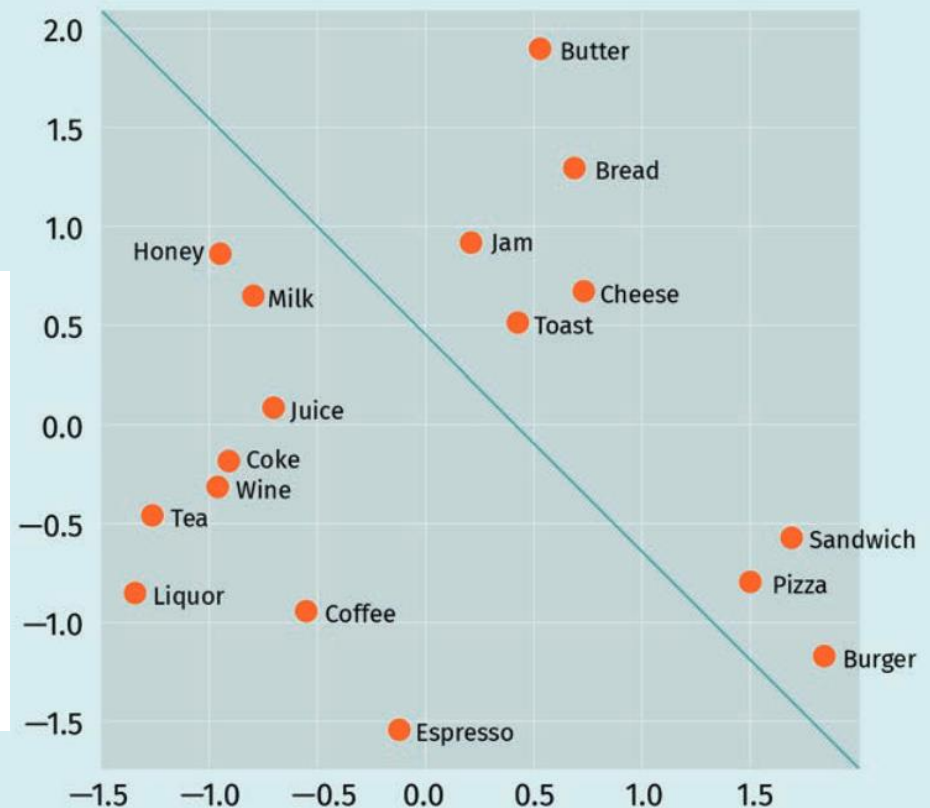
e.g., (Chandrasekaran, 2021)

 **Your answer:**
A **variable** saves a **value** at a certain **location**.

 **Model answer:**
A **location** in memory that can store a **value**.

 **Your grade:**
5 of 5 points 

PCA Scatterplot Visualizing Word Vectors of Selected Food and Beverages





KEYWORD EXTRACTION



SEMANTIC SIMILARITY

Key
e.g.
Me
Se
e.g., (Chandrasekaran, 2021)

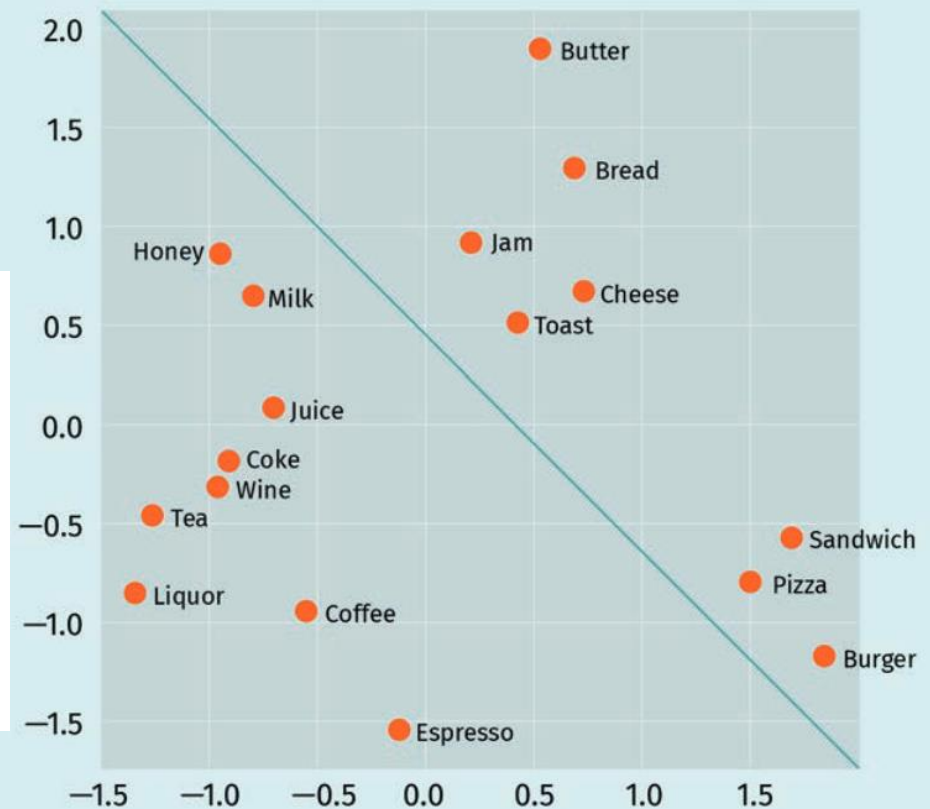
SpaCy
(Honnibal & Montani, n.d.)

 **Your answer:**
A **variable** saves a **value** at a certain **location**.

 **Model answer:**
A **location** in memory that can store a **value**.

 **Your grade:**
5 of 5 points 

PCA Scatterplot Visualizing Word Vectors of Selected Food and Beverages



TRANSLATION

MODELS

Google's Neural Machine Translation System

(Wu et al., 2016;
Aiken, 2019)

Thank you for your answer! 👍

? The question:

✎ Your answer:

📖 Model answer:

🎓 Your grade:
5 of 5 points ★★★★★

Awesome, that gives us fuel for 10 million more kilometers! 🙌

Only **36 million** km away from Mars!

🌍 🚀

15:30



Bedankt voor je antwoord!

? De vraag:

✎ Uw antwoord:

📖 Model antwoord:

🎓 Je cijfer:
3 van 5 punten ★★★

Geweldig, dat geeft ons brandstof voor

Slechts **44 miljoen** km verwijderd van

🌍 🚀

RELATED WORK

TRANSLATION

MODELS

Google's Neural Machine Translation System

(Wu et al., 2016;
Aiken, 2019)

translate.google.com

Thank you for your answer! 👍

? The question:

✎ Your answer:

📖 Model answer:

🎓 Your grade:
5 of 5 points ★★★★★

Awesome, that gives us fuel for 10 million more kilometers! 🙌

Only 36 million km away from Mars!

🌍 🚀

15:30



Translation
Model



Bedankt voor je antwoord!

? De vraag:

✎ Uw antwoord:

📖 Model antwoord:

🎓 Je cijfer:
3 van 5 punten ★★★

Geweldig, dat geeft ons brandstof voor

Slechts 44 miljoen km verwijderd van

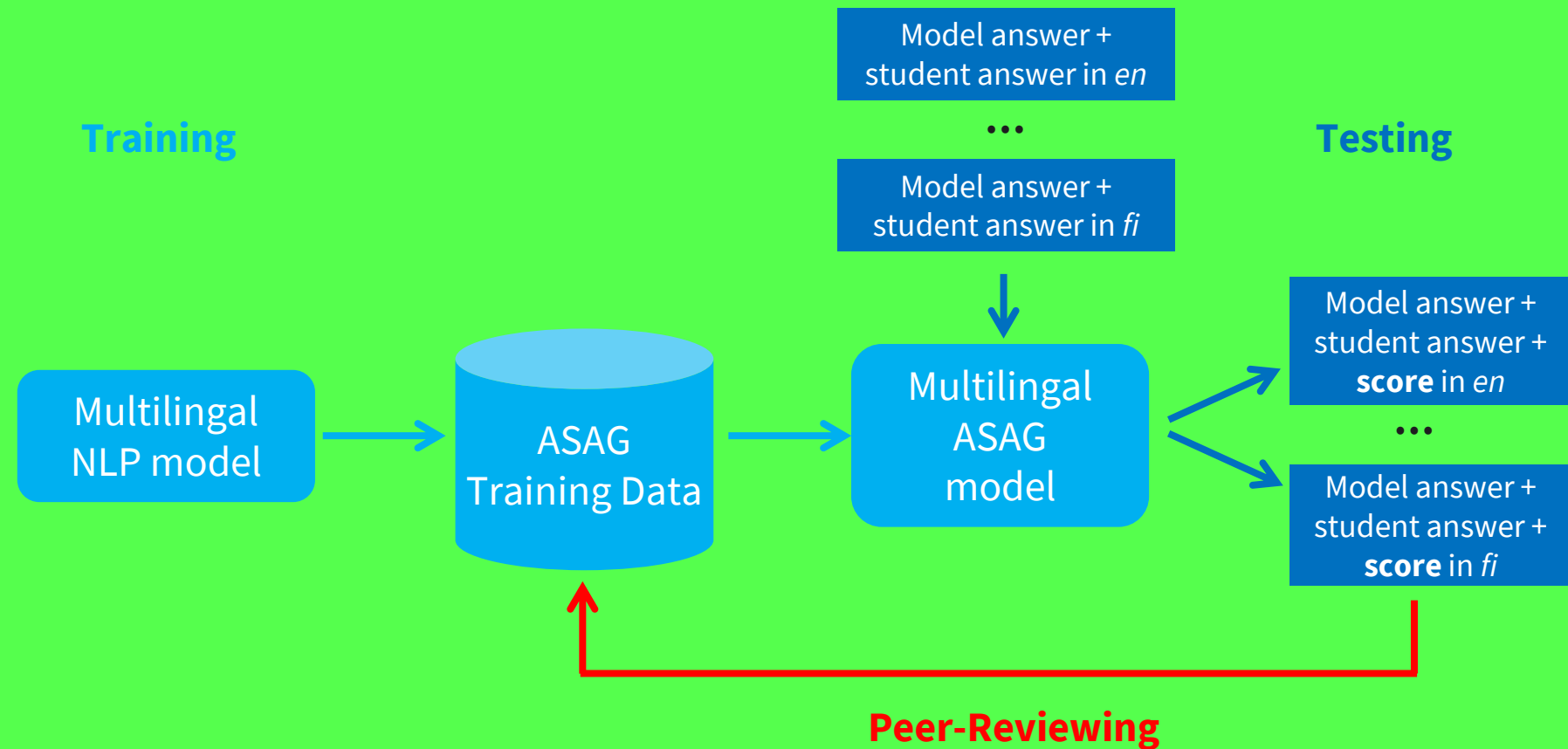
🌍 🚀

RELATED WORK

CROUDSOURCING
&
PEER-REVIEWING

Scoring another answer

e.g., Peer-based
proofreading
(Luo et al., 2014)



RELATED WORK

Wait! To use the new fuel, you need to activate the fuel pump! To do this, please **grade the following answer** of an English speaking fellow astronaut:

? The question:

What does the address (&) operator return?

The astronaut's answer:
memory address of its operand

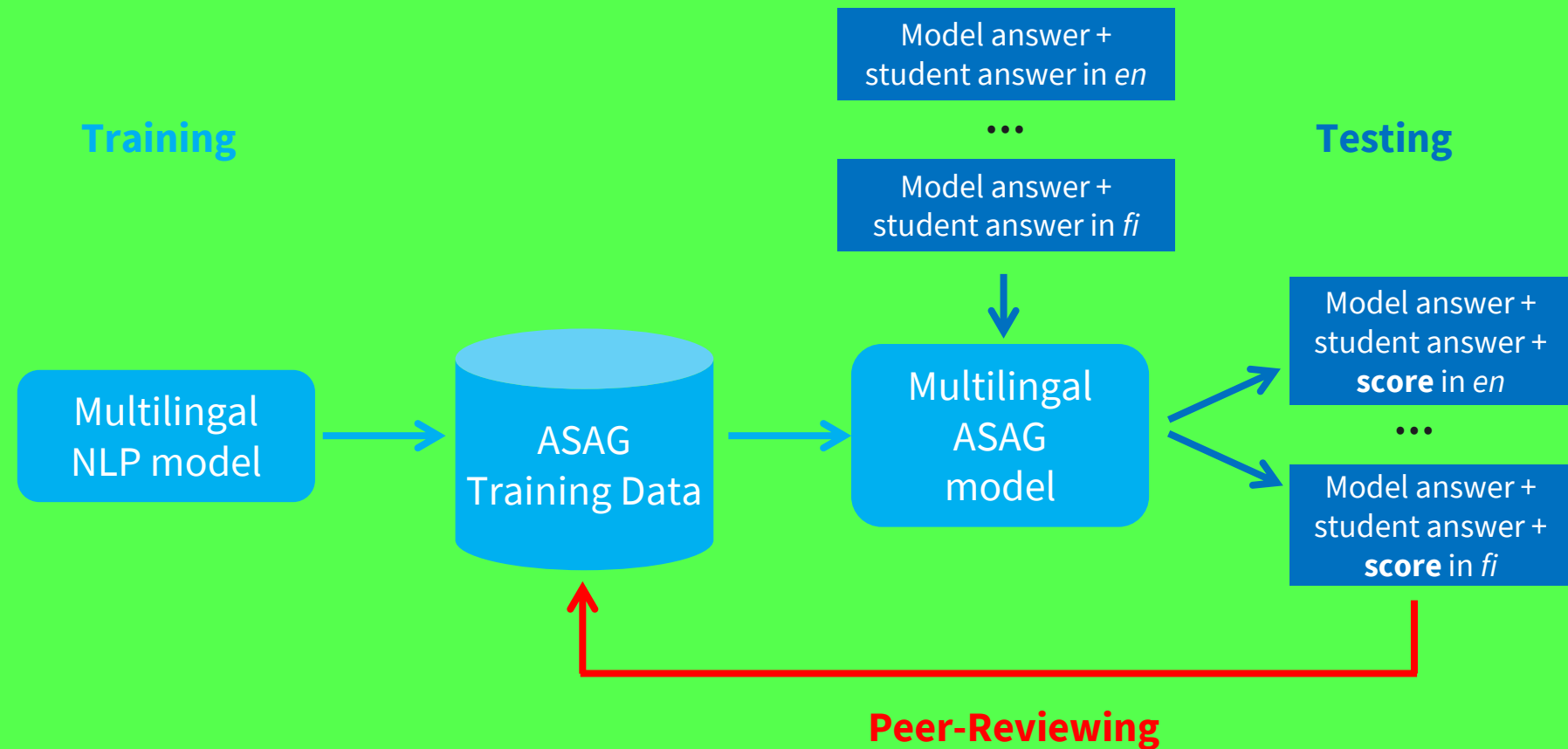
Model answer:
The memory address of its operand.

How many points would you give?
Please be fair and honest!

02:28

0 1 2 3 4 5

Training



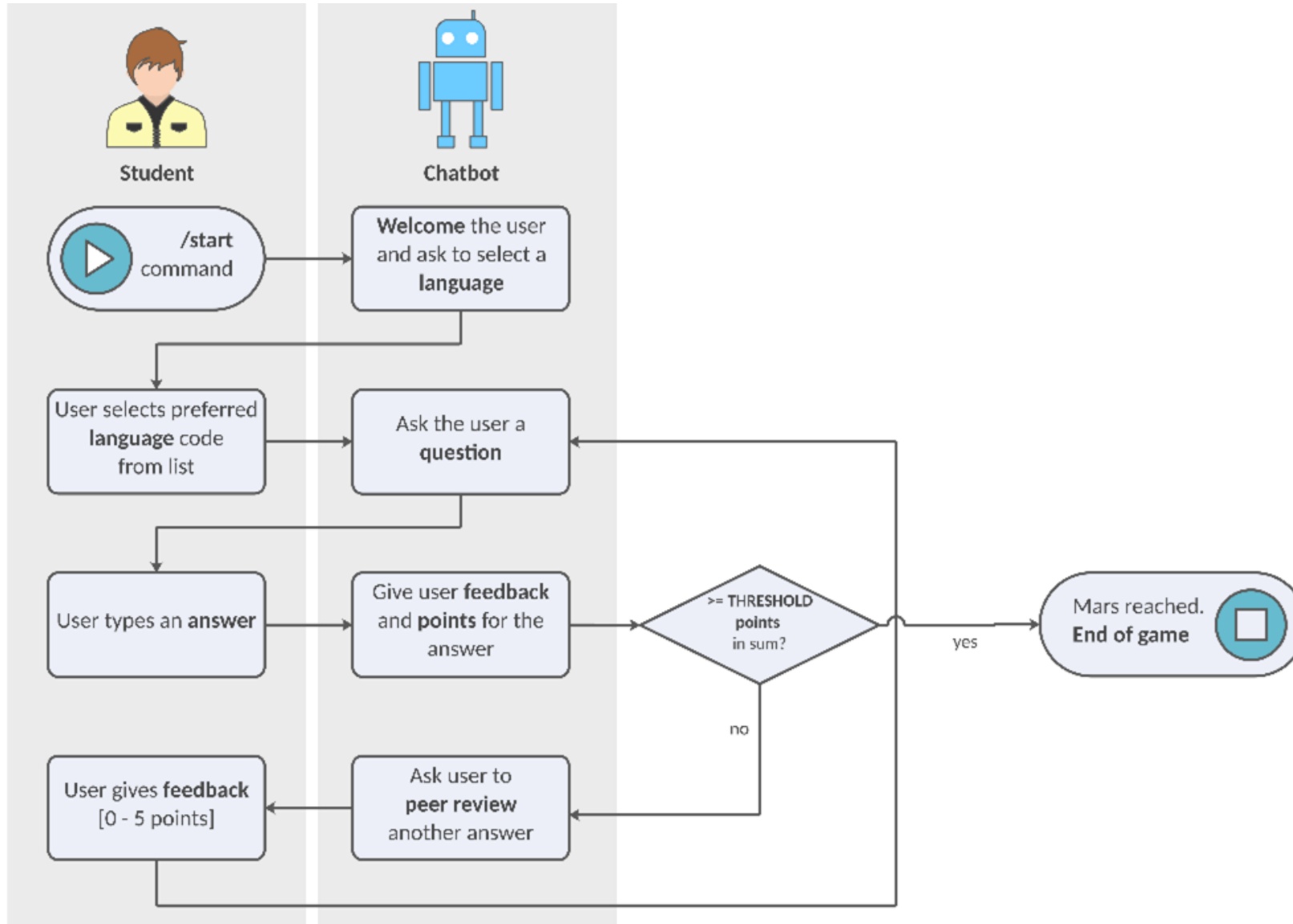
Testing

Peer-Reviewing

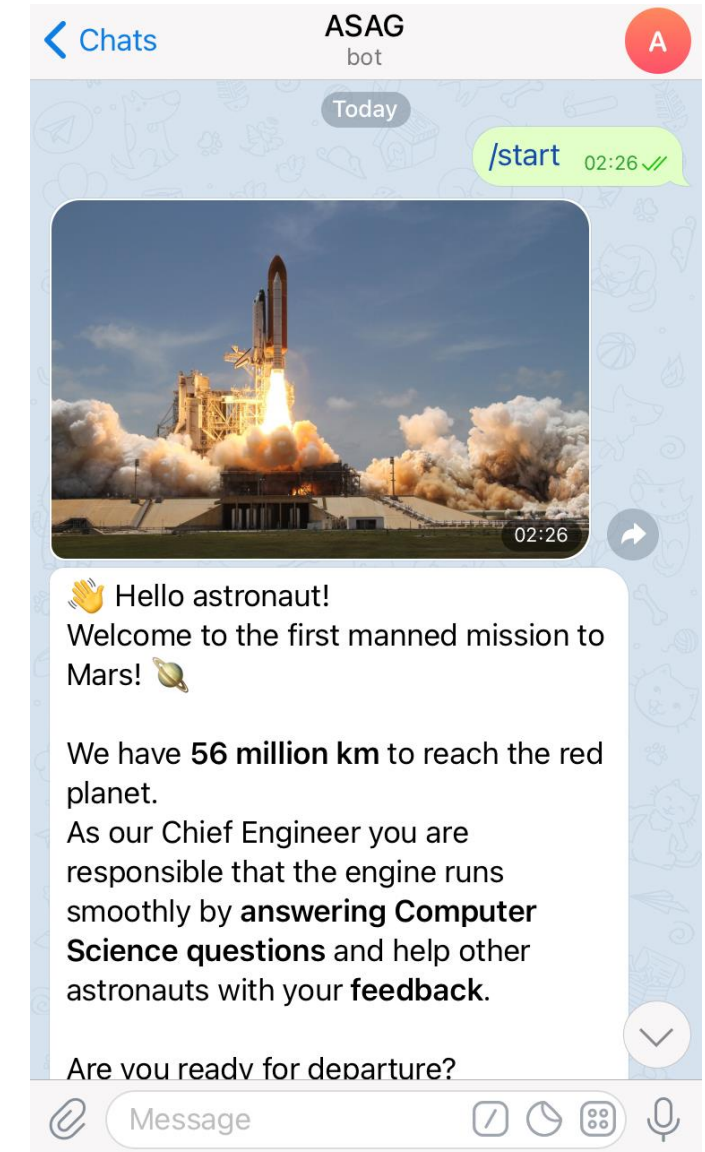
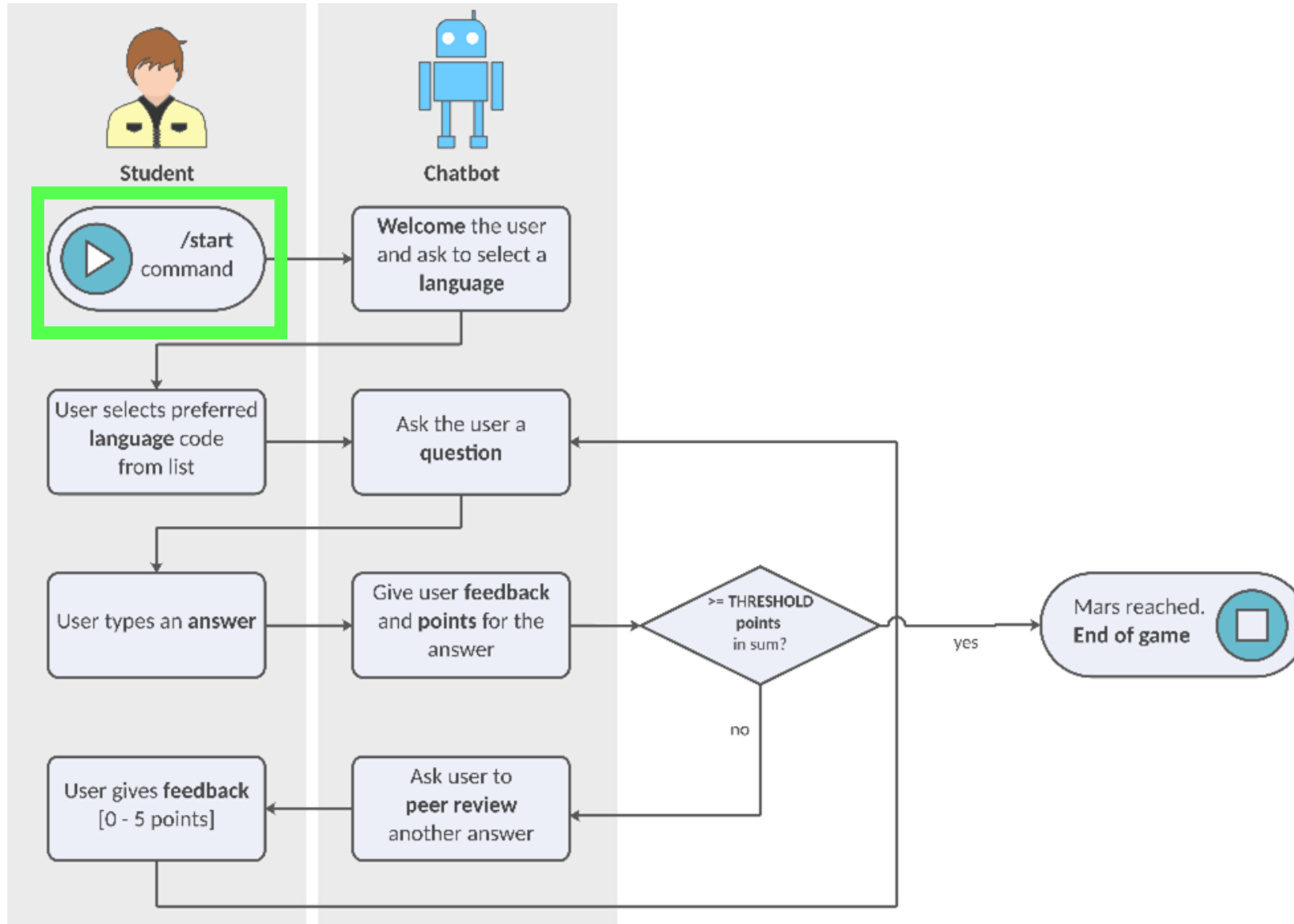
03

AI-BASED INTERACTIVE EXAM PREPARATION

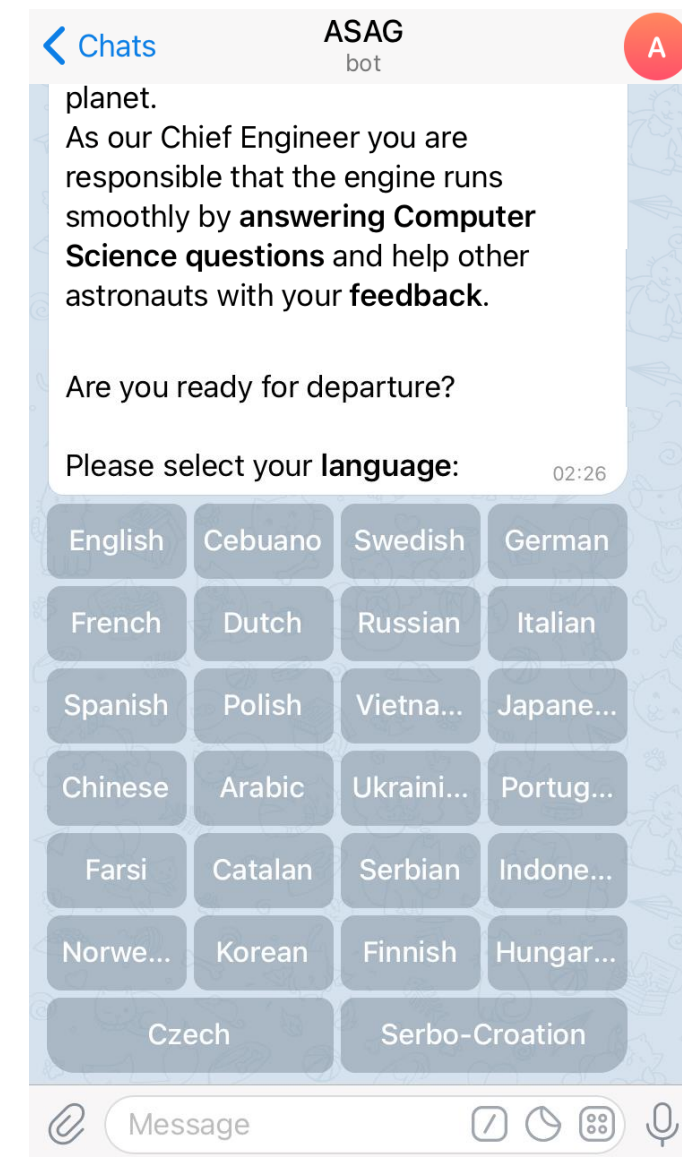
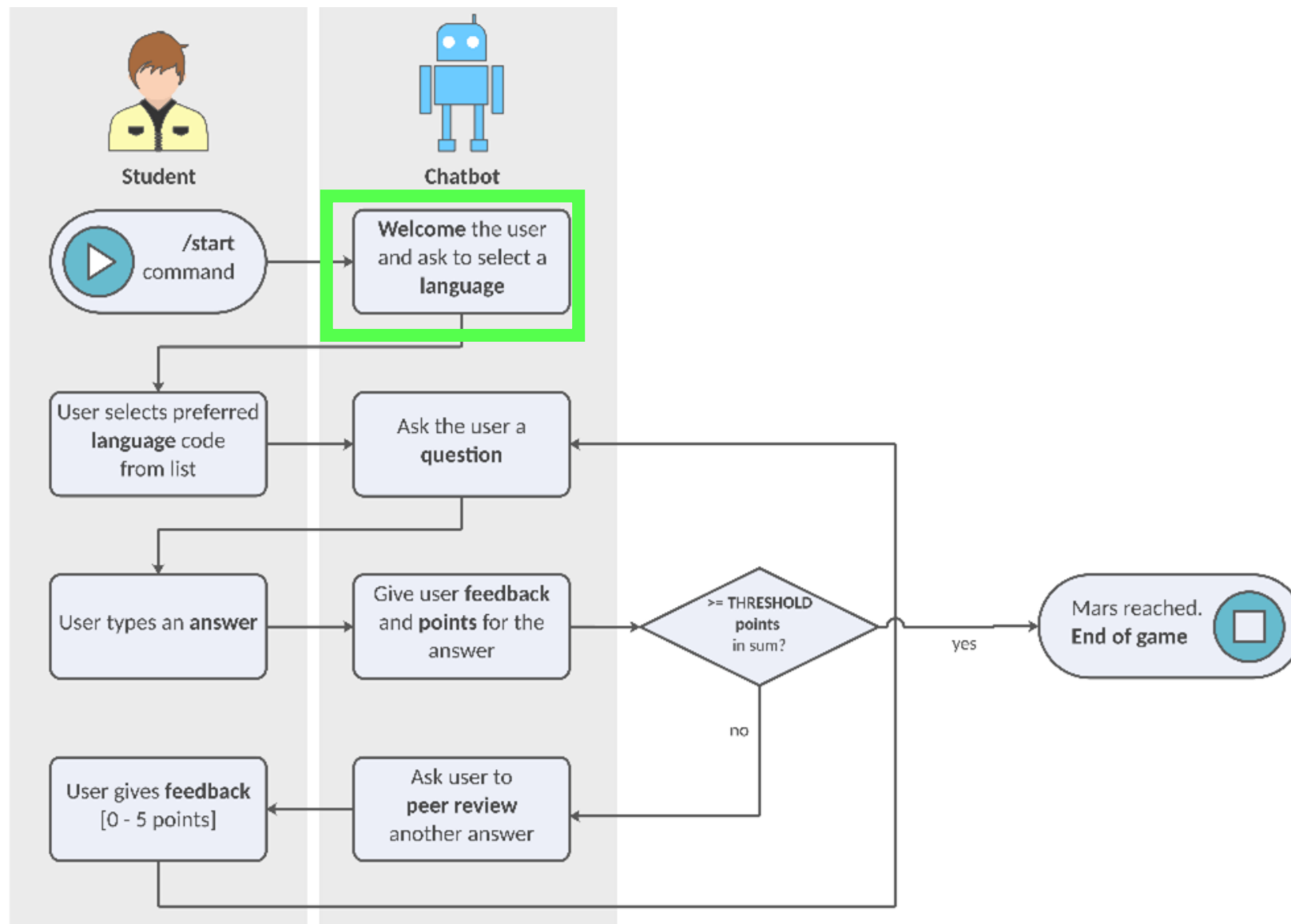
AI-BASED INTERACTIVE EXAM PREPARATION



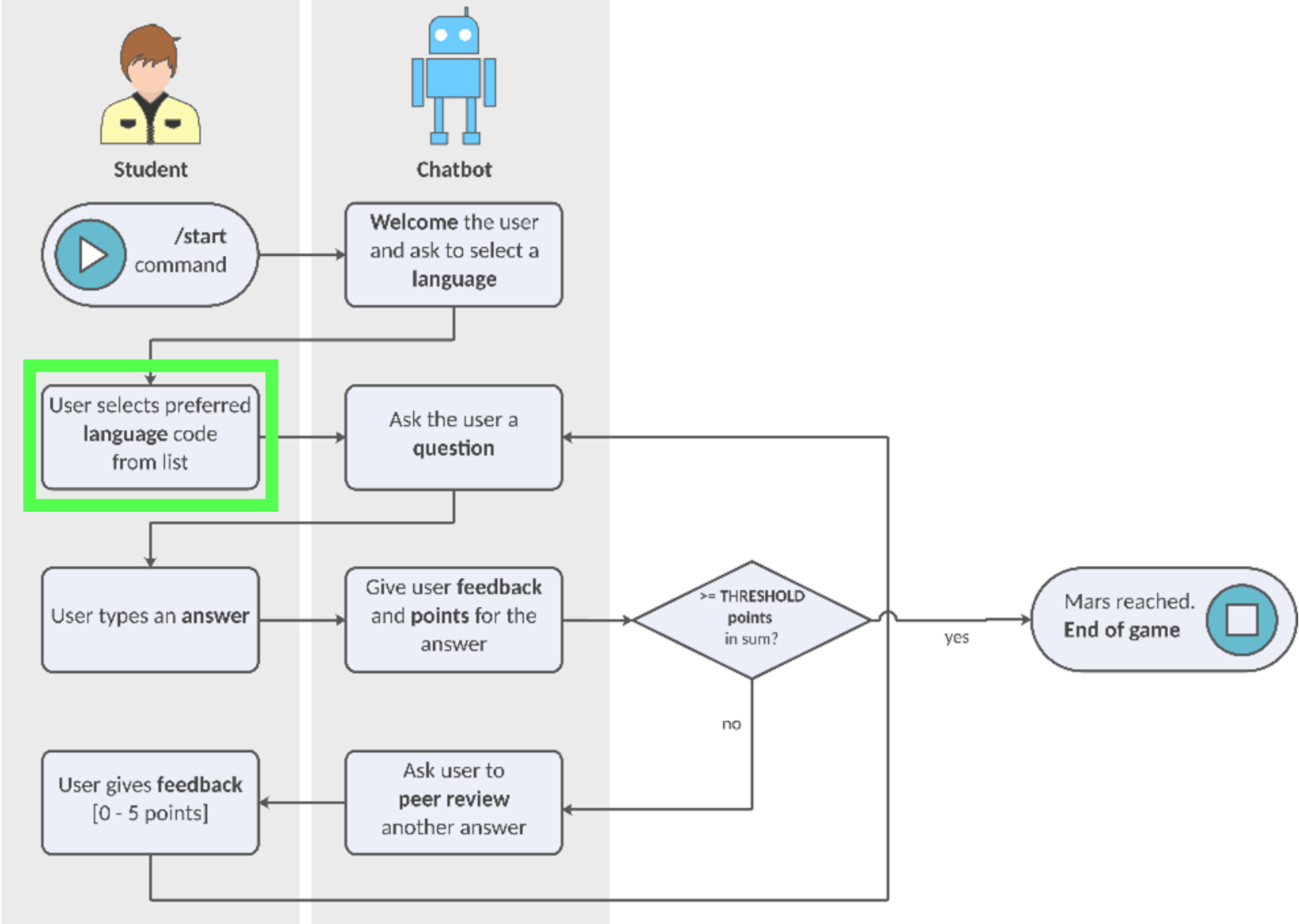
AI-BASED INTERACTIVE EXAM PREPARATION



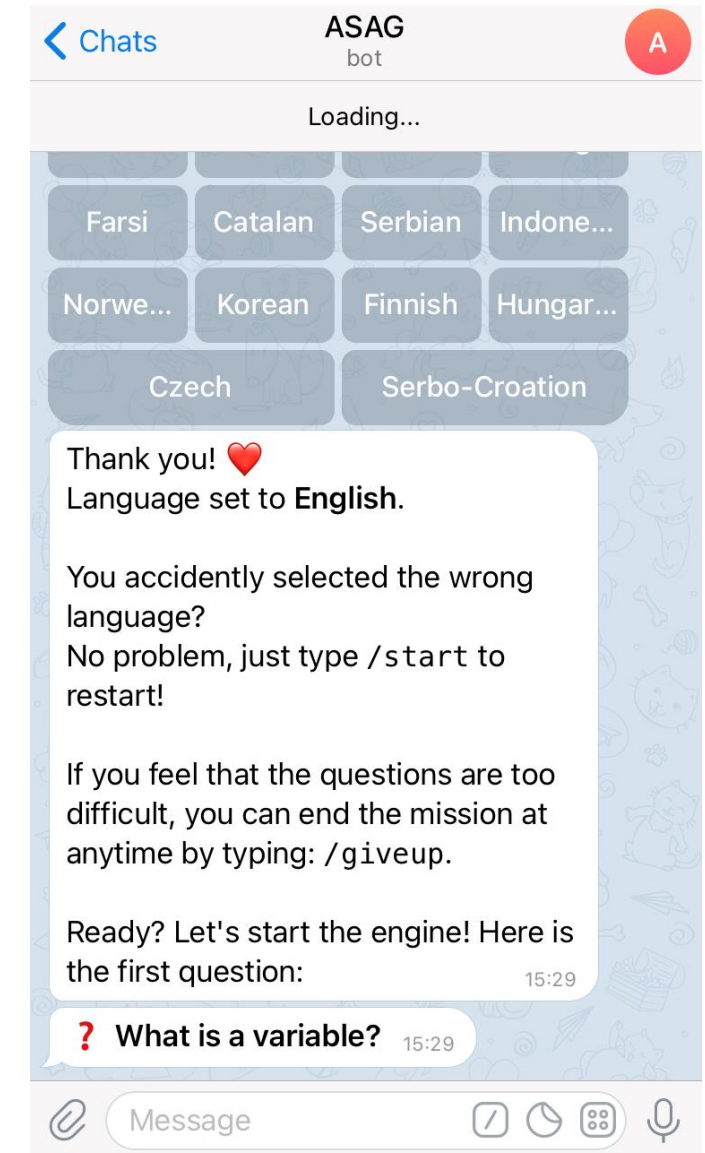
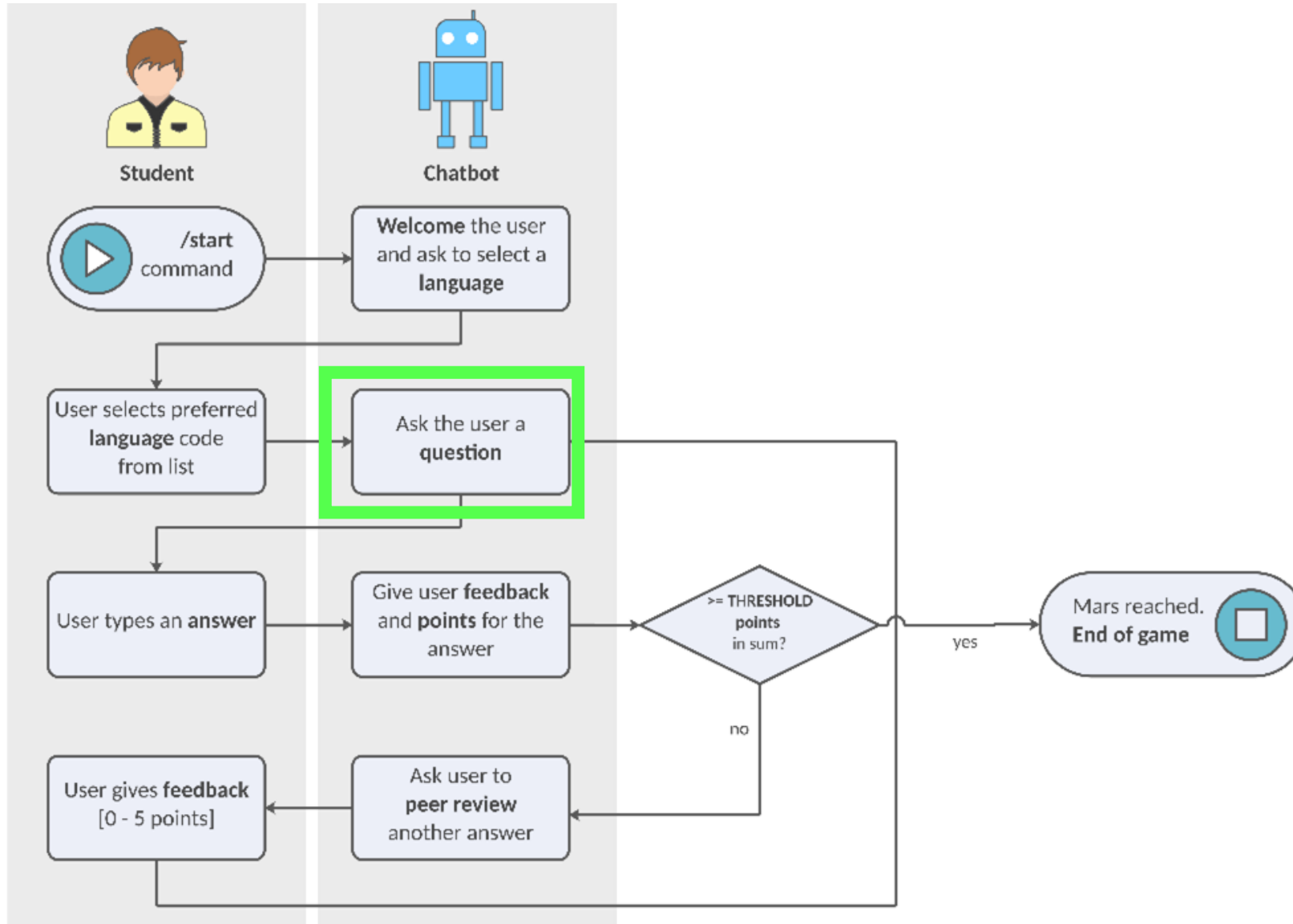
AI-BASED INTERACTIVE EXAM PREPARATION



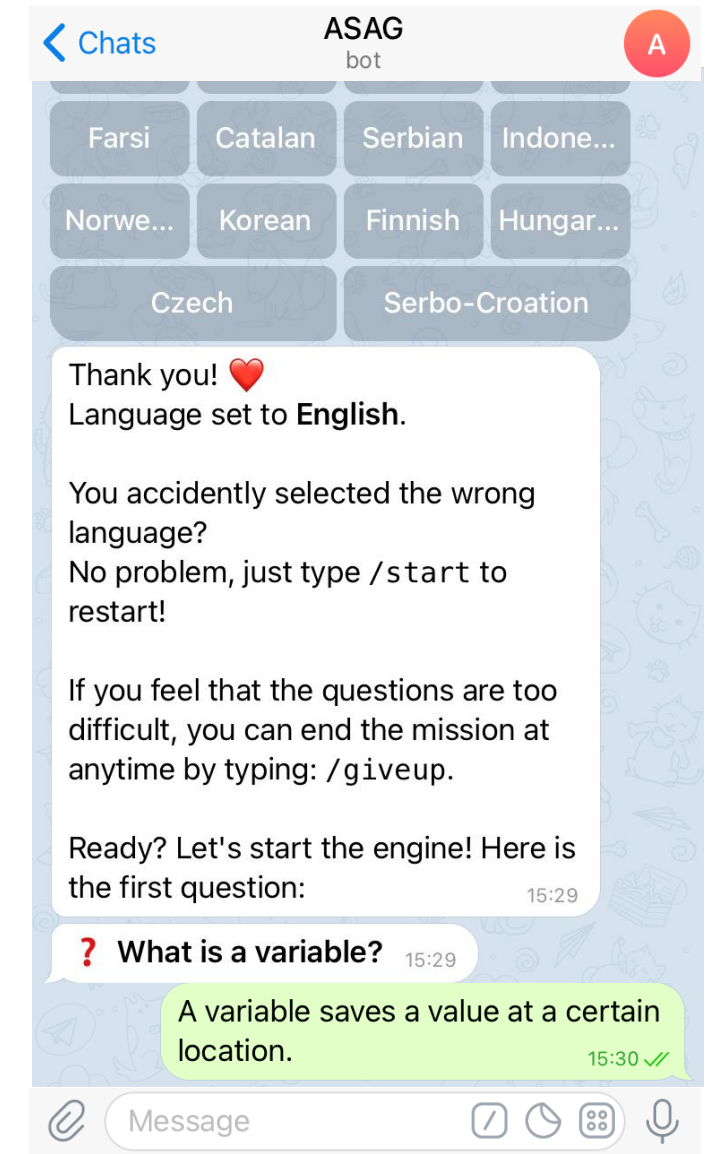
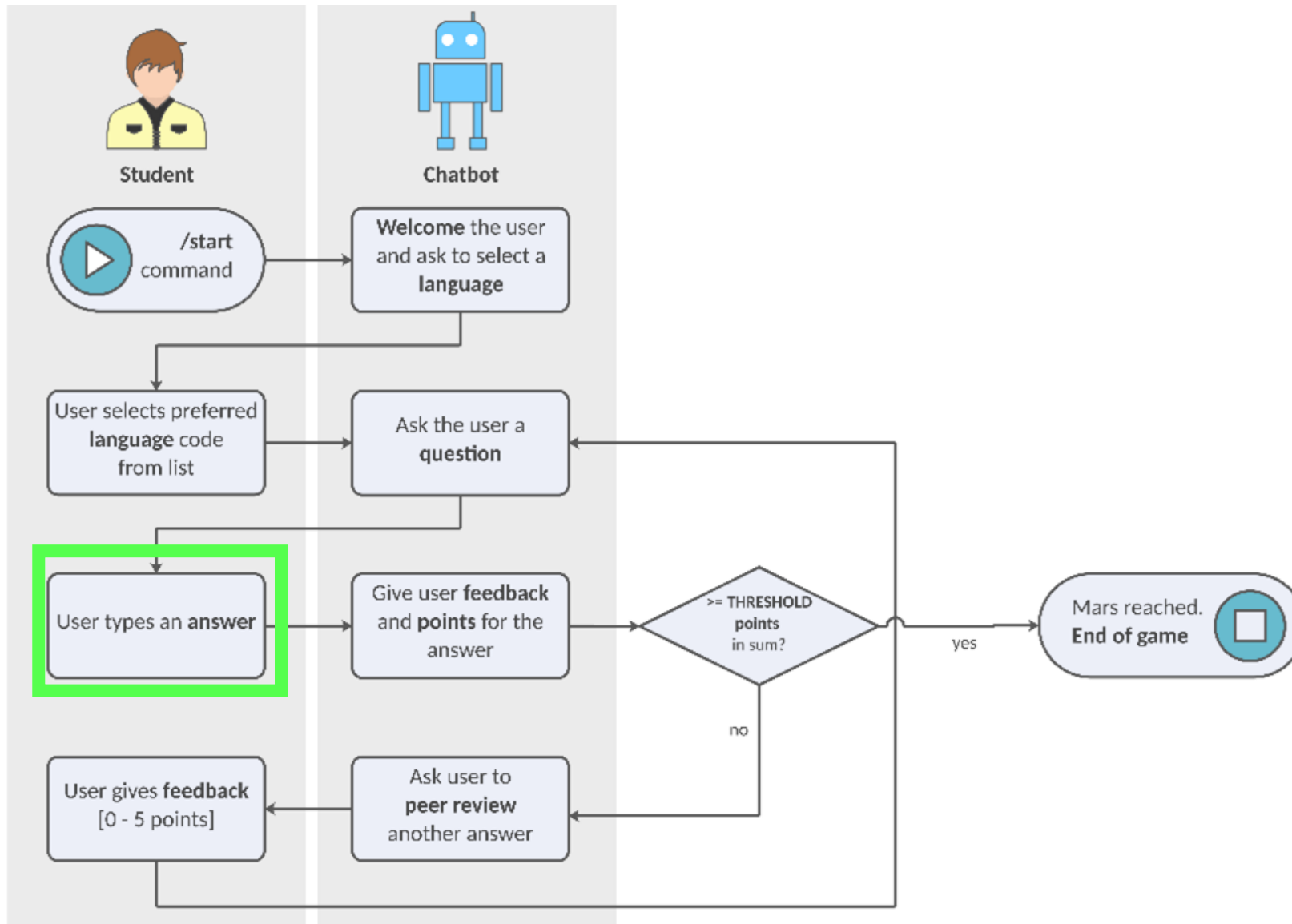
AI-BASED INTERACTIVE EXAM PREPARATION



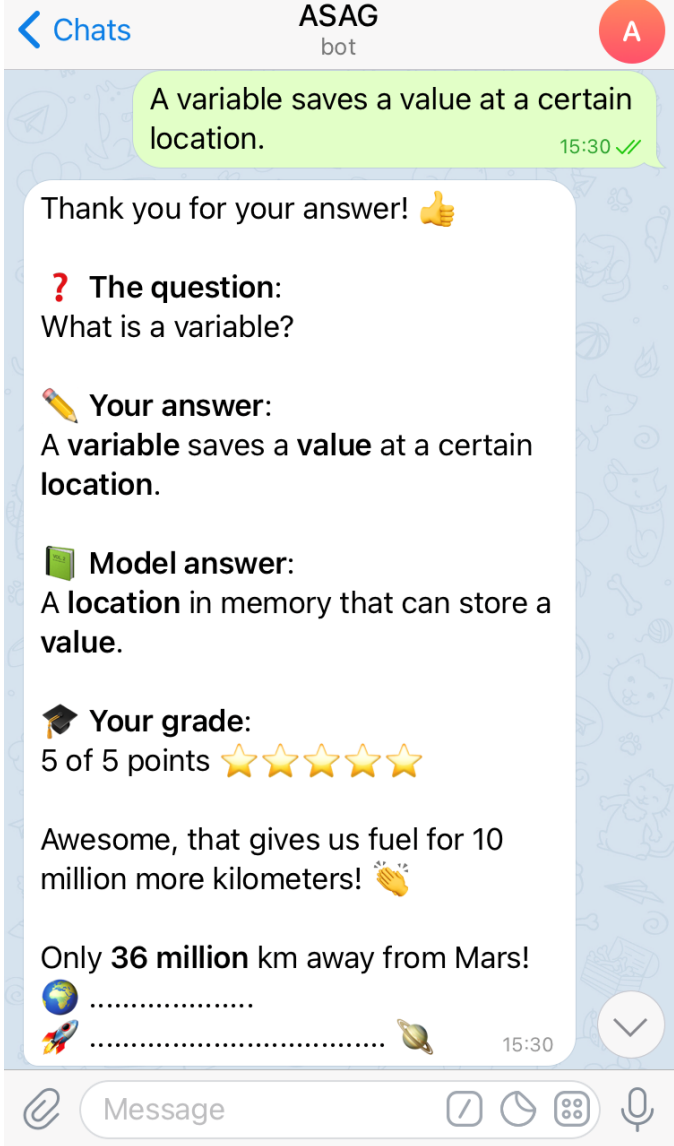
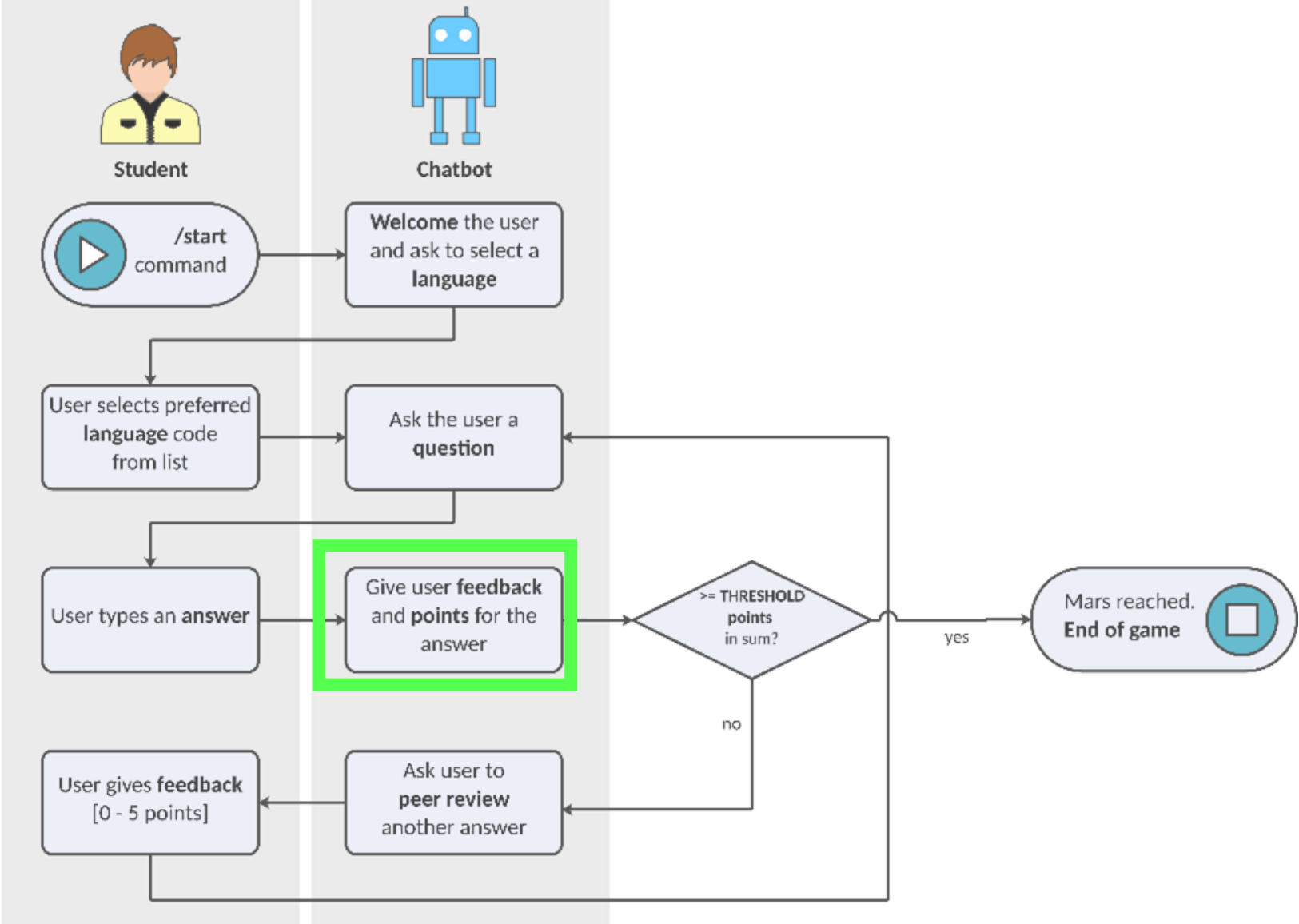
AI-BASED INTERACTIVE EXAM PREPARATION



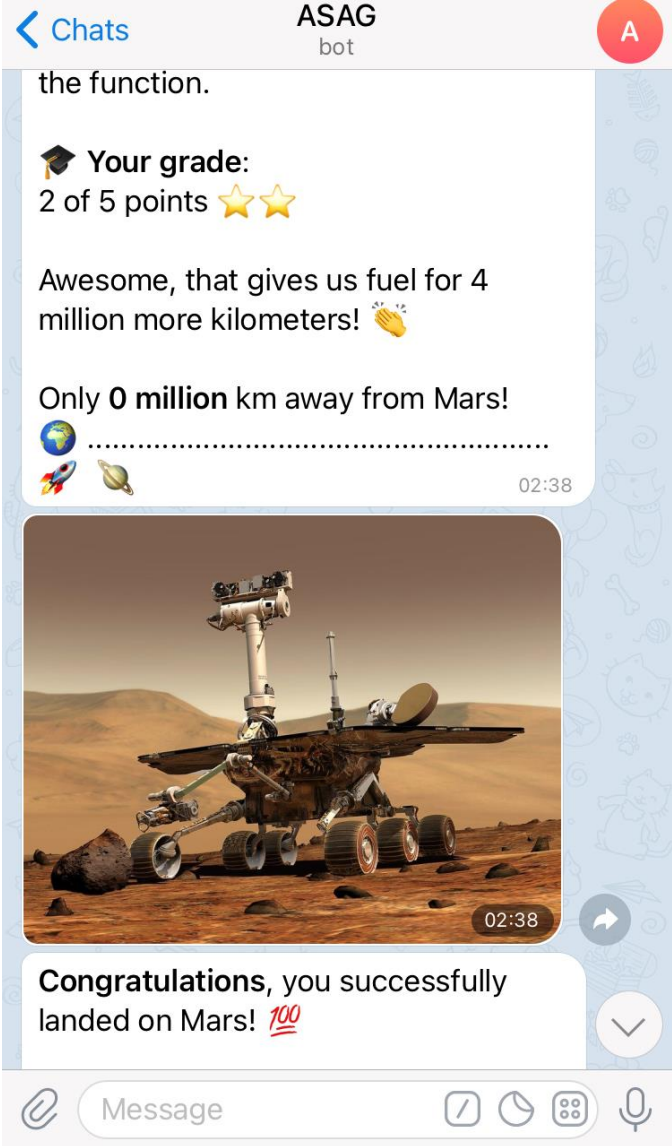
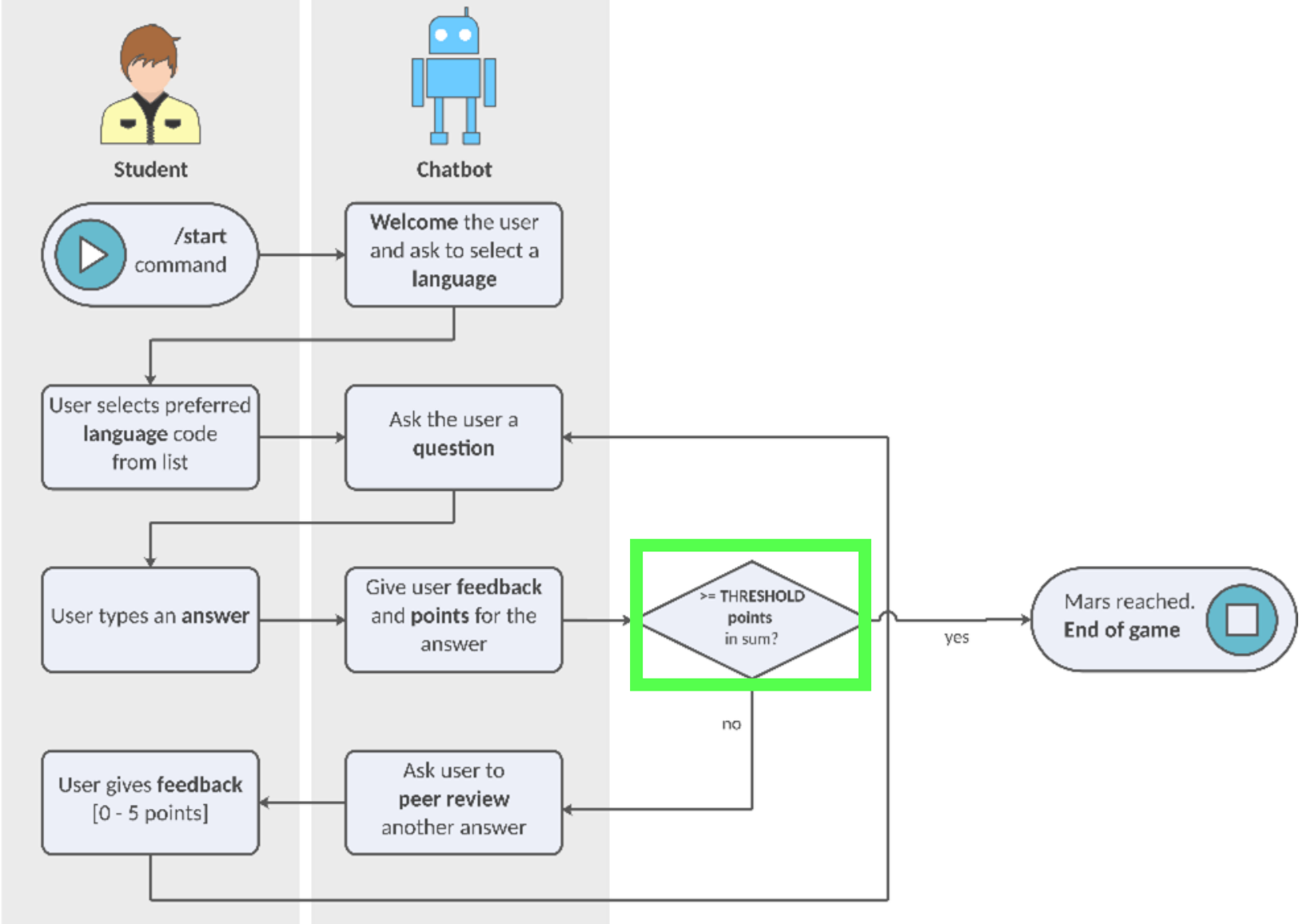
AI-BASED INTERACTIVE EXAM PREPARATION



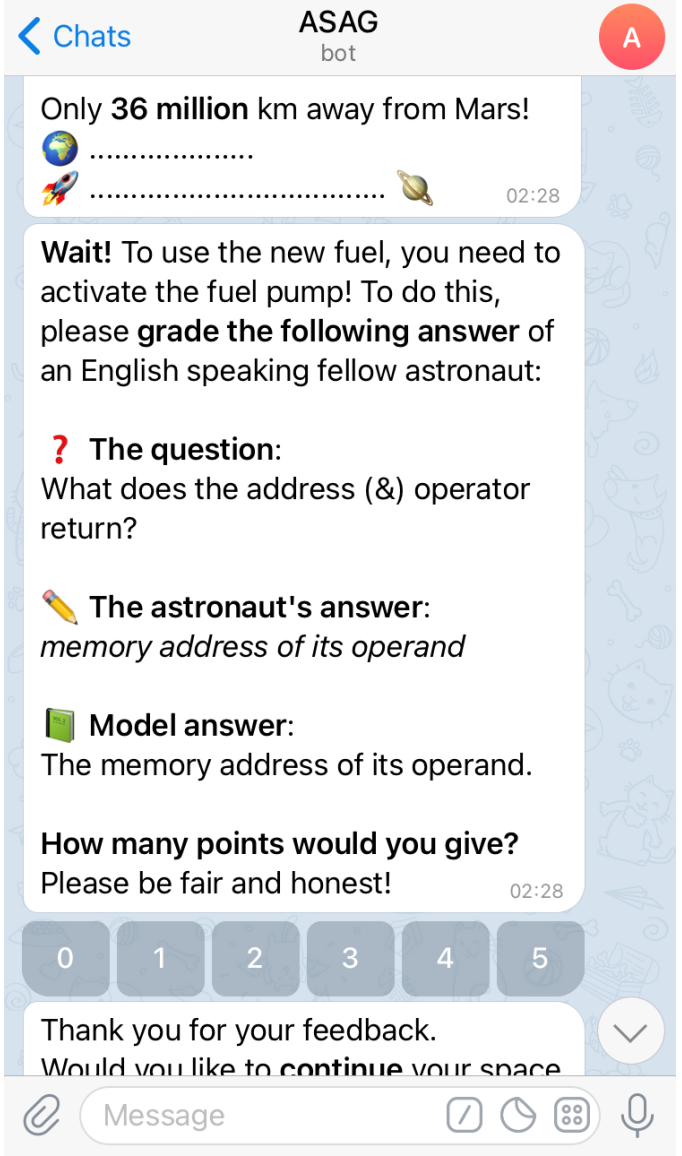
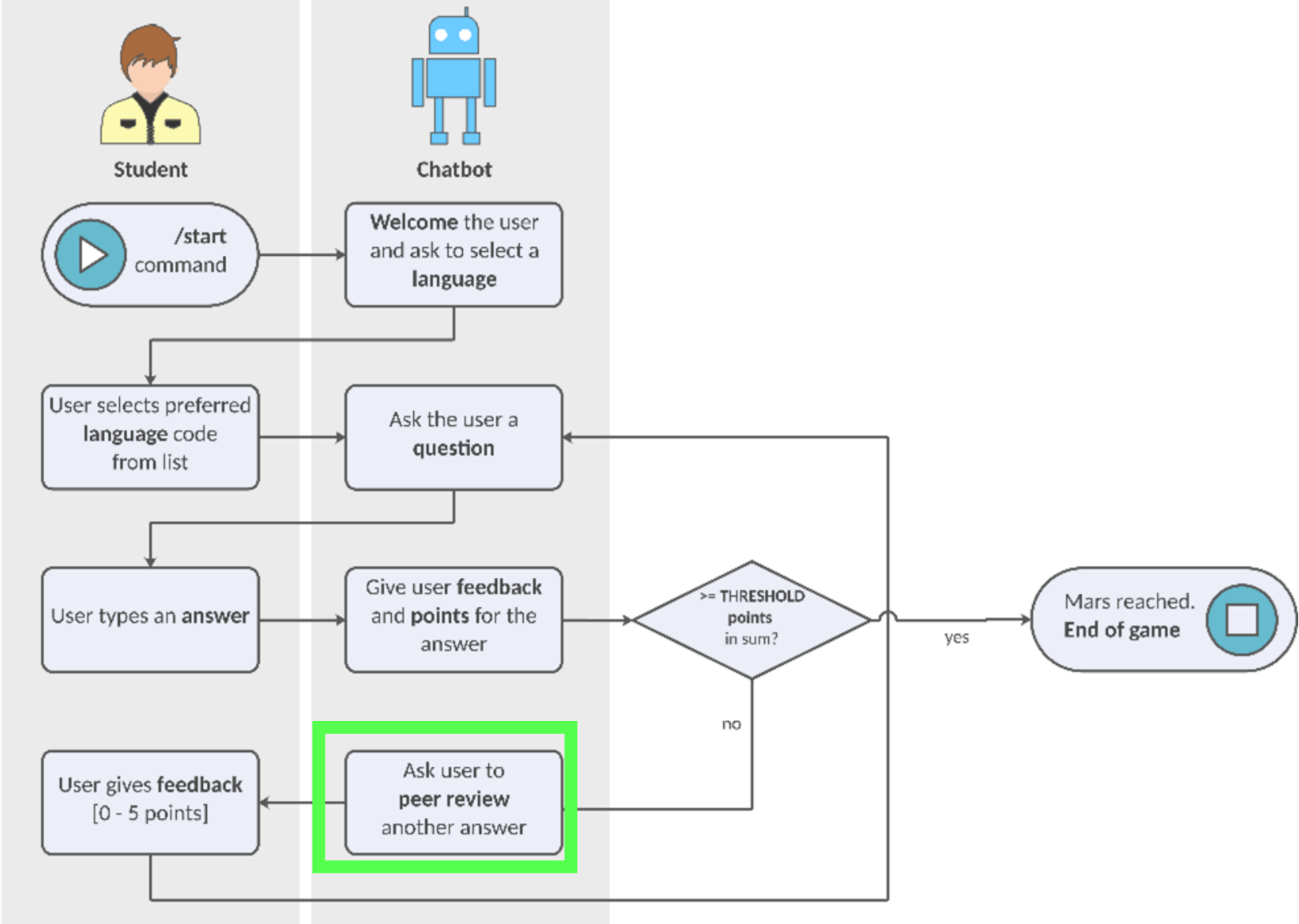
AI-BASED INTERACTIVE EXAM PREPARATION



AI-BASED INTERACTIVE EXAM PREPARATION



AI-BASED INTERACTIVE EXAM PREPARATION



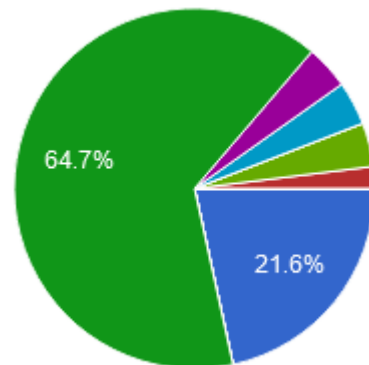
04

EXPERIMENTS AND RESULTS

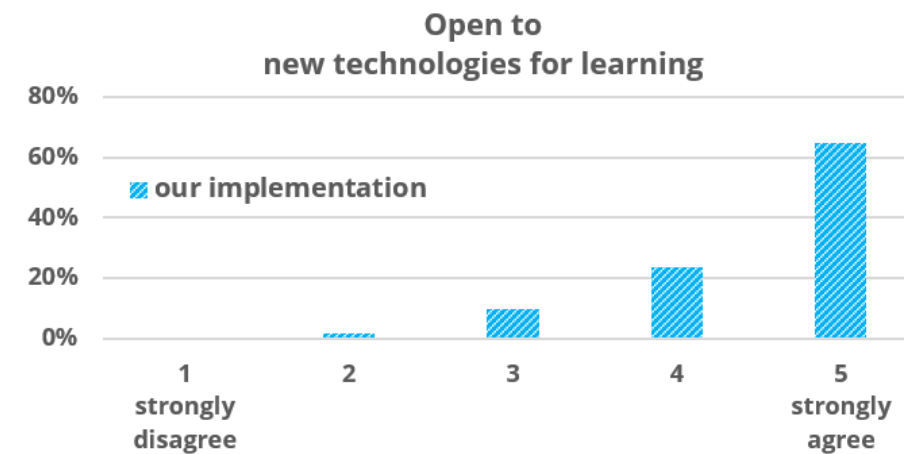
SURVEY



- 51 participants
- From 6 countries
- Most participants are students (most 18-44 years old)



- en (English)
- de (German)
- fr (French)
- nl (Dutch)
- it (Italian)
- es (Spanish)



GAMIFICATION

- Story
- Theme

USER EXPERIENCE

- Easy to use?
- Fun?

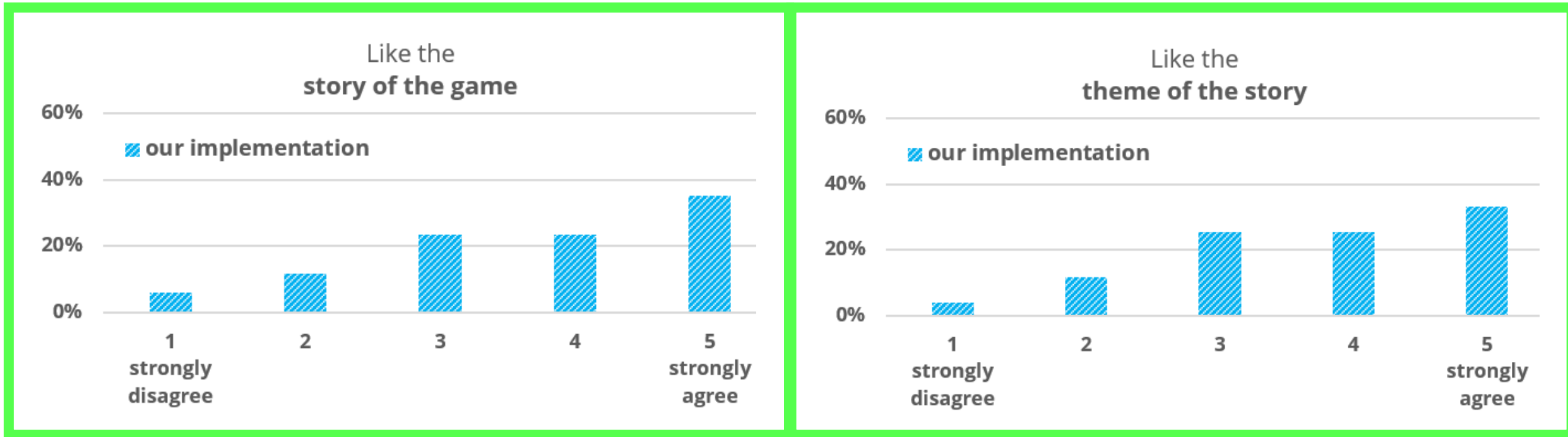
QUALITY OF THE NLP MODELS

- Translation quality?
- Scoring quality?
- Keyword highlighting?

LEARNING EXPERIENCE + MOTIVATION

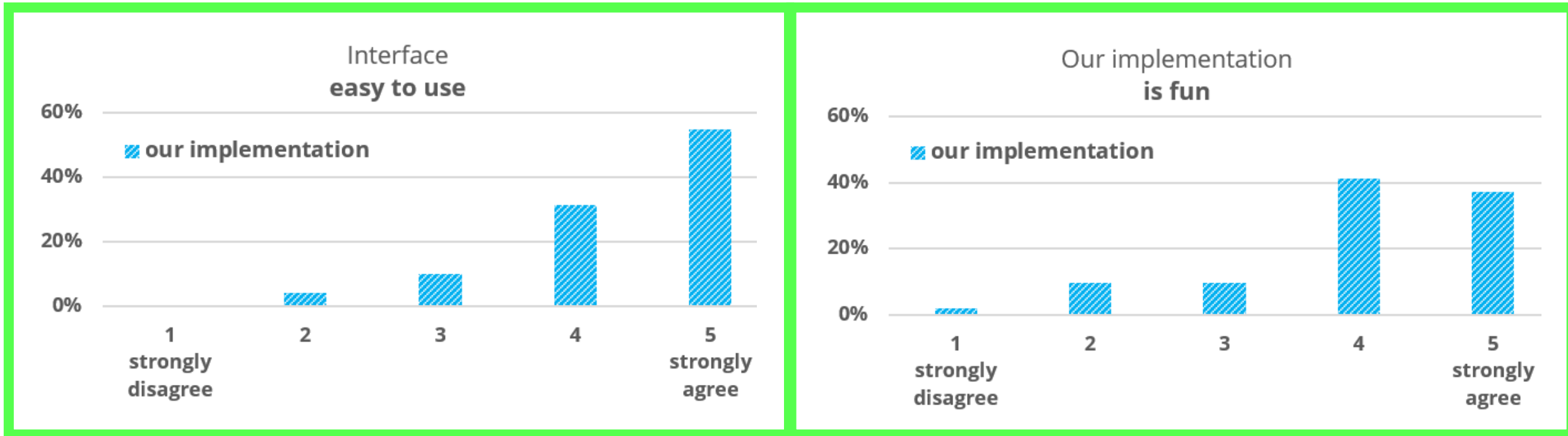
- Improved experience?
- Makes sense?
- Usage?
- Fellow students?
- Acceleration?
- Scoring ?
- Languages?
- Elderly people?
- Motivation?

SURVEY – GAMIFICATION



➔ Good gamification with a simple story and without special graphical features

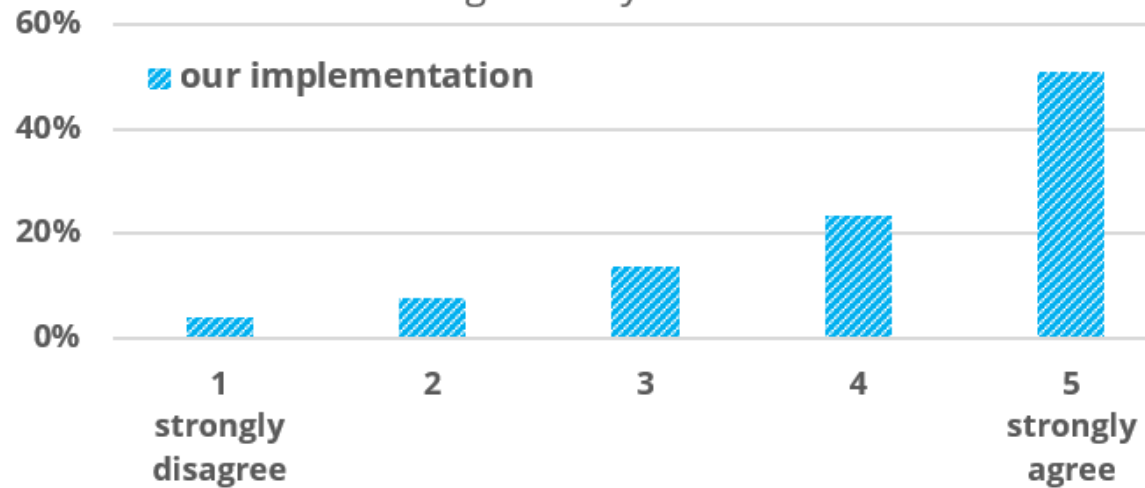
SURVEY – USER EXPERIENCE



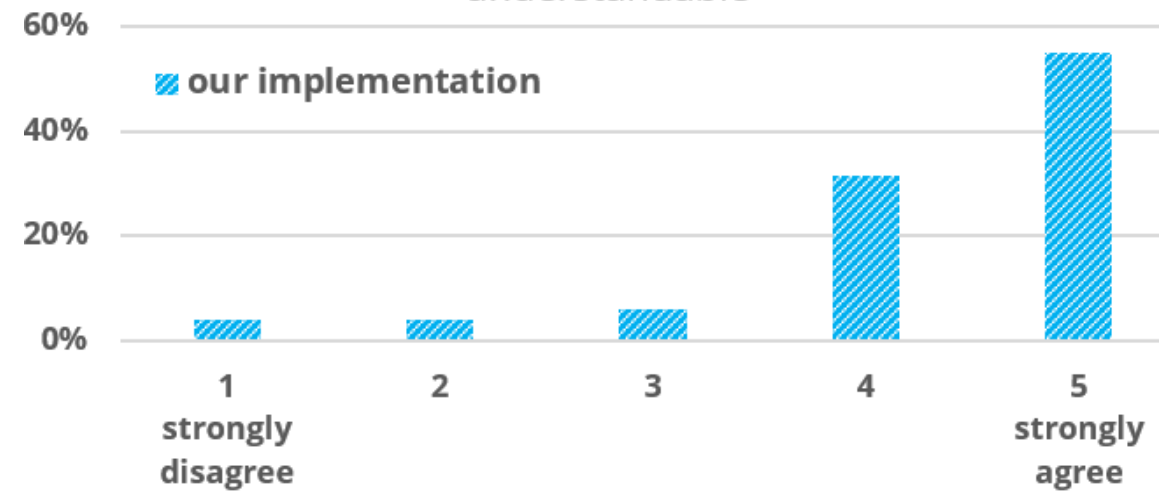
➔ Interface is easy to use and operating the exam trainer is fun

SURVEY – NLP MODELS (1)

Questions in native language
linguistically correct

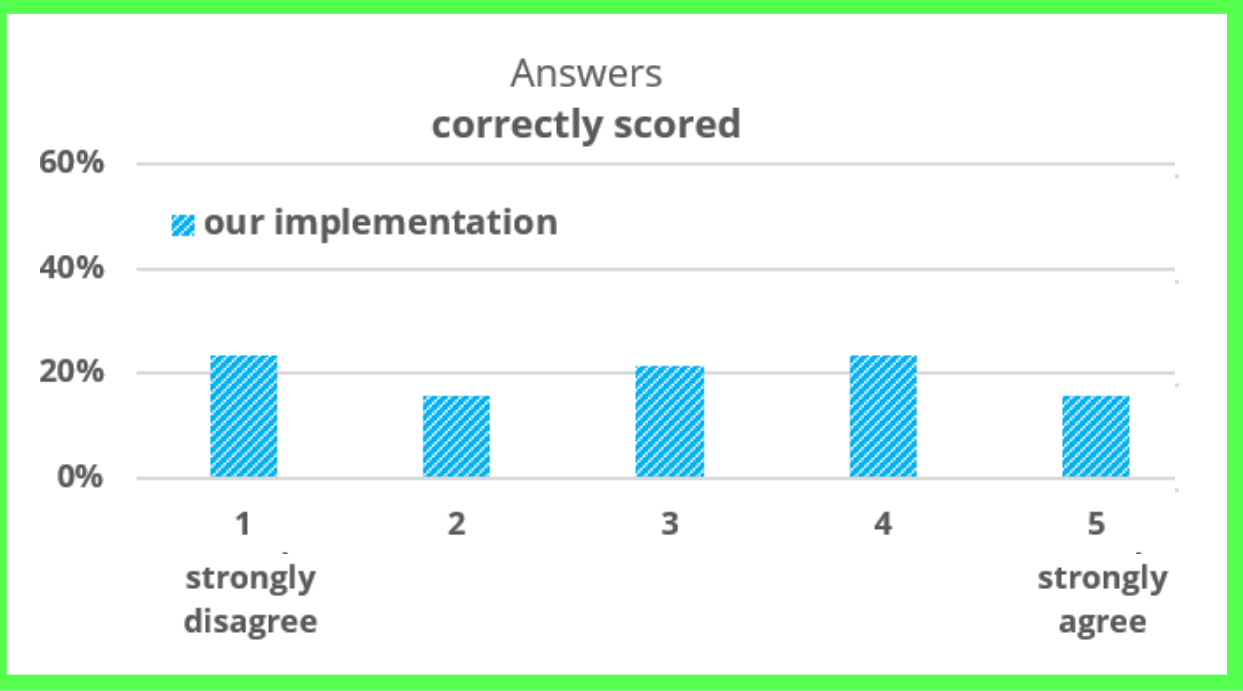


Questions in native language
understandable



➔ Majority rates the machine-translated questions as linguistically correct and understandable

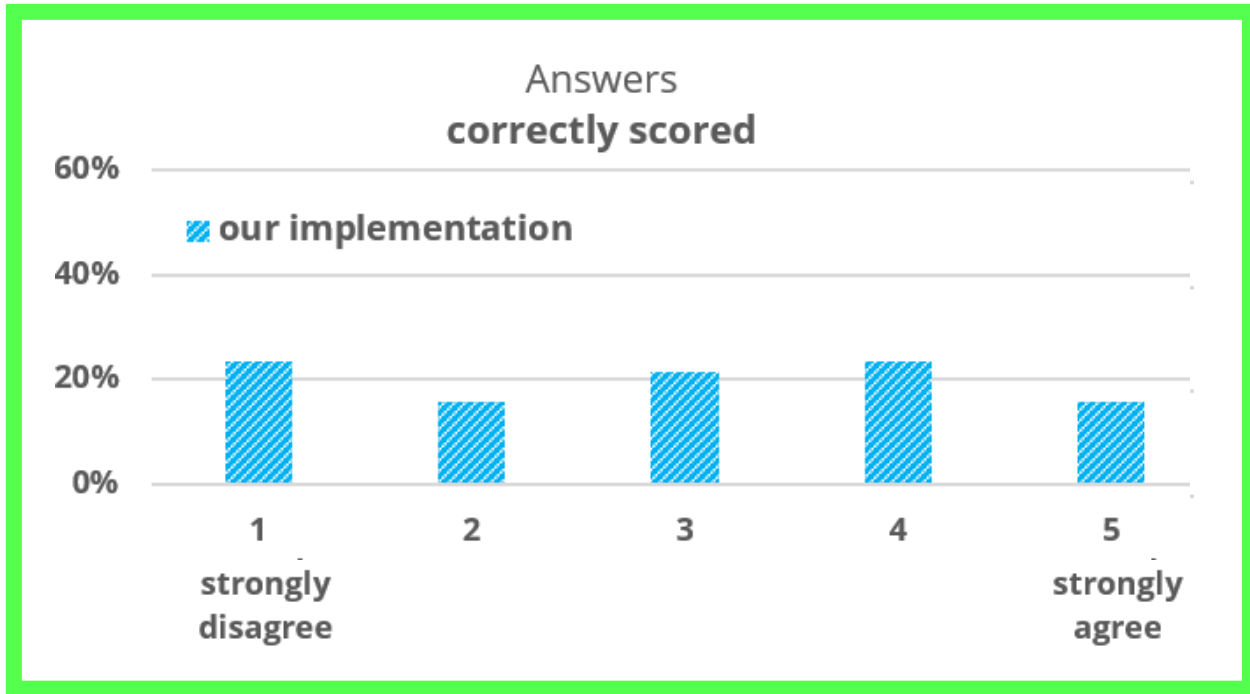
SURVEY – NLP MODELS (2)



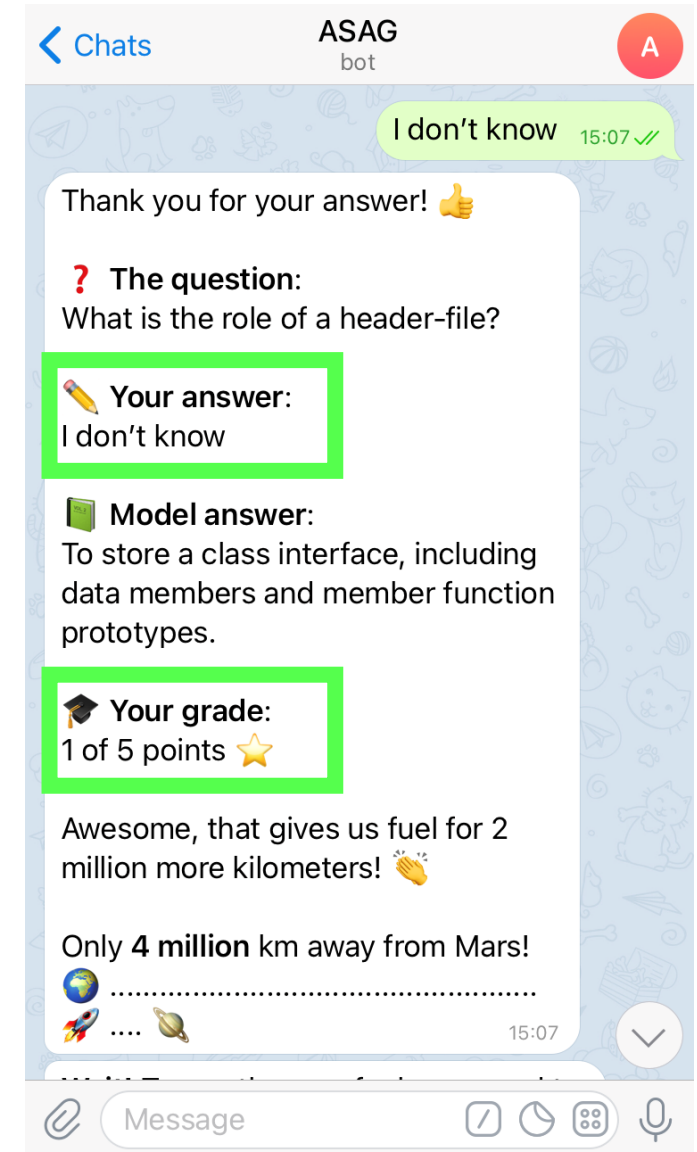
➔ Automatic scoring of the student answers was only rated average

Graphic Source: Schlippe & Sawatzki (2021).

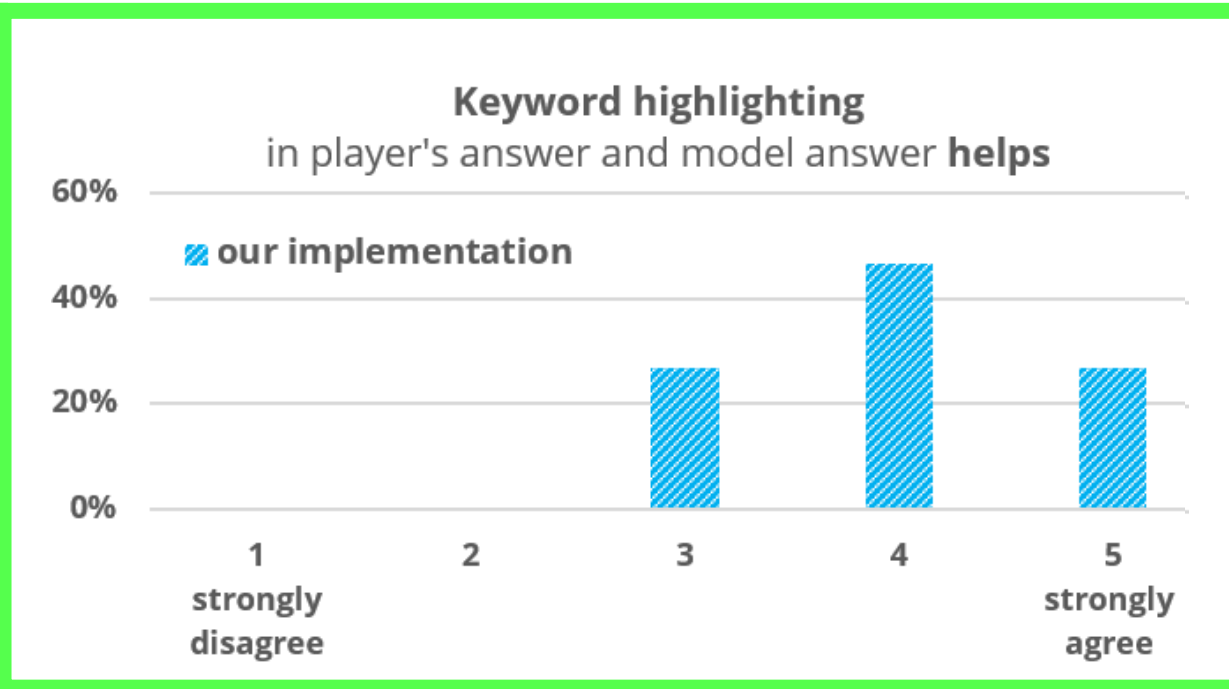
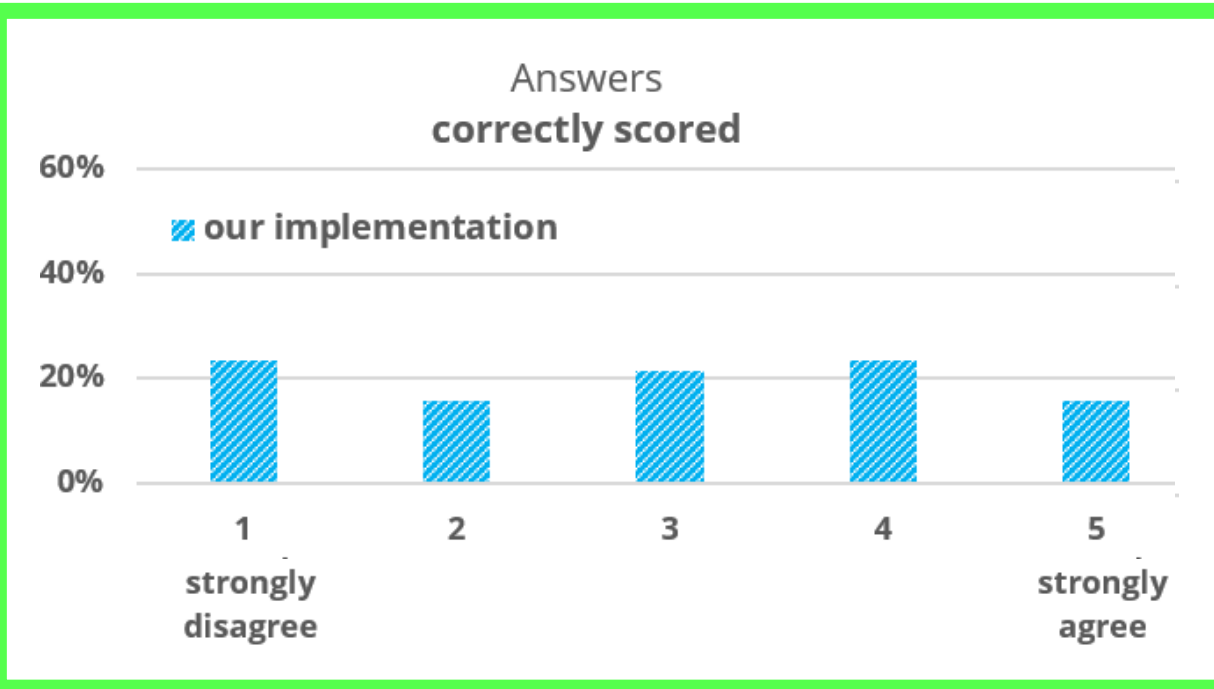
SURVEY – NLP MODELS (2)



➔ Automatic scoring of the student answers was only rated average



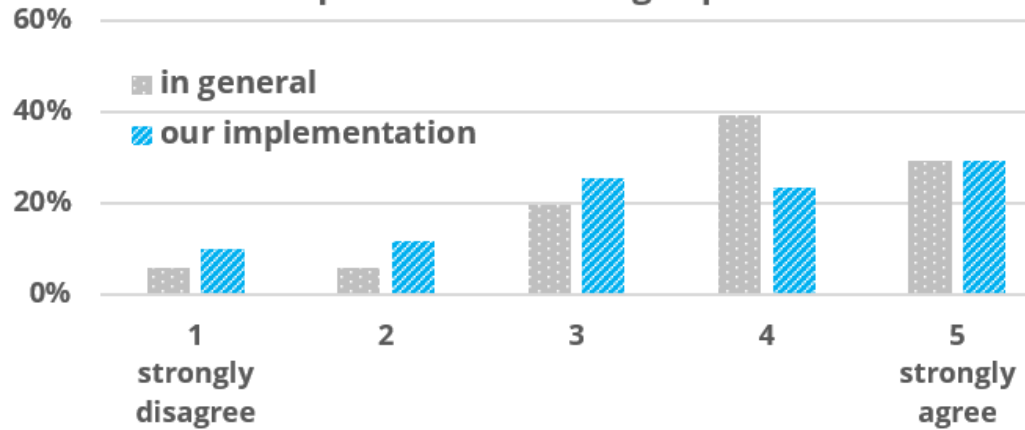
SURVEY – NLP MODELS (2)



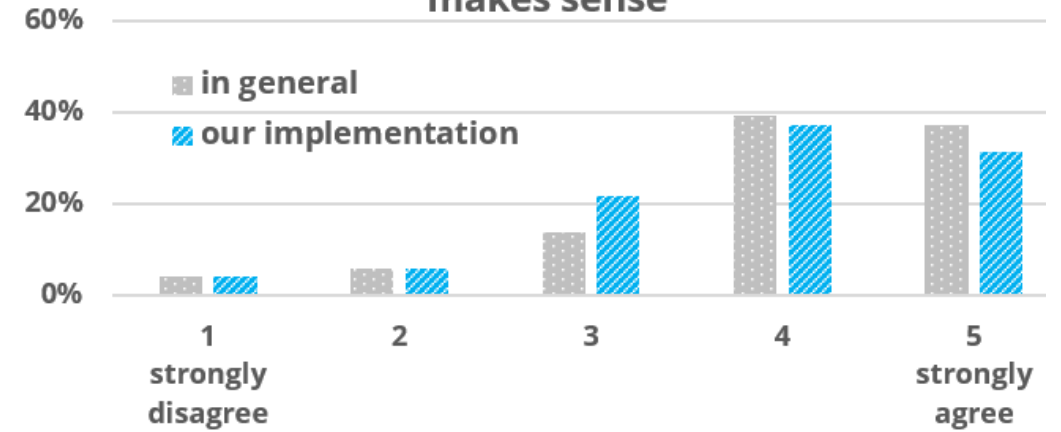
- ➔ Automatic scoring of the student answers was only rated average (Problem: Random answers)
- ➔ Explainability approach with keyword highlighting was well rated

SURVEY – LEARNING EXPERIENCE (1)

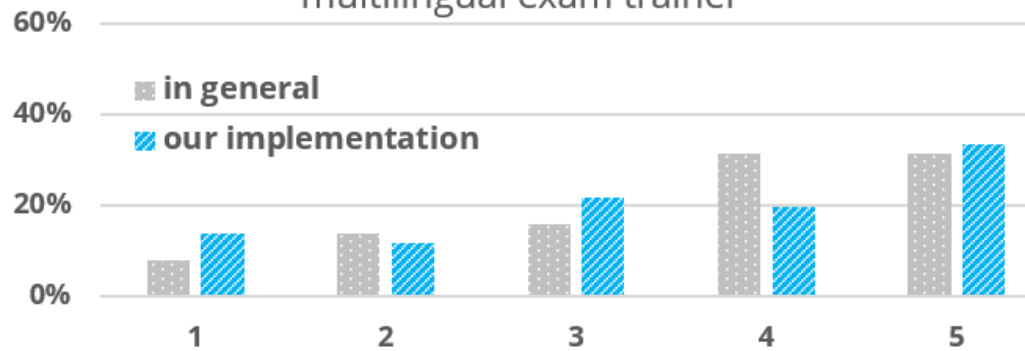
Multilingual exam trainer
improves the learning experience



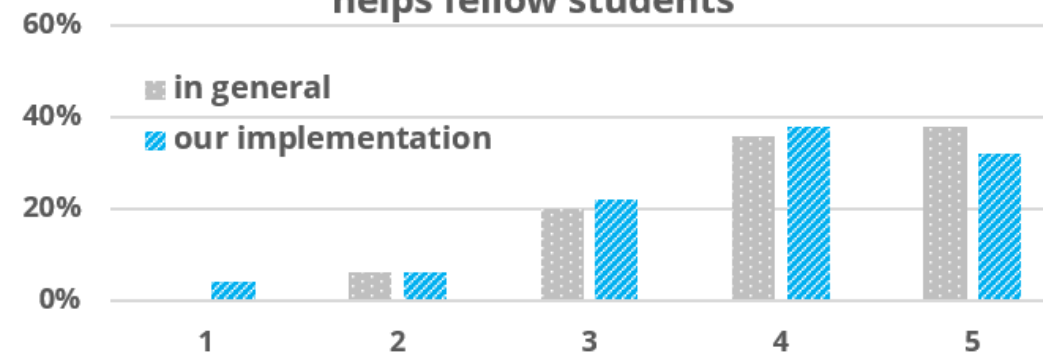
Multilingual exam trainer
makes sense



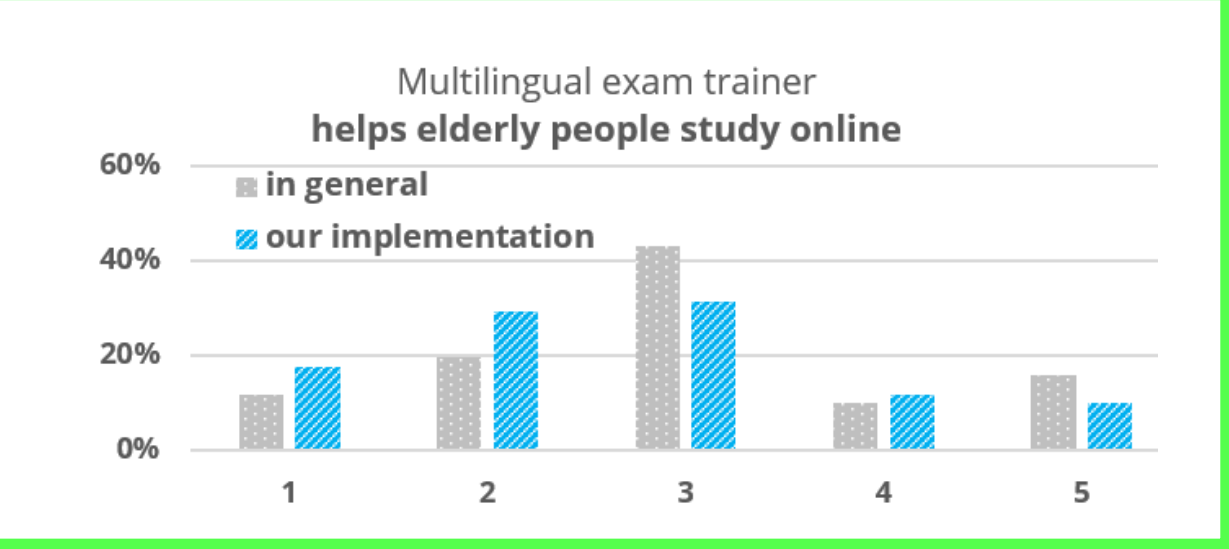
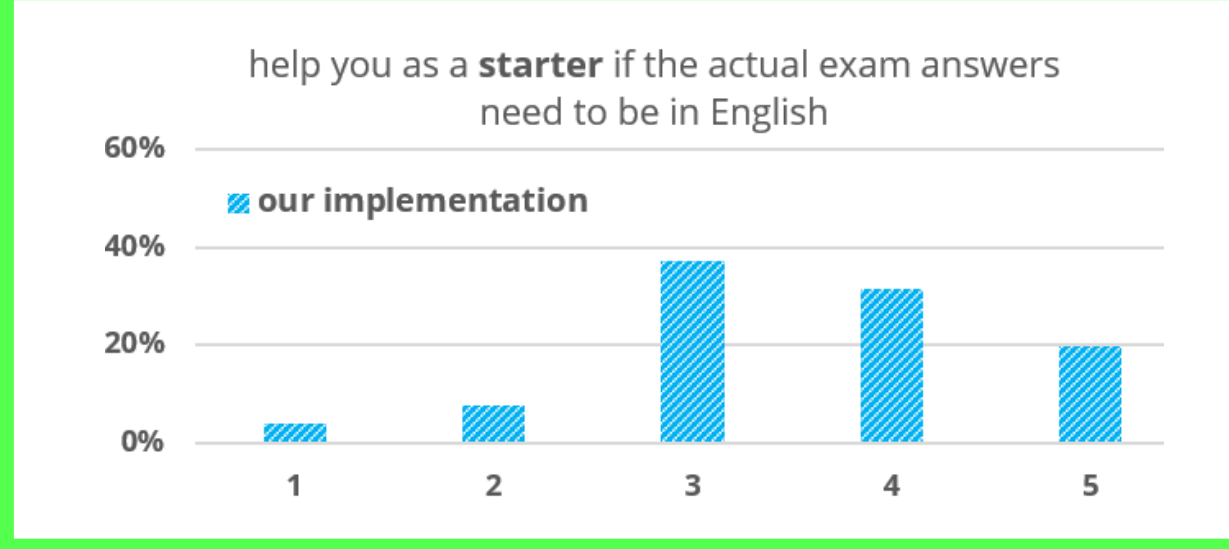
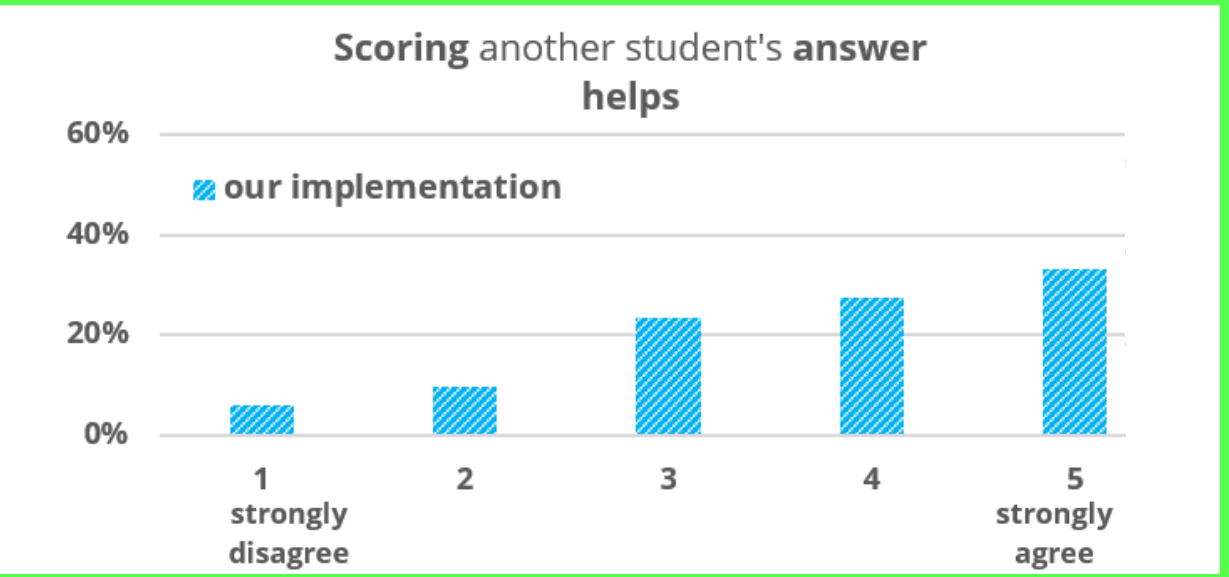
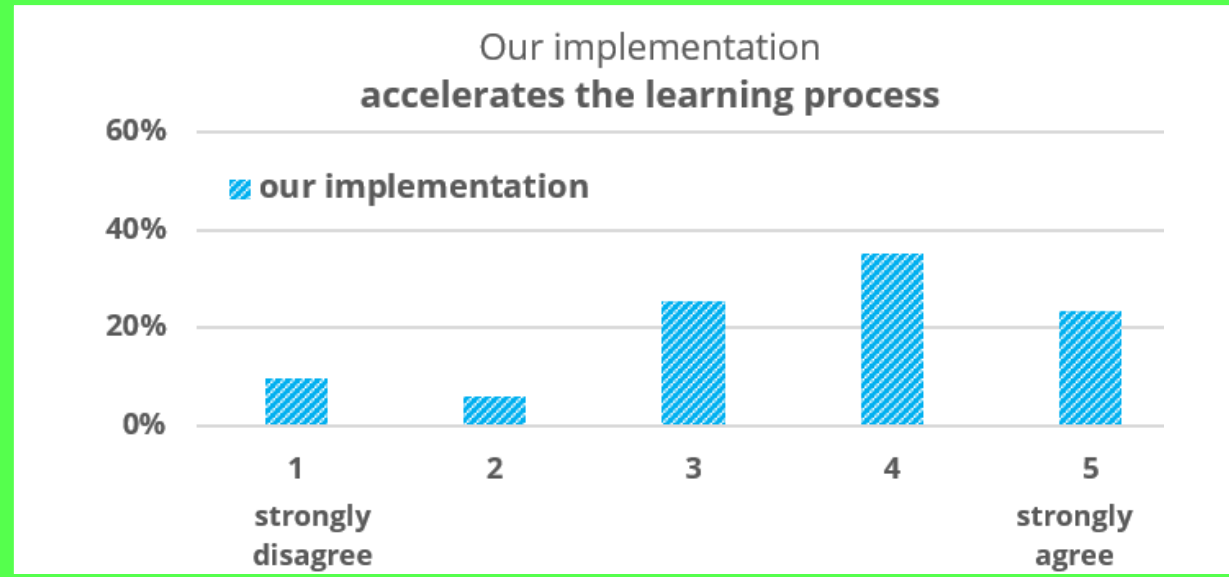
You would use a
multilingual exam trainer



Multilingual exam trainer
helps fellow students



SURVEY – LEARNING EXPERIENCE (2)



05

CONCLUSION AND FUTURE WORK

CONCLUSION AND FUTURE WORK

Conclusion

- Multilingual interactive conversational AI tutoring system for exam preparation
- combines multilingual NLP components, ASAG, conversational AI, keyword extraction, learning analytics, crowdsourcing, and gamification
- positive feedback in a survey regarding learning experience, user experience, motivation, quality of NLP models, and gamification
- proof-of-concept where users have tested 6 languages so far

CONCLUSION AND FUTURE WORK

Conclusion

- Multilingual interactive conversational AI tutoring system for exam preparation
- combines multilingual NLP components, ASAG, conversational AI, keyword extraction, learning analytics, crowdsourcing, and gamification
- positive feedback in a survey regarding learning experience, user experience, motivation, quality of NLP models, and gamification
- proof-of-concept where users have tested 6 languages so far

Future Work

- extension to other languages
- address the issue of explainability to provide even better support to the users
- more emotional dialog, e.g., (Wölfel et al., 2015; Schlippe et al., 2020)

THANK YOU

Tim Schlippe

 tim.schlippe@iu.org

REFERENCES

Literature

- **United Nations: Sustainable Development Goals: 17 Goals to Transform our World (2021):**
<https://www.un.org/sustainabledevelopment/sustainable-development-goals>
- **Statista: The Most Spoken Languages Worldwide in 2019 (2020):**
<https://www.statista.com/statistics/266808/the-most-spoken-languages-worldwide>
- **Schlippe, T. & Sawatzki, J. (2021):** *AI-based Multilingual Interactive Exam Preparation*. In: Proceedings of The Learning Ideas Conference 2021 (14th annual conference), ALICE - Special Conference Track on Adaptive Learning via Interactive, Collaborative and Emotional Approaches, New York, New York, 14-18 June 2021. In: *Advances in Intelligent Systems and Computing*, Springer, 2021.
- **Devlin, J., Chang, M.W., Lee, K., & Toutanova, K. (2019):** *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. In: NAACL-HLT, Minneapolis, Minnesota.
- **Pires, T., Schlinger, E., & Garrette, D. (2019):** *How Multilingual is Multilingual BERT?* In: ACL. Florence, Italy.
- **Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019):** *RoBERTa: A Robustly Optimized BERT Pretraining Approach*. arXiv:1907.11692.

Images

- **Images provided by OpenClipart-Vectors/154119/Pixabay.**
(<https://pixabay.com/vectors/international-project-world-154119> [last access: 16.16.2021])
- **Coursebook (NLP and Computer Vision):** IU International University of Applied Sciences. Version No.: 001-2020-1211.

Literature

- **Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., & Stoyanov, V. (2018):** *Unsupervised Cross-lingual Representation Learning at Scale*. arXiv:1911.02116.
- **Burrows, S., Gurevych, I., & Stein, B. (2014):** *The Eras and Trends of Automatic Short Answer Grading*. In: IJAIED 25, 60–117.
- **Camus, L. & Filighera, A. (2020):** *Investigating Transformers for Automatic Short Answer Grading*. AIED. Cyberspace.
- **Sawatzki, J., Schlippe, T., & Benner-Wickner, M. (2021):** *Deep Learning Techniques for Automatic Short Answer Grading: Predicting Scores for English and German Answers*. In: AIET, Wuhan, China.
- **Schlippe, T. & Sawatzki, J. (2021b):** *Cross-Lingual Automatic Short Answer Grading*. In: AIET, Wuhan, China.
- **Wölfel, M. (2021):** *Towards the Automatic Generation of Pedagogical Conversational Agents from Lecture Slides*. In: EAI ICMTTEL. Cyberspace.
- **Reyes, R., Garza, D., Garrido, L., De la Cueva, V., & Ramirez, J. (2019):** *Methodology for the Implementation of Virtual Assistants for Education Using Google Dialogflow*. In: MICAI. Xalapa, Mexico.
- **Setiaji, H. & Papatungan, I.V. (2017):** *Design of Telegram Bots for Campus Information Sharing*. In: ICITDA. Yogyakarta, Indonesia.
- **Hasan, K.S. (2014):** *Automatic Keyphrase Extraction: A Survey of the State of the Art*. In: ACL. Baltimore, Maryland, USA.

Literature

- **Merrouni, Z.A., Frikh, B., & Ouhbi, B. (2020):** *Automatic Keyphrase Extraction: A Survey and Trends*. In: JIIS. vol. 54, pp. 391–424.
- **Chandrasekaran, D. & Mago, V. (2021):** *Evolution of Semantic Similarity - A Survey*. arXiv: 2004.13820.
- **Wu, Y., Schuster, M., Chen, Z., Le, Q.V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., Klingner, J., Shah, A., Johnson, M., Liu, X., Kaiser, L., Gouws, S., Kato, Y., Kudo, T., Kazawa, H., Stevens, K., Kurian, G., Patil, N., Wang, W., Young, C., Smith, J., Riesa, J., Rudnick, A., Vinyals, O., Corrado, G., Hughes, M., & Dean, J. (2016):** *Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation*. CoRR abs/1609.08144.
- **Aiken, M. (2019):** *An Updated Evaluation of Google Translate Accuracy*. *Studies in Linguistics and Literature* 3, 253.
- **Luo, H., Robinson, A. & Park, J.-Y. (2014):** *Peer Grading in a MOOC: Reliability, Validity, and Perceived Effects*. *Online Learning: Official Journal of the Online Learning Consortium*. 18. 1-14. 10.24059/olj.v18i2.429.
- **Wölfel, M., Schlippe, T., & Stitz, A. (2015):** *Voice Driven Type Design*. In: SpeD. Bucharest, Romania.
- **Schlippe, T., Alessai, S., El-Taweel, G., Wölfel, & M., Zaghouani, W. (2020):** *Visualizing Voice Characteristics with Type Design in Closed Captions for Arabic*. In: *Cyberworlds*. Caen, France.