DCU
Ollscoil Chathair
Bhaile Átha Cliath
Dublin City University

Machine Translation Quality Assessment Lesson 3 - Evaluation Design

M.Ed. João Lucas Cavalheiro Camargo

Learning Outcomes

LO4 - Design replicable evaluation of MT systems, cognisant of the diverse evaluation approaches and types of evaluators in the process.

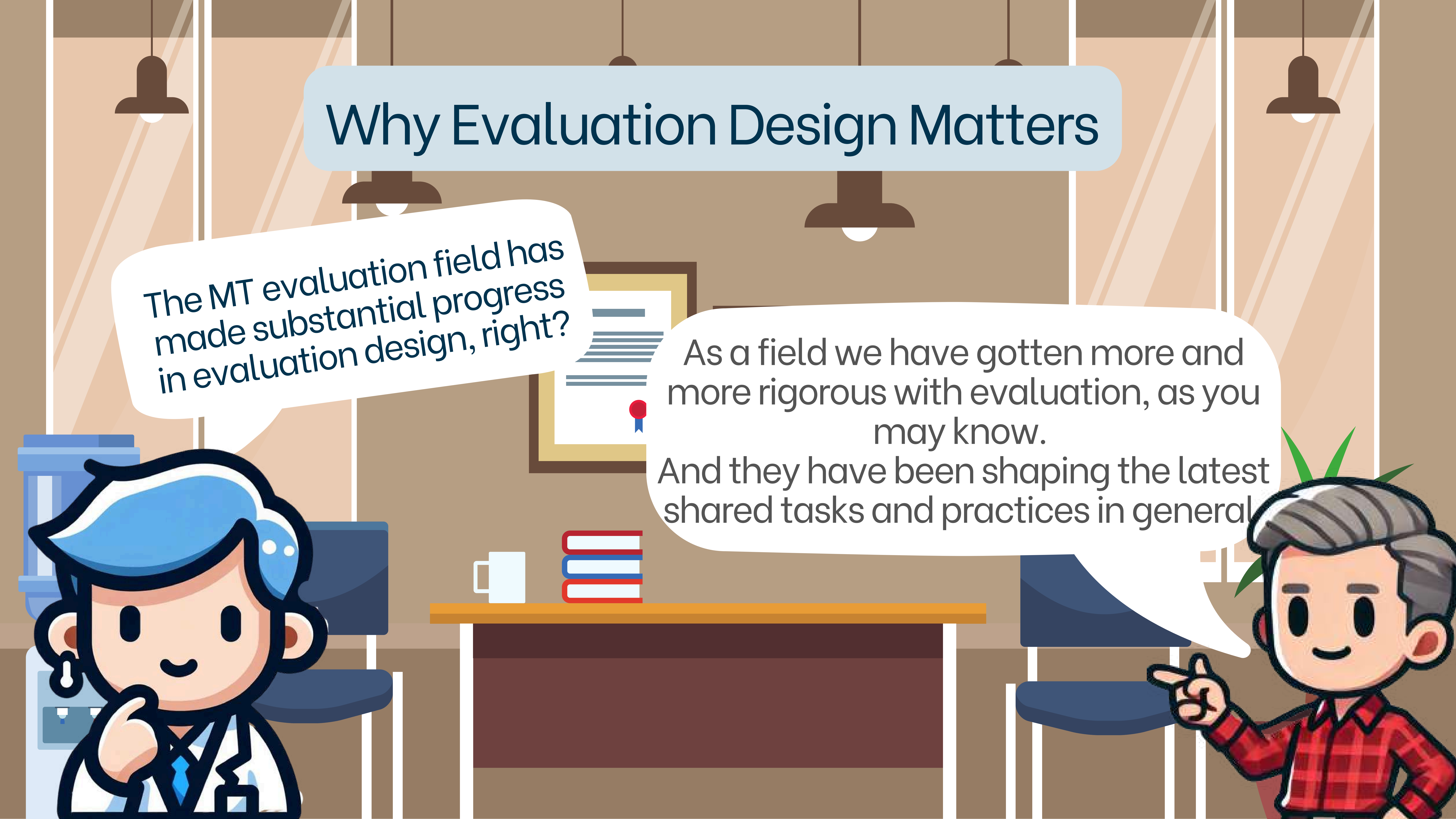
LO5 - Report results from the evaluation of an MT system addressing the context, the type of evaluators and the use case of the MT system.

Structure

1 - Recap
2 - Why Evaluation Design matters
3 - Who should Evaluate?
4 – What are guidelines for?
5 - Where should I evaluate?
6 - What scale should I use?
7 - Why should I use test suites?
8 - Why do I need to perform a risk assessment?

Why Evaluation Design matters?

Why Evaluation Design Matters



Why Evaluation Design Matters

Attaining the Unattainable? Reassessing Claims of Human Parity in Neural Machine Translation

Toral et al. (2018)

- Test sets should be the same as their source language to avoid spurious effects of translationese.
- Human evaluations should be conducted by professional translators.
- Human evaluations should consider the whole document.
- Test sets should be translated by experienced professional translators from scratch.

Why Evaluation Design Matters

A Set of Recommendations for Assessing Human-Machine Parity in Language Translation

Läubli et al. (2020)

- ✅ Choose professional translators as raters.
- ✅ Evaluate documents, not sentences.
- ✅ Evaluate fluency in addition to adequacy.
- ✅ Do not heavily edit reference translations for fluency.
- ✅ Use original source texts

Why Evaluation Design Matters

Translationese in Machine Translation Evaluation

Graham et al. (2020)

- ✅ Reverse-created data should be avoided
- ✅ Ensure high inter-annotator agreement levels or employ a reproducible method of human evaluation
- ✅ Results from the language pairs should not be generalised to other language pairs
- ✅ Results from a specific domain should not be generalised to other domains

Why Evaluation Design Matters

Translationese in Machine Translation Evaluation

Graham et al. (2020)

- ✅ The translation sample size (n) should be planned prior to the evaluation. This is to ensure a sample size sufficient to achieve strong statistical power (at least 80%).
- ✅ Human evaluation sample size (V) should be reported.
- ✅ Ensure the right statistical tests are applied and make sure to only group systems (or not) if they are statistically significant.

Why Evaluation Design Matters

An Example

Findings of the WMT 2023 Shared Task on Automatic Post-Editing

Pushpak Bhattacharyya
IIT Bombay
pb@cse.iitb.ac.in

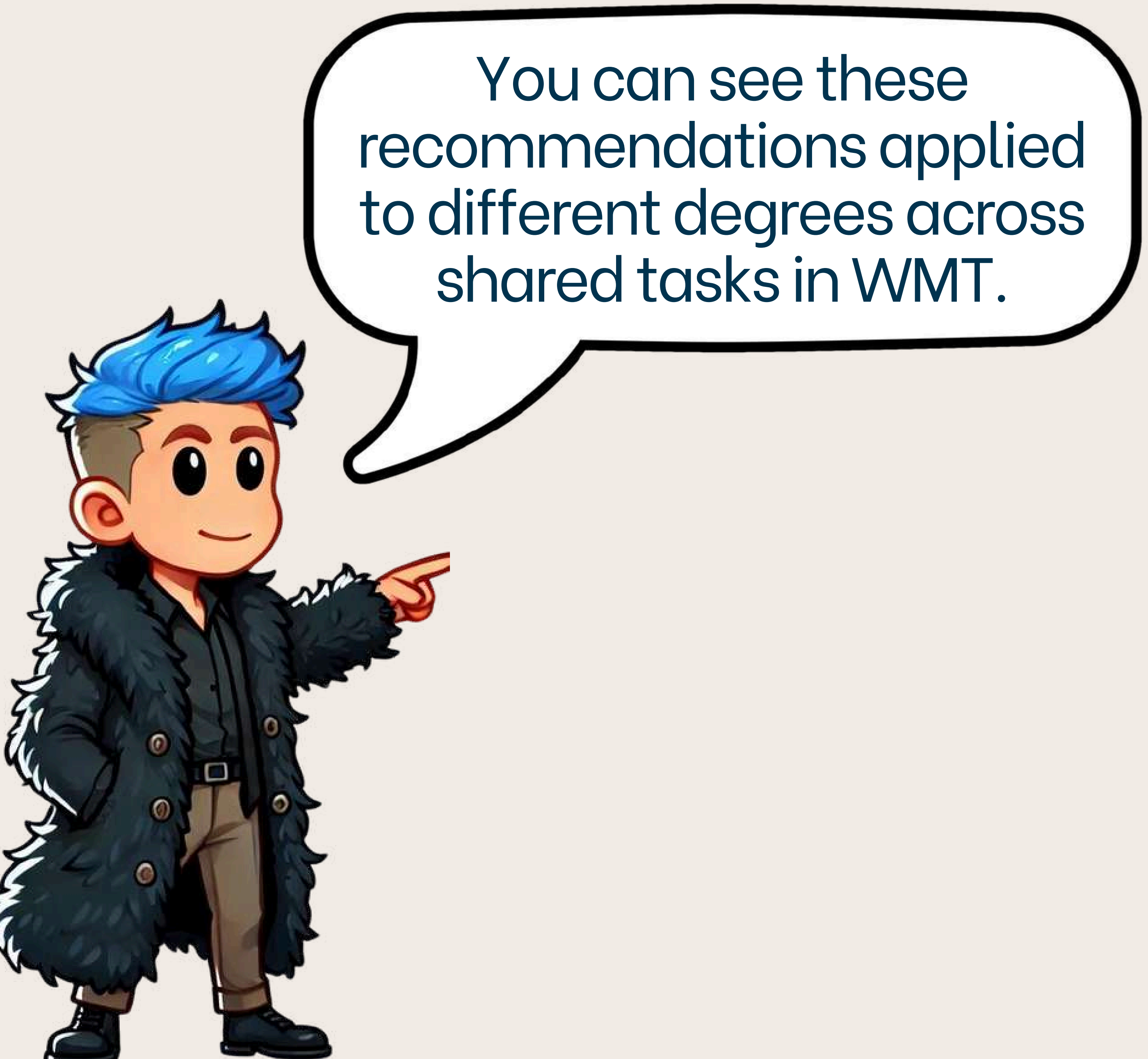
Rajen Chatterjee
Apple Inc.
rajen_c@apple.com

Markus Freitag
Google
freitag@google.com

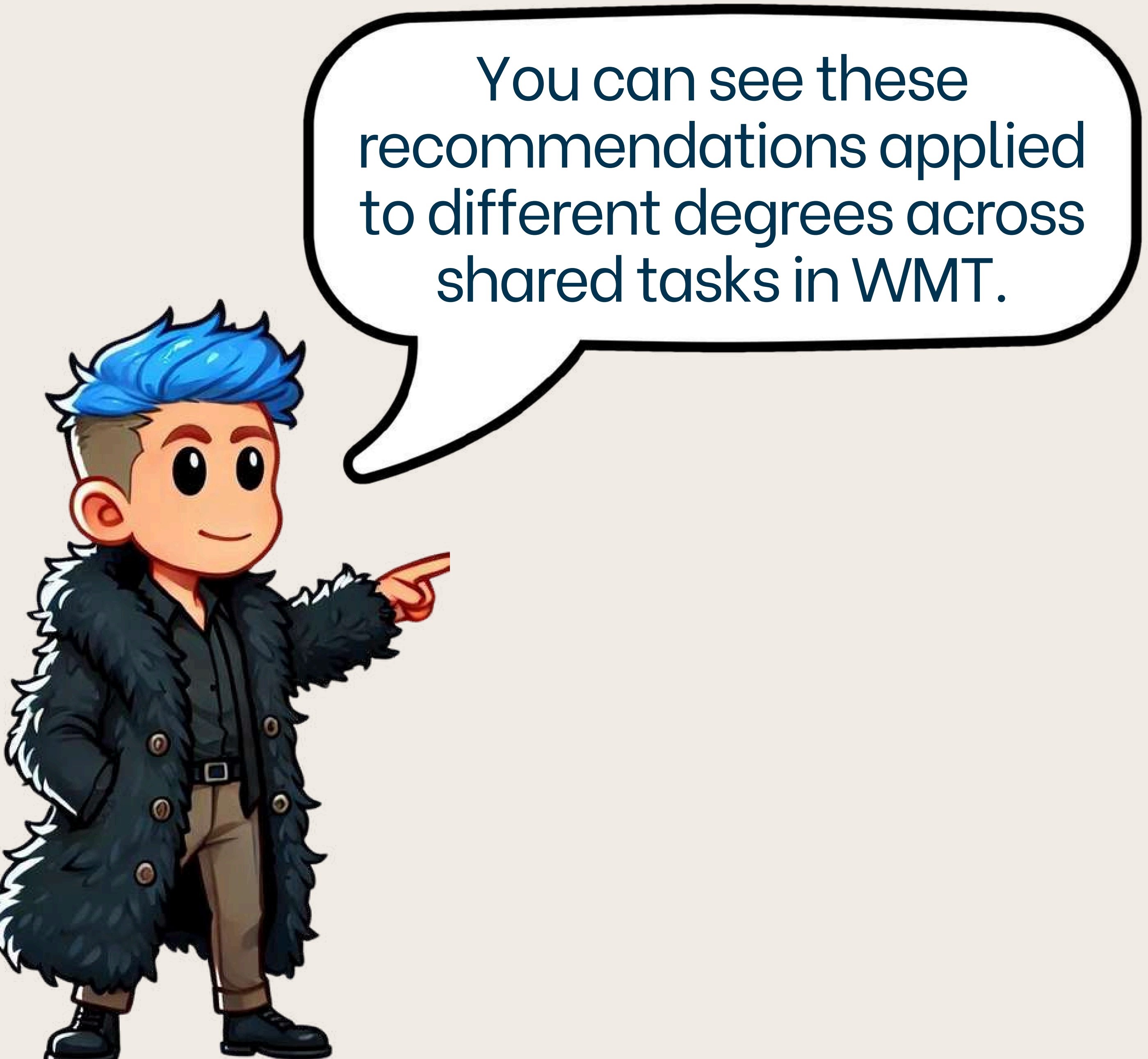
Diptesh Kanojia
University of Surrey
d.kanojia@surrey.ac.uk

Matteo Negri
Fondazione Bruno Kessler
negri@fbk.eu

Marco Turchi
Zoom Video Communications
marco.turchi@zoom.us



Why Evaluation Design Matters



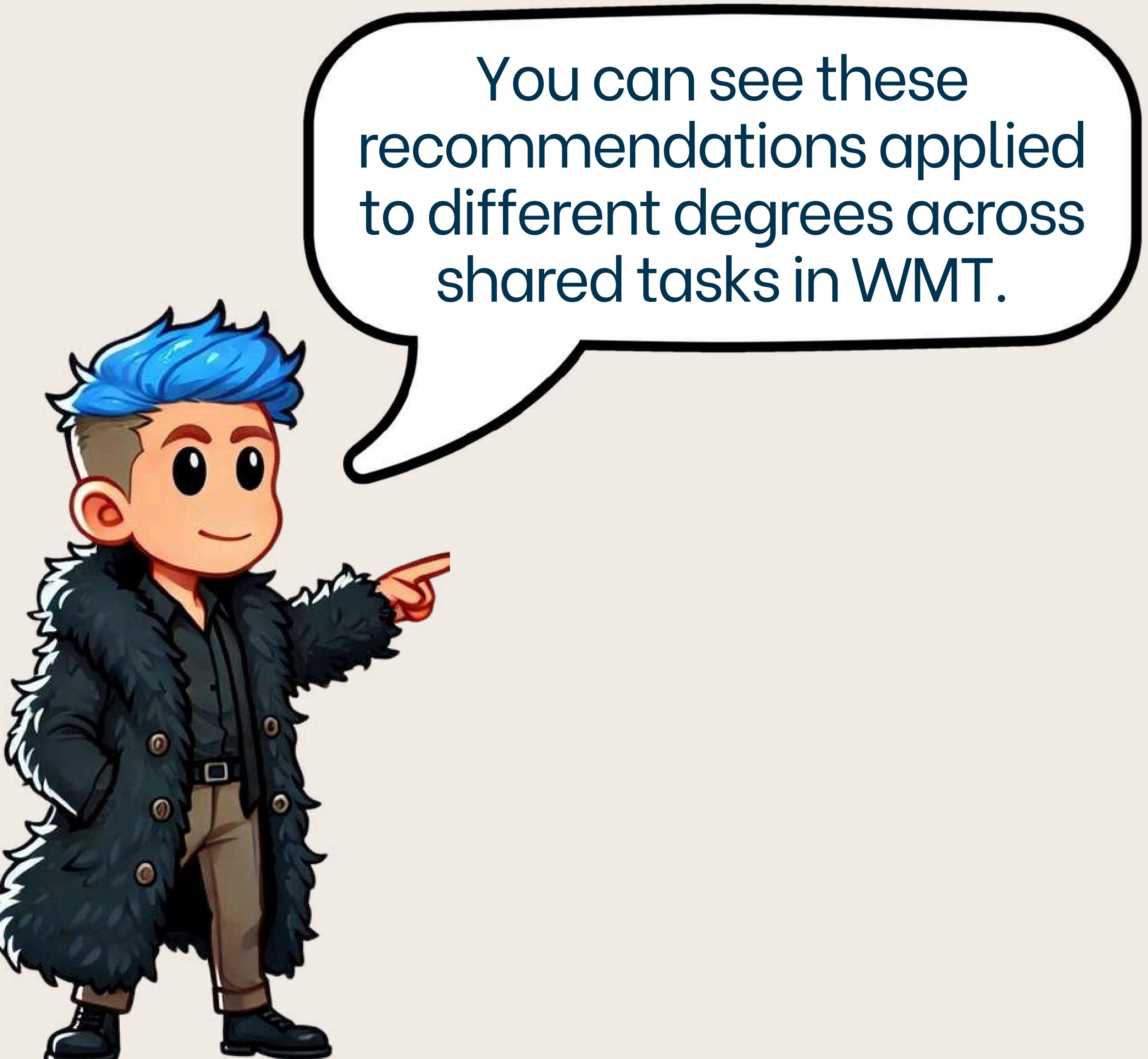
and the APE system output. We hired 4 translators to evaluate the two primary system submissions (KU_UP & KAISTAI), manually post-edited segments (*test.pe*), and the MT Output (*test.mt*). We chose to allocate an equal number of instances to each translator after shuffling, and only a single DA annotation was collected for each instance (Toral, 2020). Shuffling the instances before allocation helps prevent annotator bias towards a single system in the direct assessments.

Why Evaluation Design Matters

You can see these recommendations applied to different degrees across shared tasks in WMT.

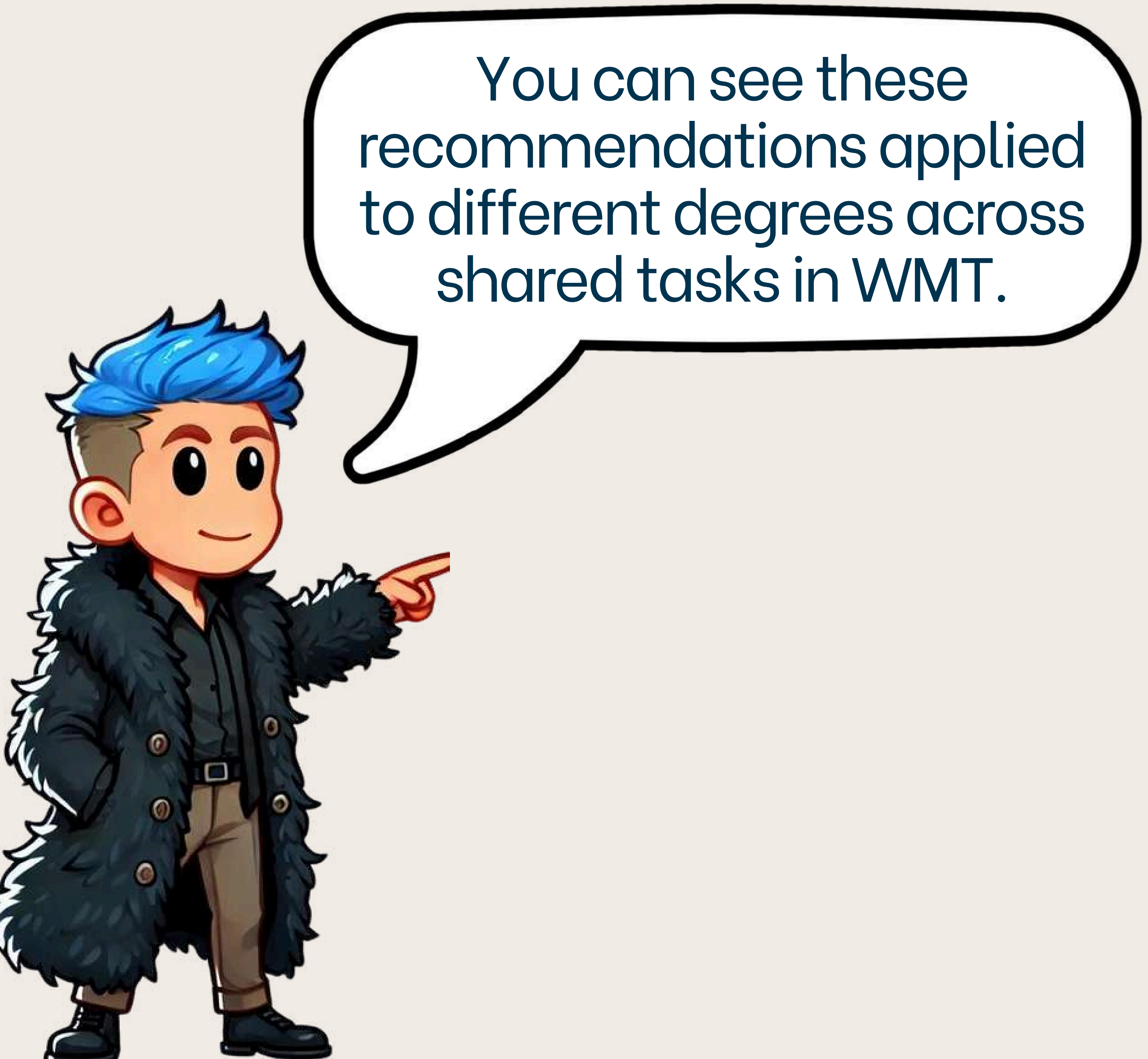
and the APE system output. We hired 4 translators to evaluate the two primary system submissions (KU_UP & KAISTAI), manually post-edited segments (*test.pe*), and the MT Output (*test.mt*). We chose to allocate an equal number of instances to each translator after shuffling, and only a single DA annotation was collected for each instance (Toral, 2020). Shuffling the instances before allocation helps prevent annotator bias towards a single system in the direct assessments.

Why Evaluation Design Matters



The annotation guidelines provide a detailed description of potential adequacy and fluency-based errors based on which the translator could estimate the direct assessment score range. However, the translators were additionally instructed to prioritize adequacy errors and focus on assessing the semantic similarity between the source and the system output. The annotators manually entered the DA score between 0-100. The collected DA annotations were unshuffled based on the segment IDs, which were unknown to the translators. We expected the human post-editing to be of higher quality compared to APE system submissions and, consequently, better than the MT baseline.

Why Evaluation Design Matters



The annotation guidelines provide a detailed description of potential adequacy and fluency-based errors based on which the translator could estimate the direct assessment score range. However, the translators were additionally instructed to prioritize adequacy errors and focus on assessing the semantic similarity between the source and the system output. The annotators manually entered the DA score between 0-100. The collected DA annotations were unshuffled based on the segment IDs, which were unknown to the translators. We expected the human post-editing to be of higher quality compared to APE system submissions and, consequently, better than the MT baseline.

Why Evaluation Design Matters

Another example

Findings of the 2023 Conference on Machine Translation (WMT23): LLMs Are Here But Not Quite There Yet

Tom Kocmi Microsoft
Eleftherios Avramidis DFKI
Rachel Bawden Inria, Paris
Ondřej Bojar Charles University
Anton Dvorkovich Dubformer
Christian Federmann Microsoft
Mark Fishel University of Tartu
Markus Freitag Google
Thamme Gowda Microsoft
Roman Grundkiewicz Microsoft
Barry Haddow University of Edinburgh
Philipp Koehn Johns Hopkins University
Benjamin Marie 4i.ai
Christof Monz University of Amsterdam
Makoto Morishita NTT
Kenton Murray Johns Hopkins University
Masaaki Nagata NTT
Toshiaki Nakazawa University of Tokyo
Martin Popel Charles University
Maja Popović Dublin City University
Mariya Shmatova Dubformer
Jun Suzuki Tohoku University

Why Evaluation Design Matters

Another example

2.2 Human preprocessing of test data

Although testing of robustness of MT is an important task, the noisy data introduces problems for human translators and annotators. Therefore, we decided to discard data considered too noisy. Furthermore, publicly available data often contains inappropriate content, which can stress either human translators or human annotators, leading to a decrease in the quality (for example, translators refuse to translate political content considered censored in their countries).

Therefore, we asked humans to check collected data and carry out minor corrections (mainly checking sentence splits and discarding similar or repeated content). This was sufficient for the news domain because it was often clean and without serious problems. However, with the expansion towards general MT, we find ourselves running into an issue of source data being noisier and less well formatted and that therefore needs to be handled before translation. Furthermore, we asked them to remove shortest documents to keep longer context. The source data for test sets therefore goes through

Why Evaluation Design Matters

Another example

2.2 Human preprocessing of test data

Although testing of robustness of MT is an important task, the noisy data introduces problems for human translators and annotators. Therefore, we decided to discard data considered too noisy. Furthermore, publicly available data often contains inappropriate content, which can stress either human translators or human annotators, leading to a decrease in the quality (for example, translators refuse to translate political content considered censored in their countries).

Therefore, we asked humans to check collected data and carry out minor corrections (mainly checking sentence splits and discarding similar or repeated content). This was sufficient for the news domain because it was often clean and without serious problems. However, with the expansion towards general MT, we find ourselves running into an issue of source data being noisier and less well formatted and that therefore needs to be handled before translation. Furthermore, we asked them to remove shortest documents to keep longer context. The source data for test sets therefore goes through

Why Evaluation Design Matters

Another example

5 Human Evaluation

Human evaluation for all language translation directions is performed with source-based ("bilingual") Direct Assessment (DA, Graham et al., 2013) of individual segments in document context with Scalar Quality Metrics (SQM) guidelines, mostly following the setup established at WMT22 (DA+SQM, Kocmi et al., 2022). DA+SQM asks the annotators to provide a score between 0 and 100 on a sliding scale, but the slider is presented with seven labelled tick marks, as demonstrated in Figure 1.

Two different annotation platforms and four distinct pools of annotators (Table 3) are used for annotation of different language pairs. We use the open-source framework Appraise (Federmann, 2018) for the evaluation of English$\rightarrow$Czech, English$\leftrightarrow${Chinese, German, Japanese}, and Czech$\rightarrow$Ukrainian. Toloka AI[21] hosts the evaluation of English$\leftrightarrow${Hebrew, Russian, Ukrainian} using their own implementation of the source-based

Why Evaluation Design Matters

Another example

5 Human Evaluation

Human evaluation for all language translation directions is performed with source-based ("bilingual") Direct Assessment (DA, Graham et al., 2013) of individual segments in document context with Scalar Quality Metrics (SQM) guidelines, mostly following the setup established at WMT22 (DA+SQM, Kocmi et al., 2022). DA+SQM asks the annotators to provide a score between 0 and 100 on a sliding scale, but the slider is presented with seven labelled tick marks, as demonstrated in Figure 1.

Two different annotation platforms and four distinct pools of annotators (Table 3) are used for annotation of different language pairs. We use the open-source framework Appraise (Federmann, 2018) for the evaluation of English→Czech, English↔{Chinese, German, Japanese}, and Czech→Ukrainian. Toloka AI[21] hosts the evaluation of English↔{Hebrew, Russian, Ukrainian} using their own implementation of the source-based

Why Evaluation Design Matters

Another example

5.1 Human annotators

Annotations for different language pairs are provided by four different parties with their pool of annotators of distinct profiles as presented in Table 3. We shift towards more professional or semi-professional annotators' pools and decide not to use MTurk annotations as in past years for reference-based DA evaluation for into-English language directions.

Assessments for English↔{Chinese, German, Japanese} are provided by Microsoft and their pool of bilingual target-language native speakers, professional translators or linguists, highly experienced in MT evaluation. Microsoft monitors the annotators' performance over time and permanently removes from the pool those who fail quality control, which increases the overall quality of the human assessment.

Why Evaluation Design Matters

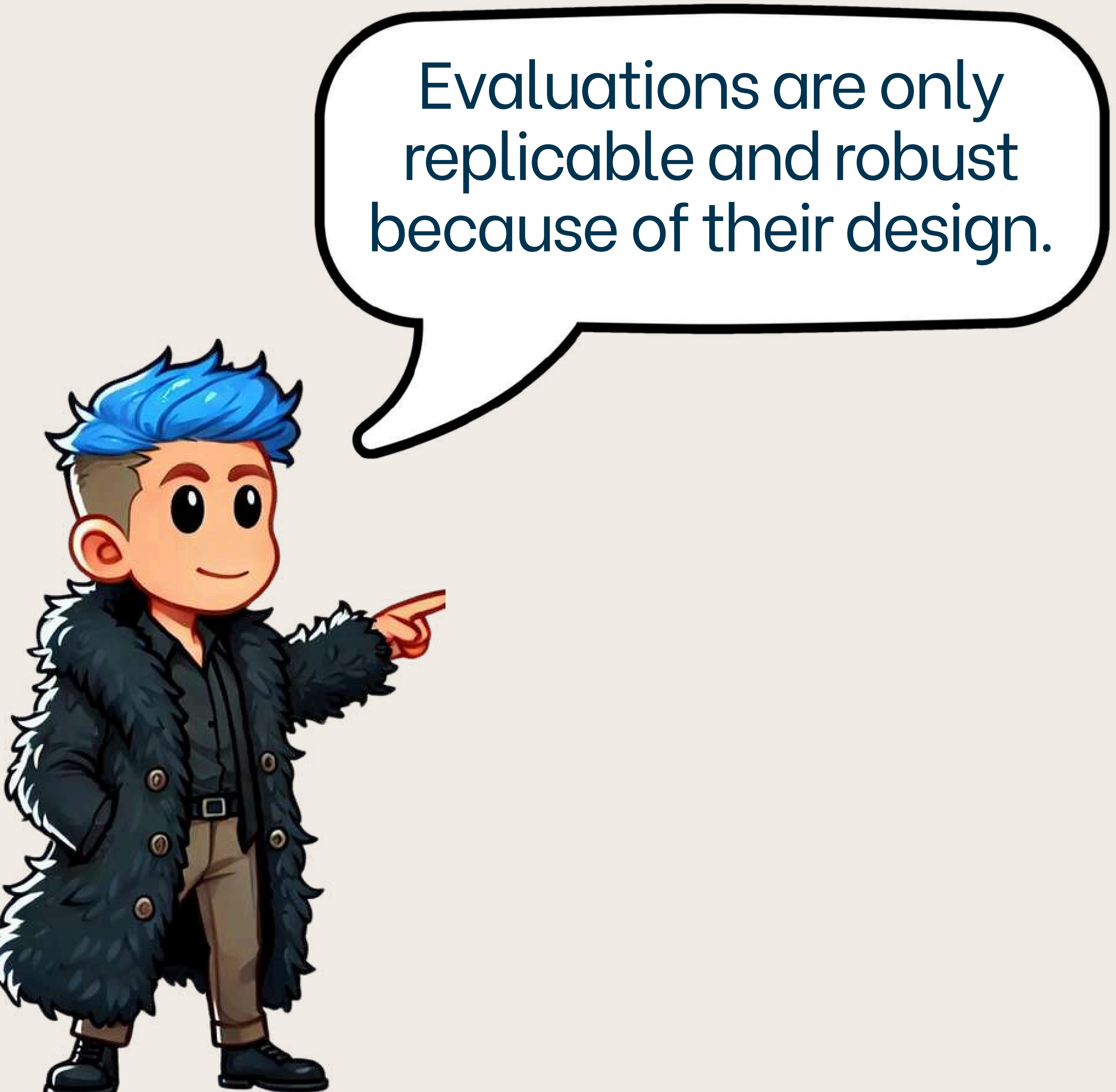
Another example

5.1 Human annotators

Annotations for different language pairs are provided by four different parties with their pool of annotators of distinct profiles as presented in Table 3. We shift towards more professional or semi-professional annotators' pools and decide not to use MTurk annotations as in past years for reference-based DA evaluation for into-English language directions.

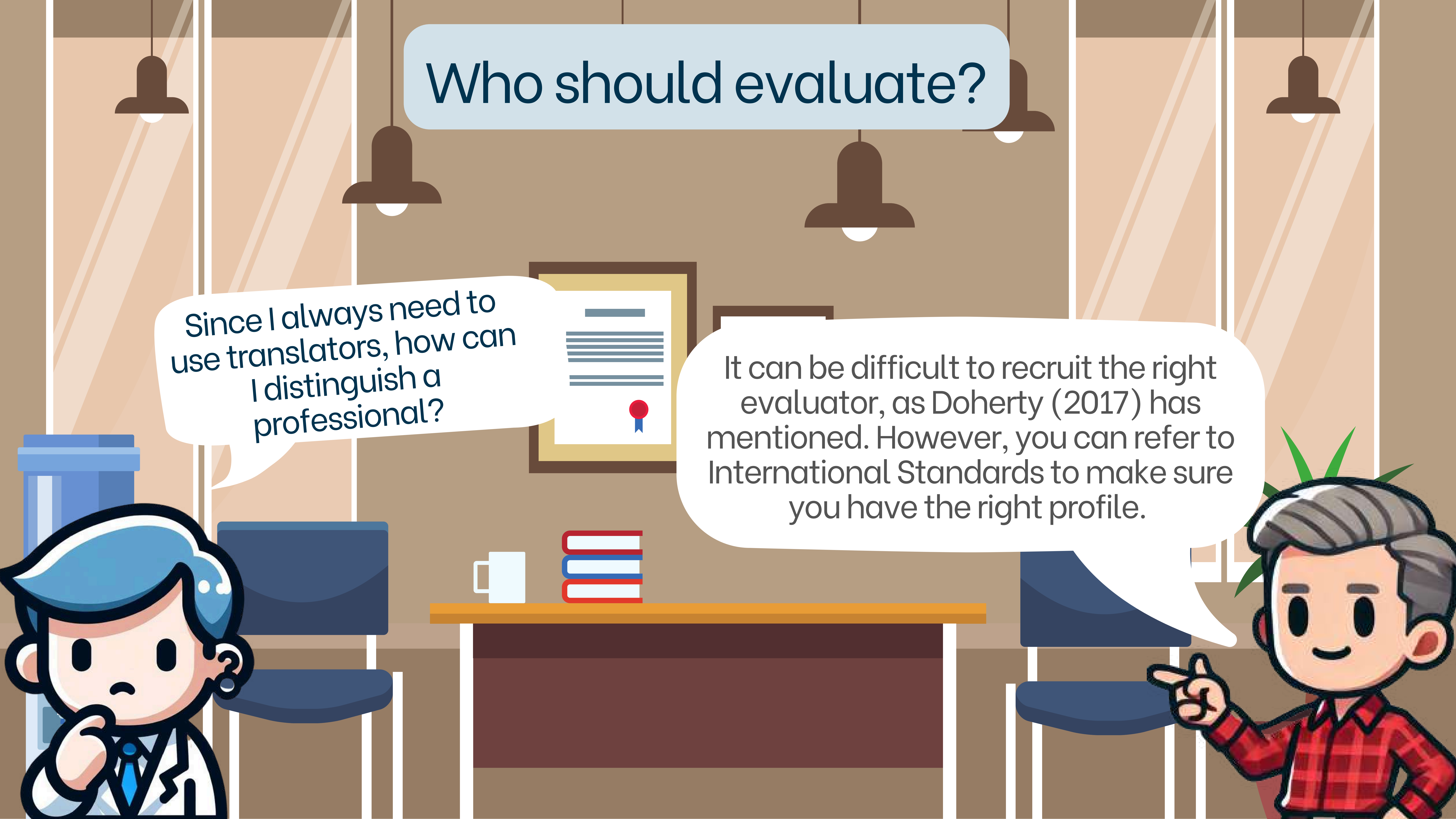
Assessments for English↔{Chinese, German, Japanese} are provided by Microsoft and their pool of bilingual target-language native speakers, professional translators or linguists, highly experienced in MT evaluation. Microsoft monitors the annotators' performance over time and permanently removes from the pool those who fail quality control, which increases the overall quality of the human assessment.

Why Evaluation Design Matters



- Purpose of MT system
- Test suites
- Evaluators
- Likert Scales
- Risk assessment
- Evaluation platforms
- Guidelines

Who should evaluate?



Who should evaluate? - International Standards

ISO 17100 Translation Services

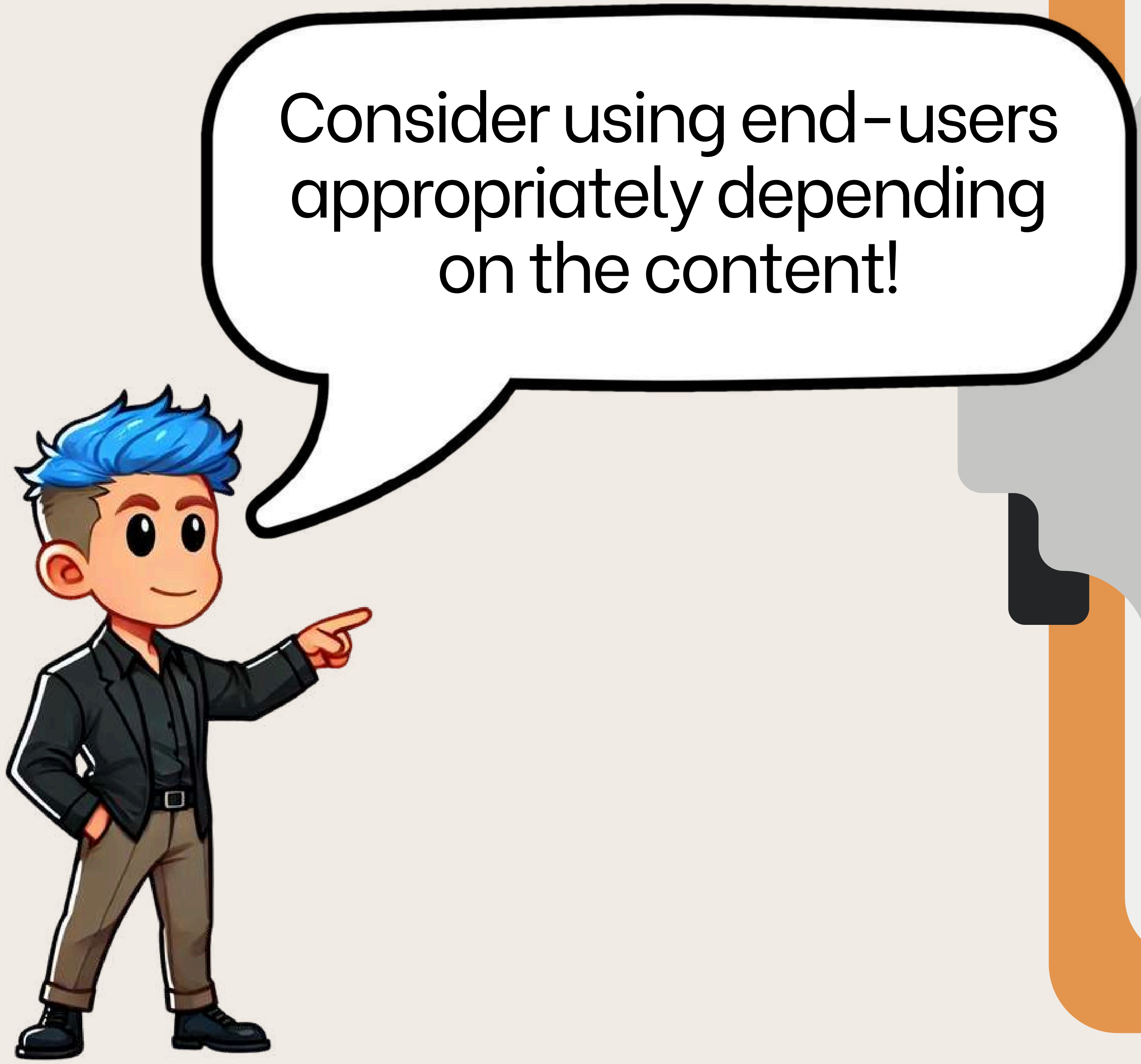
- a) Has obtained a degree in translation, linguistics or language studies or an equivalent degree that includes significant translation training, from a recognised institution of higher education
- b) Has obtained a degree in any other field from a recognised institution of higher education and has the equivalent of two years of full-time professional experience in translating;
- c) Has the equivalent of five years of full-time professional experience in translating.

Who should evaluate? - International Standards

ASTM F2475-23

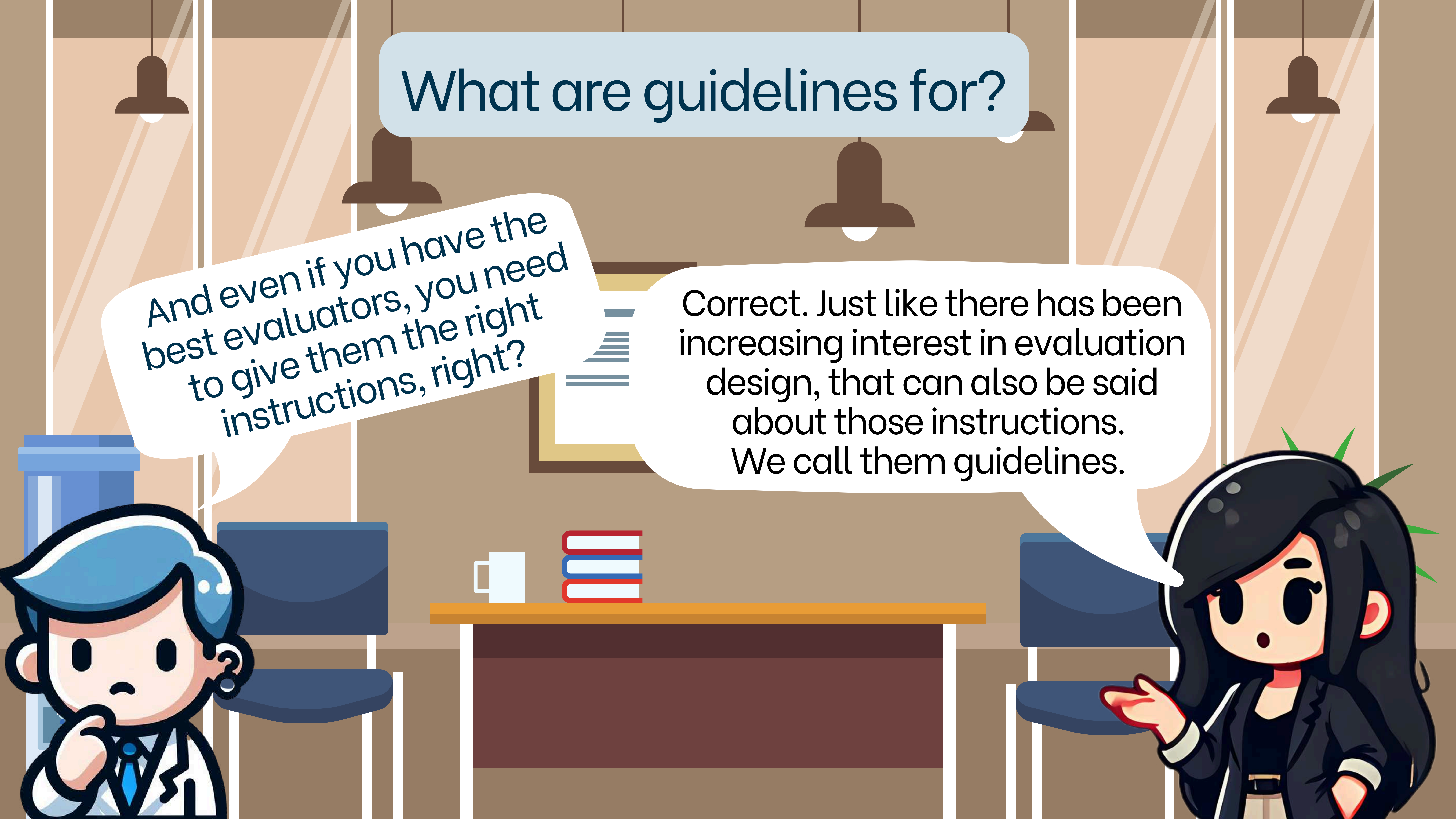
- 3.1.21 subject matter expert, n—person responsible for conducting a monolingual review of the target text to ensure domain accuracy and appropriateness of terminology and cultural nuances in the target language.
- 3.1.27 third-party evaluator, n—content expert consulted for their feedback on the finalized translation.
- 3.1.27.1 Discussion—Third-party evaluators should have similar credentials to the translator.

Who should evaluate?



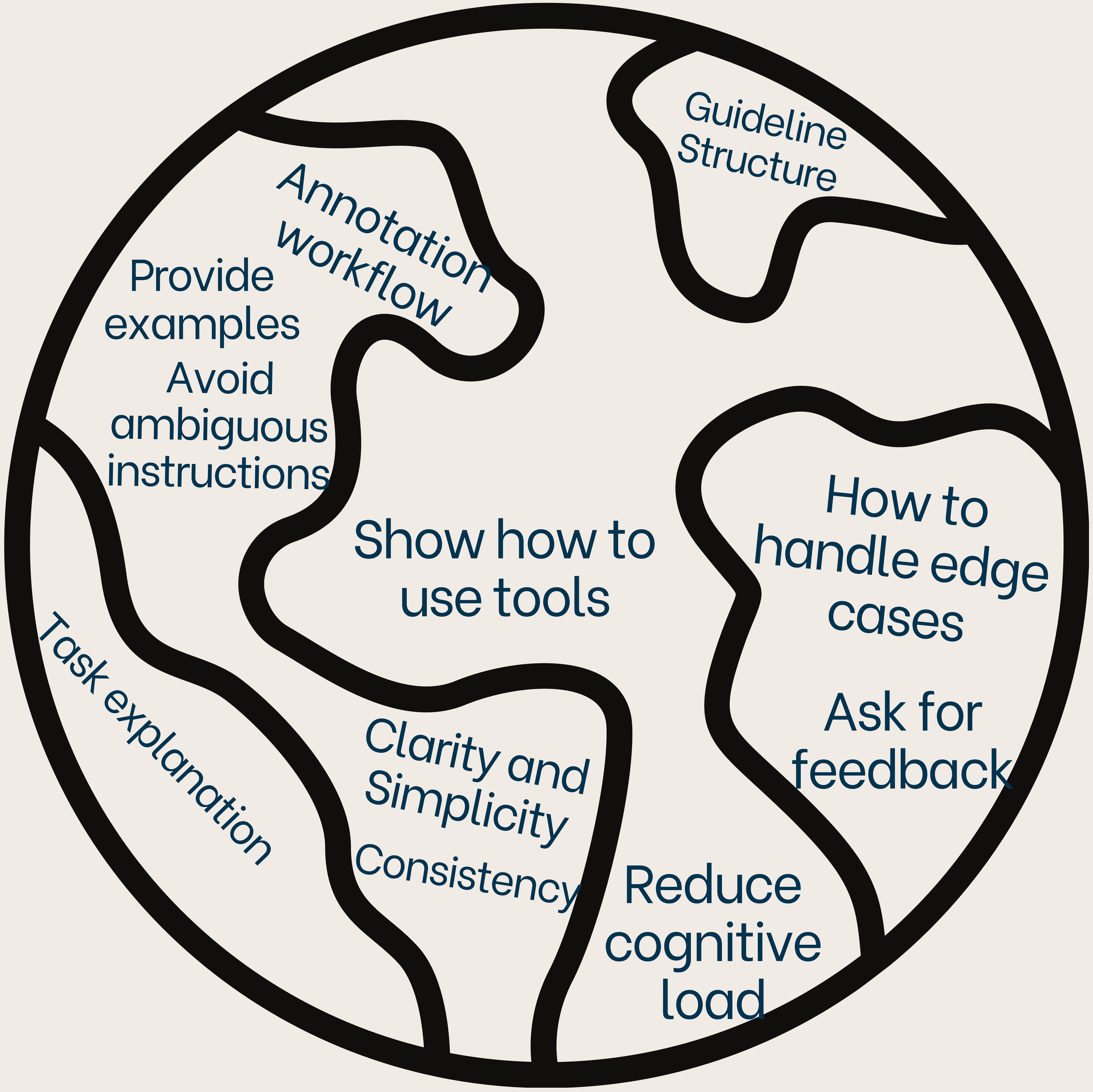
- **Expert end user** - What expertise will they be using and what aspects of translation quality assessment are tasked?

- **Non-expert end user** - If there is no expertise, what kind of product is your user evaluating? Will you emphasise the usability of the product with your desired MT system?


What are guidelines for?
And even if you have the best evaluators, you need to give them the right instructions, right?
Correct. Just like there has been increasing interest in evaluation design, that can also be said about those instructions. We call them guidelines.

What are guidelines for?

How do you create guidelines? You "planet"!



What are guidelines for?

Contributions from Knowles and Lo (2024)

Challenges

- Variability between evaluators
- Strict evaluators vs Lenient evaluators
- Lack of clear instructions on how to handle ambiguous/ incomplete translations leads to different interpretations
- User interfaces may lead to physical factors, such as sensitivity.

Context

- Intersequential context is relevant. Translations require context for accurate assessment to consider tense, number or gender.
- Consider calibration sets.

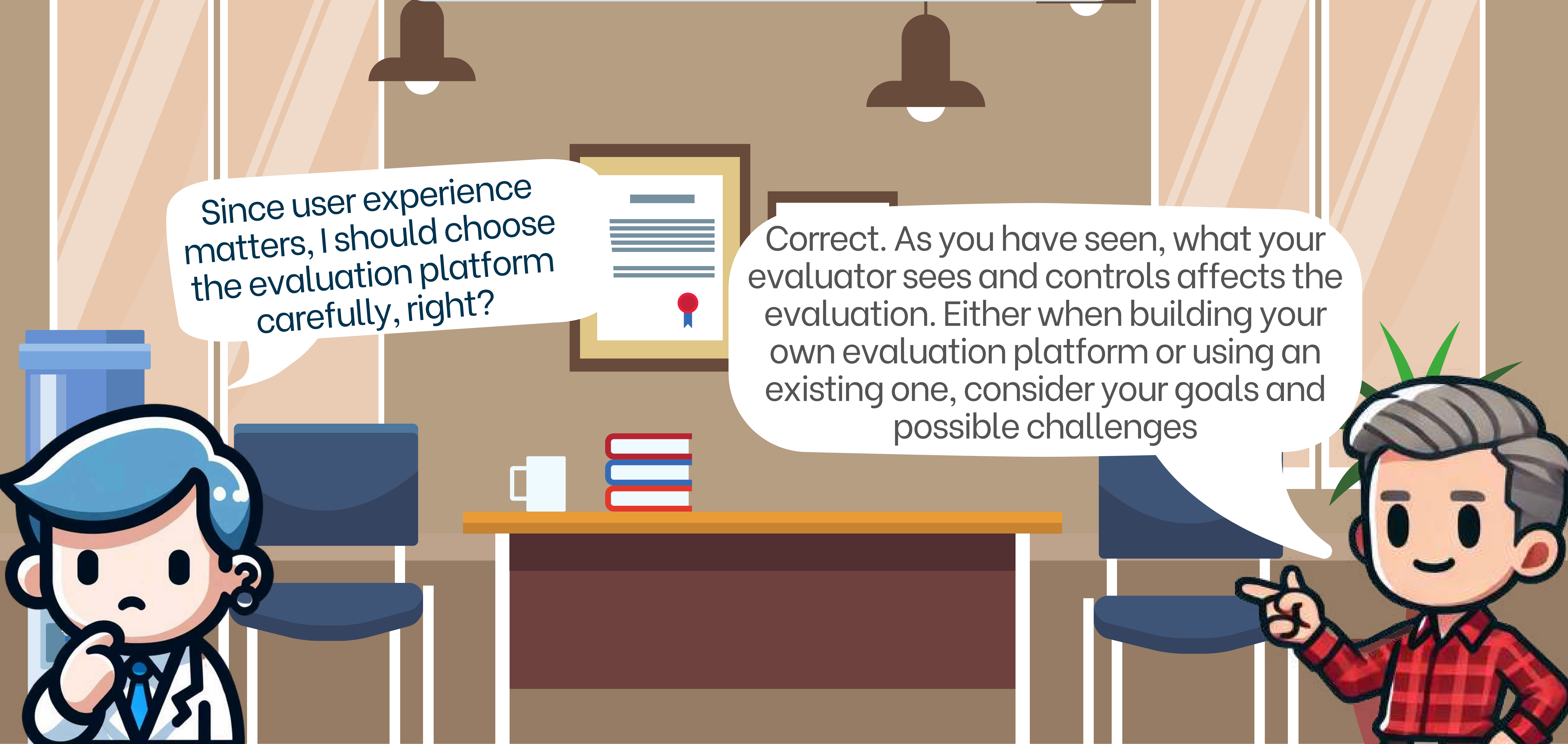
Variation and calibration

- Self-consistency can also be an issue: same annotator might score segments differently.
- Fatigue or a shift of focus can lead to degradation of consistency.
- Sliders can affect inconsistencies.

Recommendations

1. **Include context.** Protocols must provide sufficient context (preceding and following)
2. **Include calibration tasks.** Use a controlled set of translations for training and consistency checks.
3. **Provide a user friendly Interface.** Adjust the sensitivity of sliders to prevent inadequate scoring due to small hand movements.
4. **Balance Annotator workload.** More context leads to more fatigue. Consider more sessions that are shorter to maintain the quality of the annotation.

Where should I evaluate?



Where should I evaluate?

Crowdsourcing Graphical Perception: Using mechanical turk to assess visualisation design

Heer and Bostock (2010)

- Researchers looked into the perception of annotators in regards to graphics.
- They found that position, length, and color are effective visual encodings for conveying quantitative information.
- Luminance constrast demonstrated improved legibility.
- Position-based elements like sliders are important for precise user input

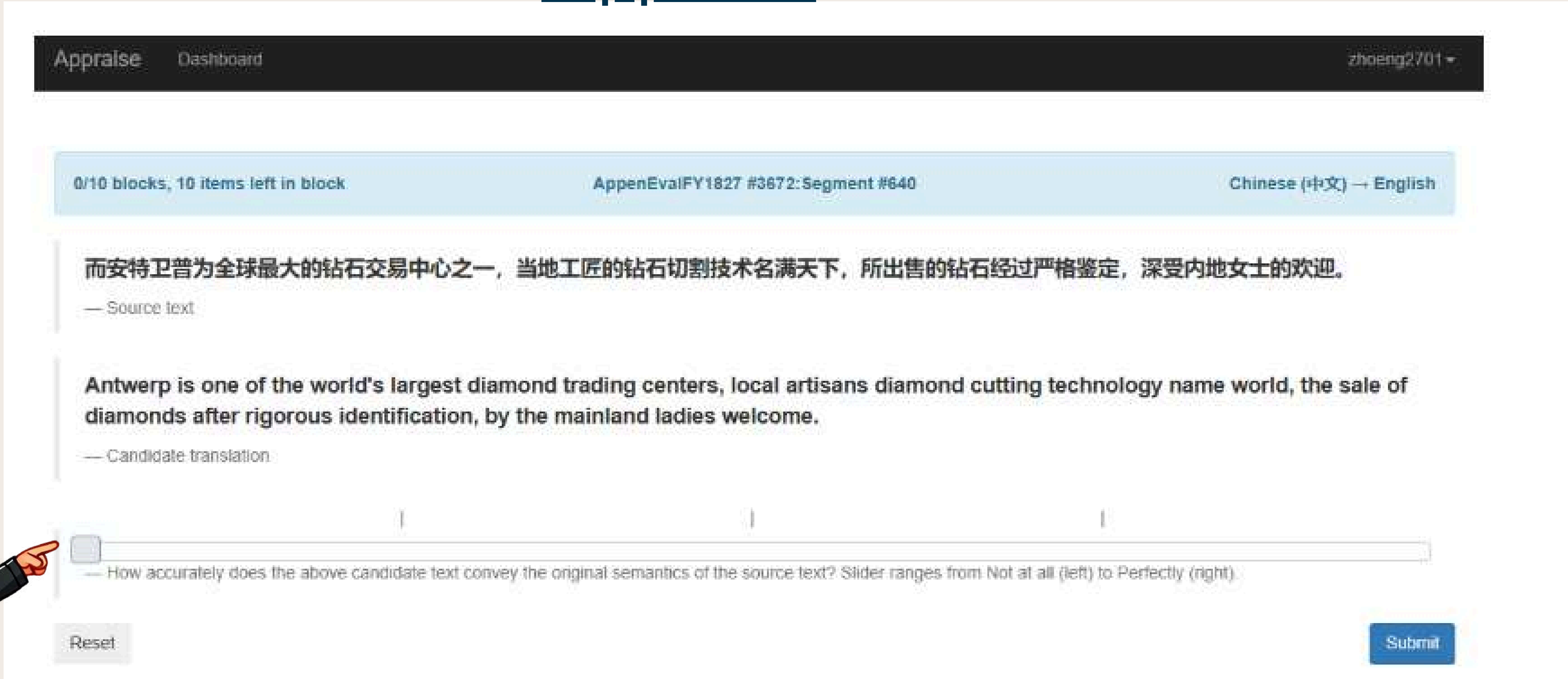
Where should I evaluate?

The impact of traditional and interactive post-editing on Machine Translation User Experience, quality, and productivity

Briva-Iglesias et al. (2023)

- Researchers proposed MT user experience.
- They measured how users feel about the system, including efficiency, control, attractiveness, and stimulation.
- Interactive post-editing led to higher user satisfaction and user experience.
- When creating your own evaluation platform, consider these aspects from evaluators. Consider user control in the interface as necessary.

Where should I evaluate? [Appraise]



Federmann (2018)

Where should I evaluate?
KantanAI

Warning! Please select value or fill all compulsory KPIs ✕

Samples	Fluency*	Adequacy*	Ranking
Para ser claros, se necesitan barandillas. Pero deberían aplicarse a las aplicaciones de IA, no a la tecnología de IA de uso general.	★★★★☆	★★★★☆	★★☆☆
Para ser claros, se necesitan medidas de protección. Pero estas deben aplicarse a las aplicaciones de IA, no a la tecnología de IA de uso general.	★★★★★	★★★★★	★★★★
Para ser claro, se necesitan salvaguardias. Pero deben aplicarse a las plicaciones de IA, no a la tecnología de IA de propósito general.	★★★★★	★★★★★	★★★☆
Para que quede bien claro, se necesitan salvaguardas. Pero se deben aplicar a la licaciones de inteligencia artificial, no a la tecnología de inteligencia icial de propósito general.	★★★★★	★★★★★	★★★☆

Where should I evaluate? Amazon Turk

Rate the quality of a translation by listening to a sample audio and comparing it with reference text in English.

Requester: Sami Ul Haq **Reward:** $0.70 per task **Tasks available:** 36

Qualifications Required: Location is US , HIT Approval Rate (%) for all Requesters' HITs greater than 95

This HIT consists of 100 English assessments. You have completed 3.

Read the text and listen the audio below. Use the slider to indicate how much do you agree with the statement written in blue box?

That is no satisfactory answer and the holding time was not worth it

0:00 0:04

The **audio** adequately expresses the meaning of the **text** written in English.

strongly disagree **strongly agree**

NEXT

Showing Task 1 of 36 Next HIT

Where should I evaluate?
MATEO

- MATEO
- Translate
- Evaluate
- Background
- Visualize

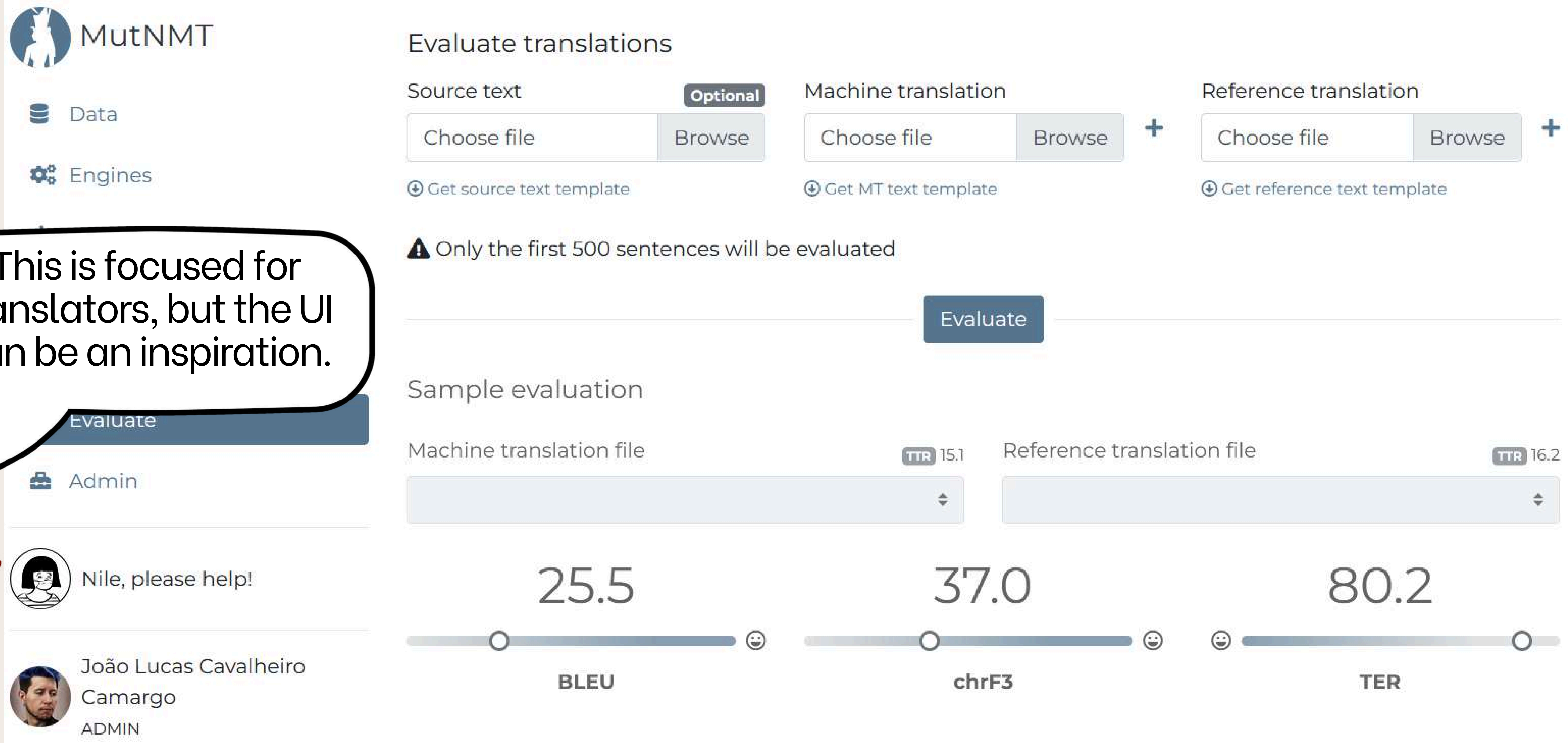
MATEO: MAchine Translation Evaluation Online

v1.1.3

MAchine Translation Evaluation Online (MATEO) brings automatic machine translation evaluation to the masses with an accessible user-interface. It is being developed at Ghent University, in the Language and Translation Technology Team (LT3).

MATEO was built to cater to both experts and non-experts. Users can be system builders, MT users and researchers, and also people from Social Sciences and Humanities (SSH), as well as teachers and students. As such, MATEO can play a crucial role in research *and* education by streamling and simplifying the evaluation aspect of MT research on the one hand and enhancing digital literacy on the other.

Where should I evaluate? MutNMT



What scale should I use?

- Direct Assessment (DA) has been used in shared tasks (Graham et al. 2013)
- Continuous scales of scores 1 - 100
- The decision of Likert scales must be made when deciding the UI of your human evaluation.

What scale should I use?

Pay attention to how your evaluators use scales

- Wide distributions (Scoring anywhere from 0 to 100)
- Narrow distributions (Scoring between 50 and 70, for example)
- Discretisation (evaluators only use the same scores, for example, they may have chosen only 0, 50 and 70)

What scale should I use?

In the guidelines, teach your evaluators how to use scales! Pay attention to:

- How the slider works. (How sensitive it is using a mouse/finger if evaluation is done with a phone)
- Tick marks (how many will there be?)
- Labels (how many will be shown or if labels will be shown)
- Score space (starting position or to show number to evaluator or not)

What scale should I use?

	5. All of it	4. Most of it	3. Some of it	2. Little of it	1. None of it
Adequacy	○	○	○	○	○

No Fluency 0

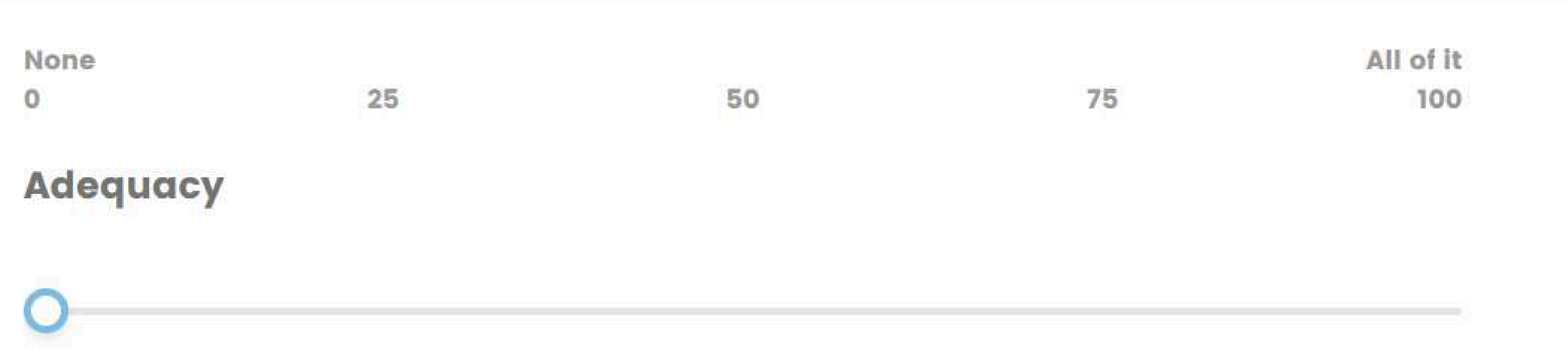
Native 100

Fluency

ADEQUACY
How much of the meaning expressed in the source appears in the translation?

4. All of it
3. Most of it
2. Little of it
1. None of it

What scale should I use?



What scale should I use?

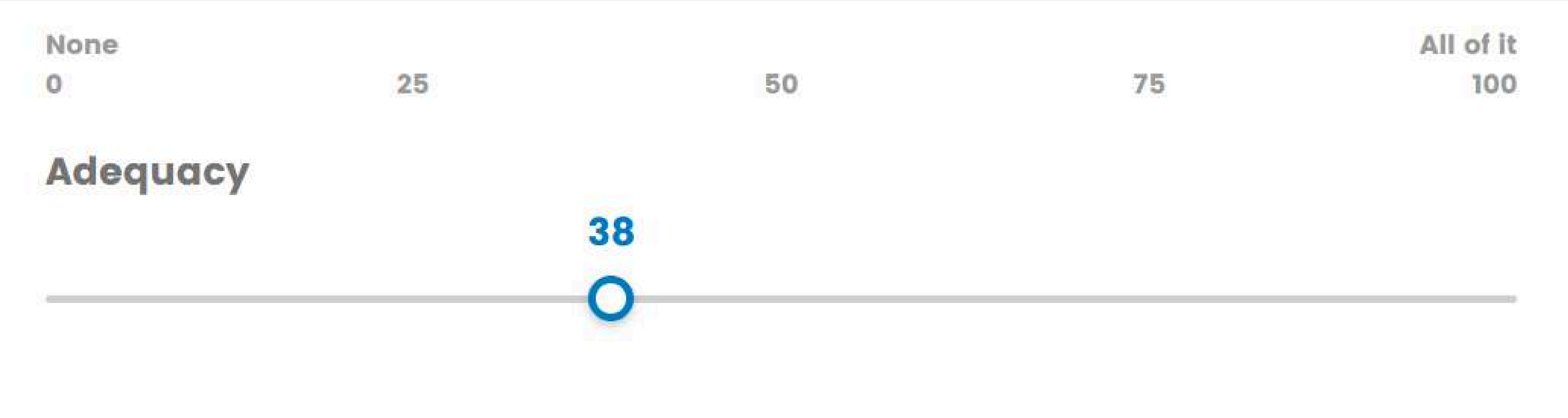
None | Little of it | Some of it | Most of it | All of it
0 | | | | 100

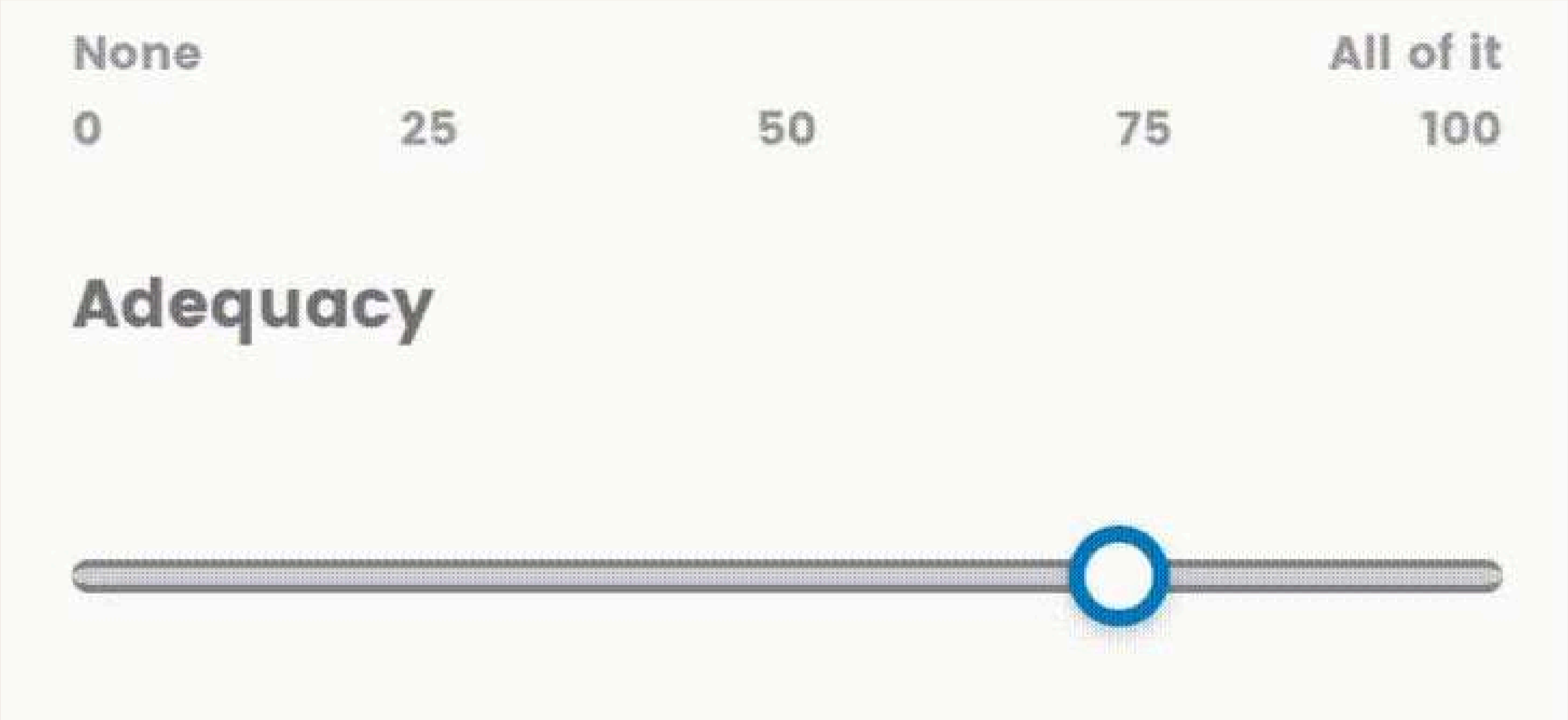
Adequacy

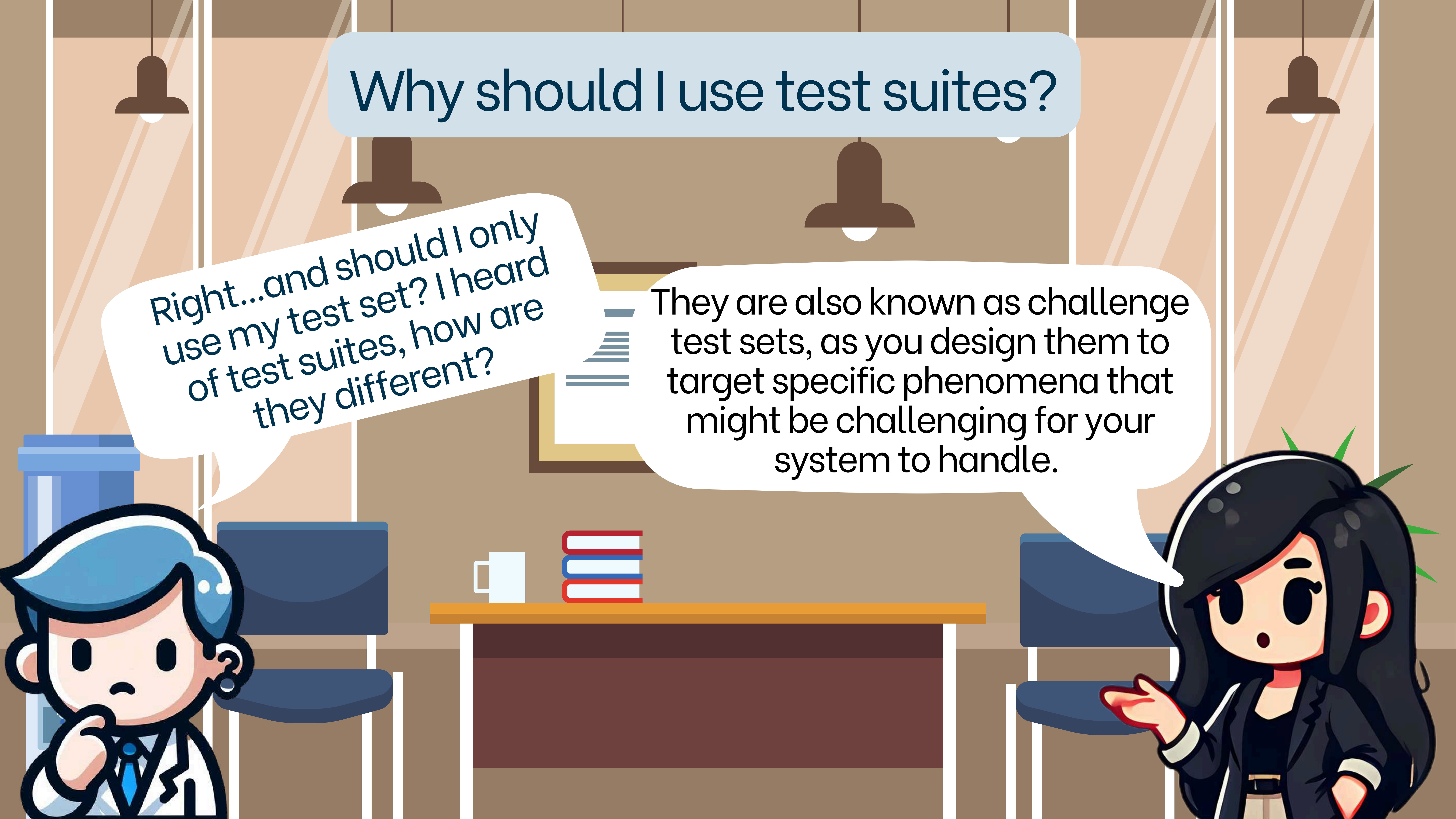
None | | | | All of it
0 | 25 | 50 | 75 | 100

Adequacy

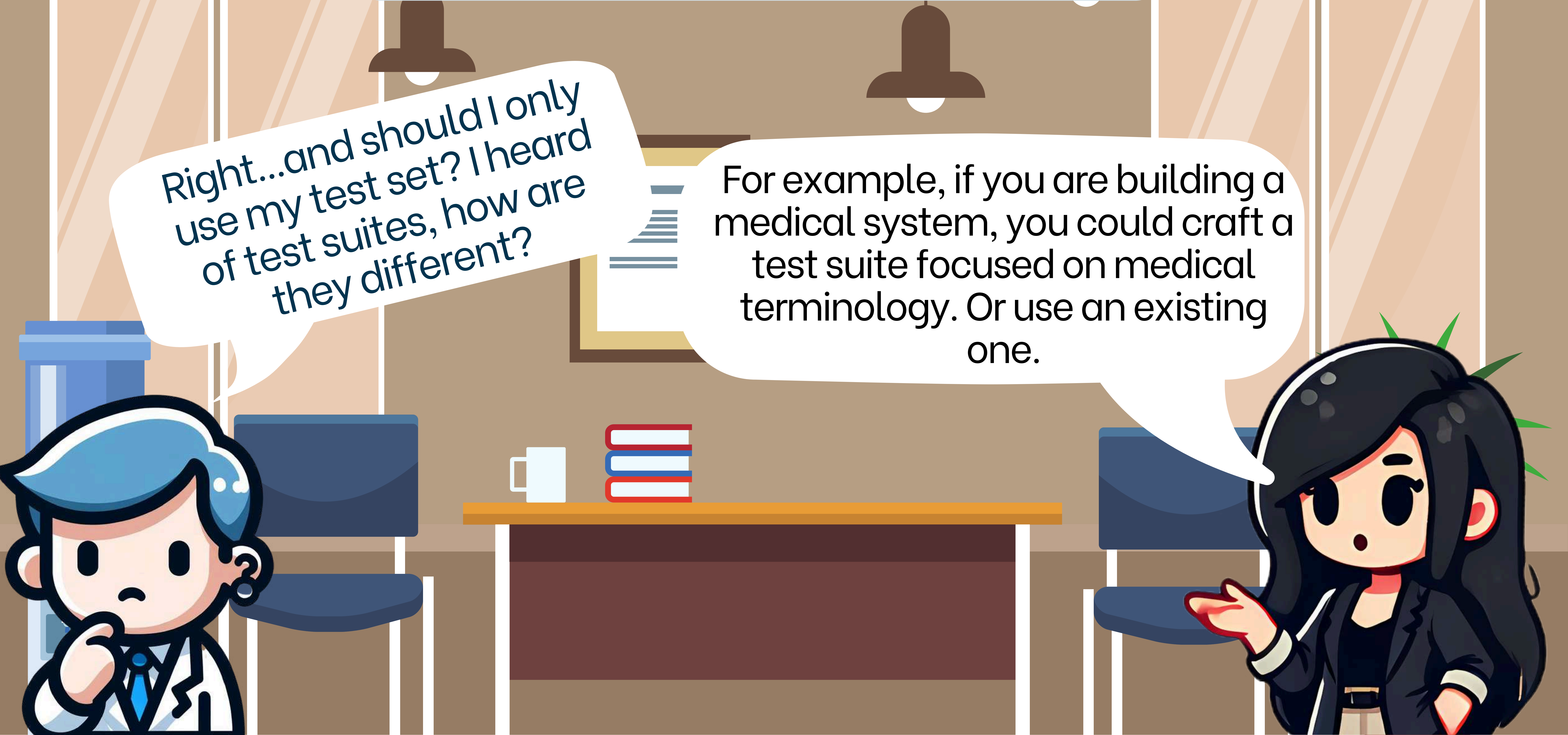
What scale should I use?

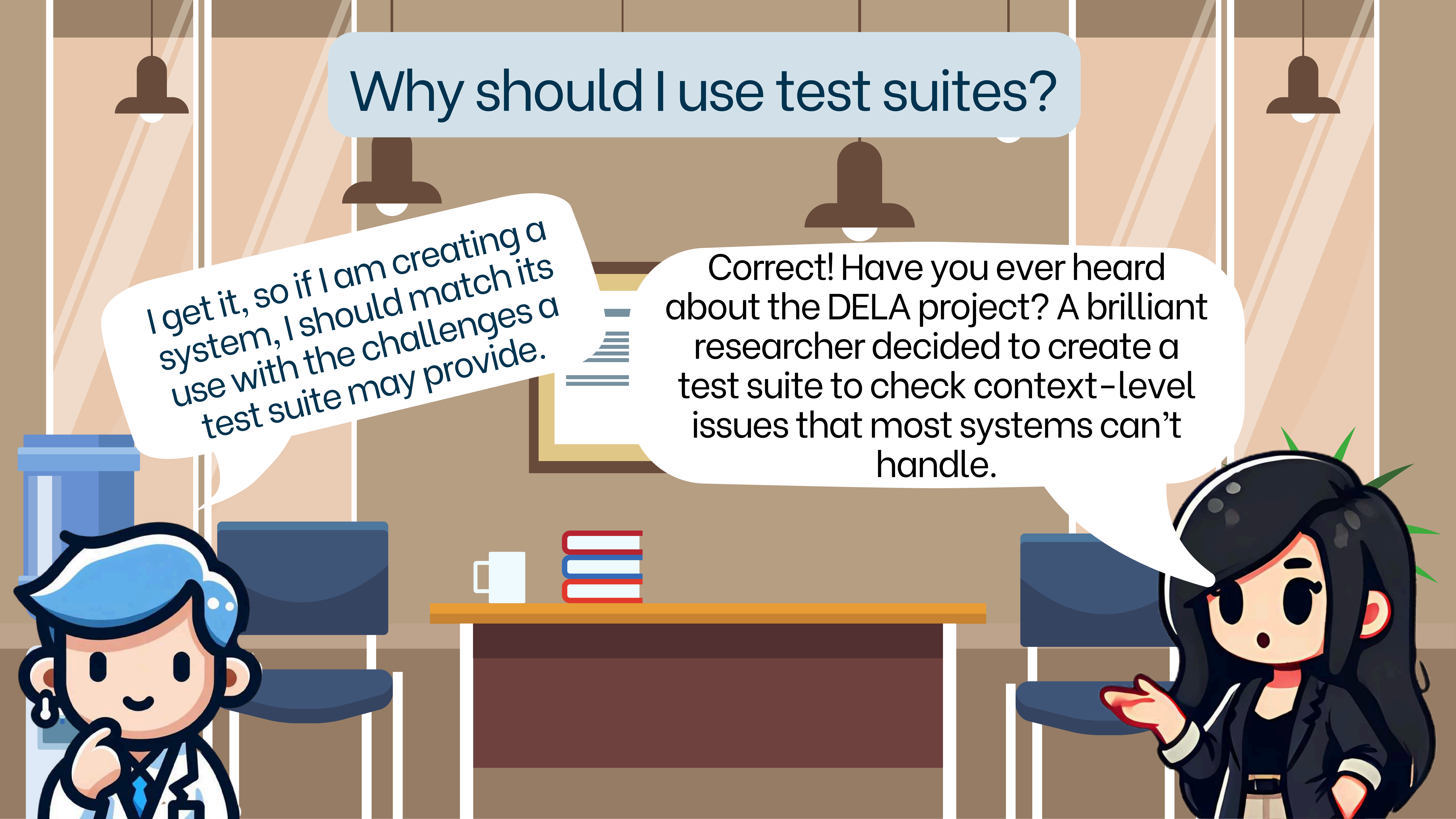





Why should I use test suites?
Right...and should I only use my test set? I heard of test suites, how are they different?
They are also known as challenge test sets, as you design them to target specific phenomena that might be challenging for your system to handle.

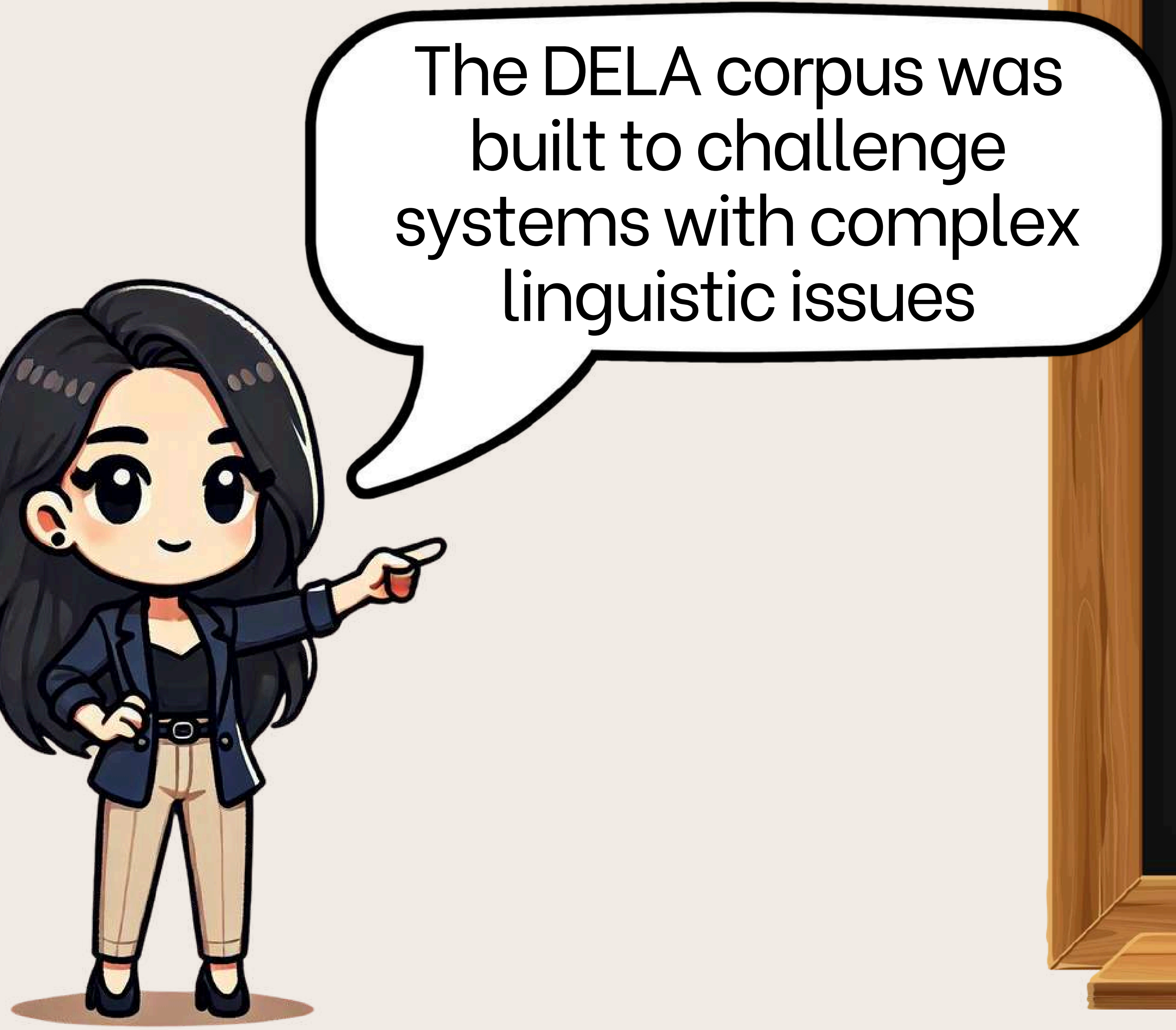
Why should I use test suites?




Why should I use test suites?
I get it, so if I am creating a system, I should match its use with the challenges a test suite may provide.
Correct! Have you ever heard about the DELA project? A brilliant researcher decided to create a test suite to check context-level issues that most systems can't handle.

Test Suites - DELA Corpus

Castilho et al. (2021)

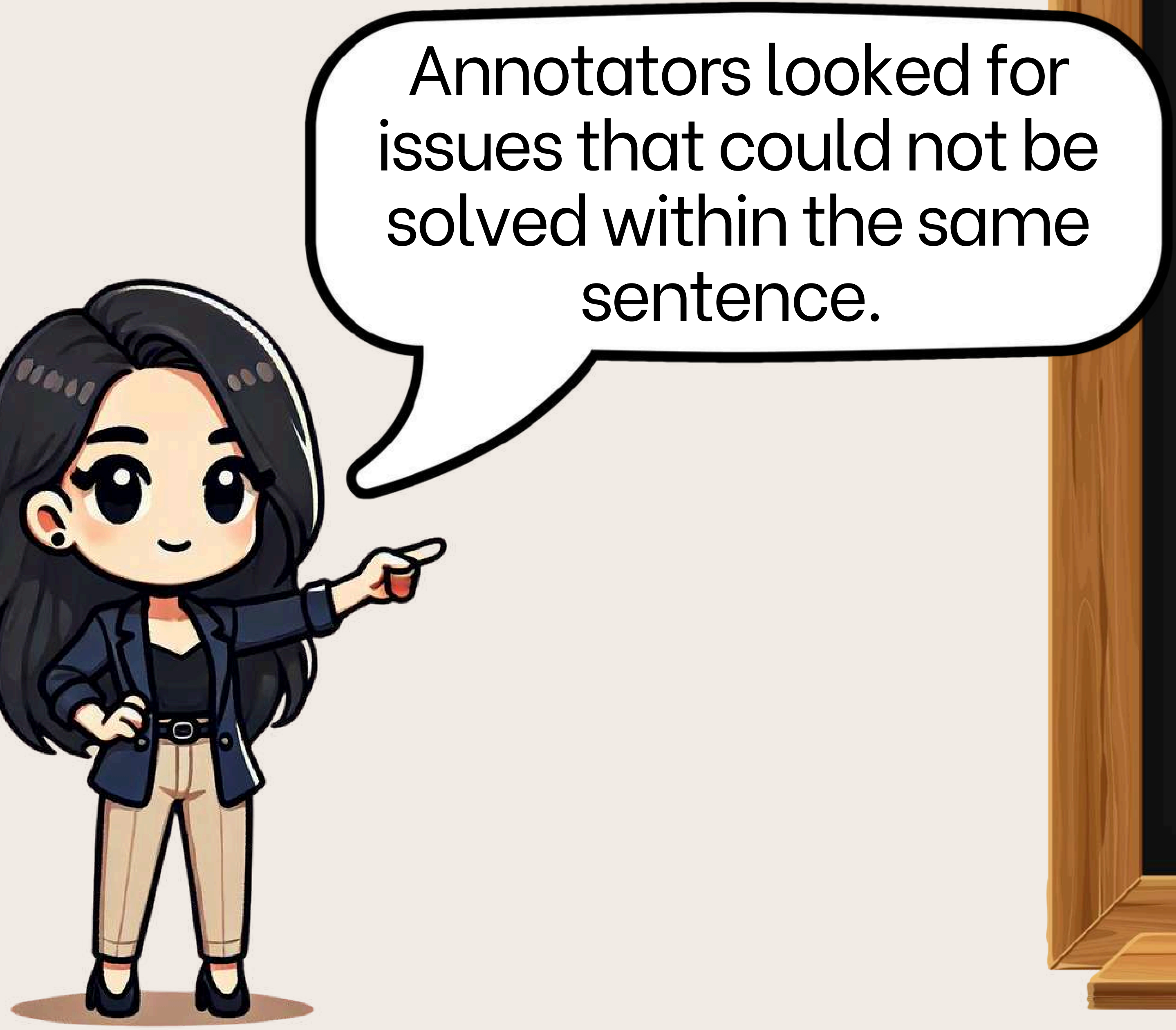


Methodology for the Corpus

- The corpus was collected from a variety of freely available sources.
- A list of context issues found in Castilho et al. (2020) was used for annotators to search for challenging texts.
- 60 full documents (57217 tokens) were collected from six different domains: literary, subtitles, news, reviews, medical and legislation). (p. 3)

Test Suites - DELA Corpus

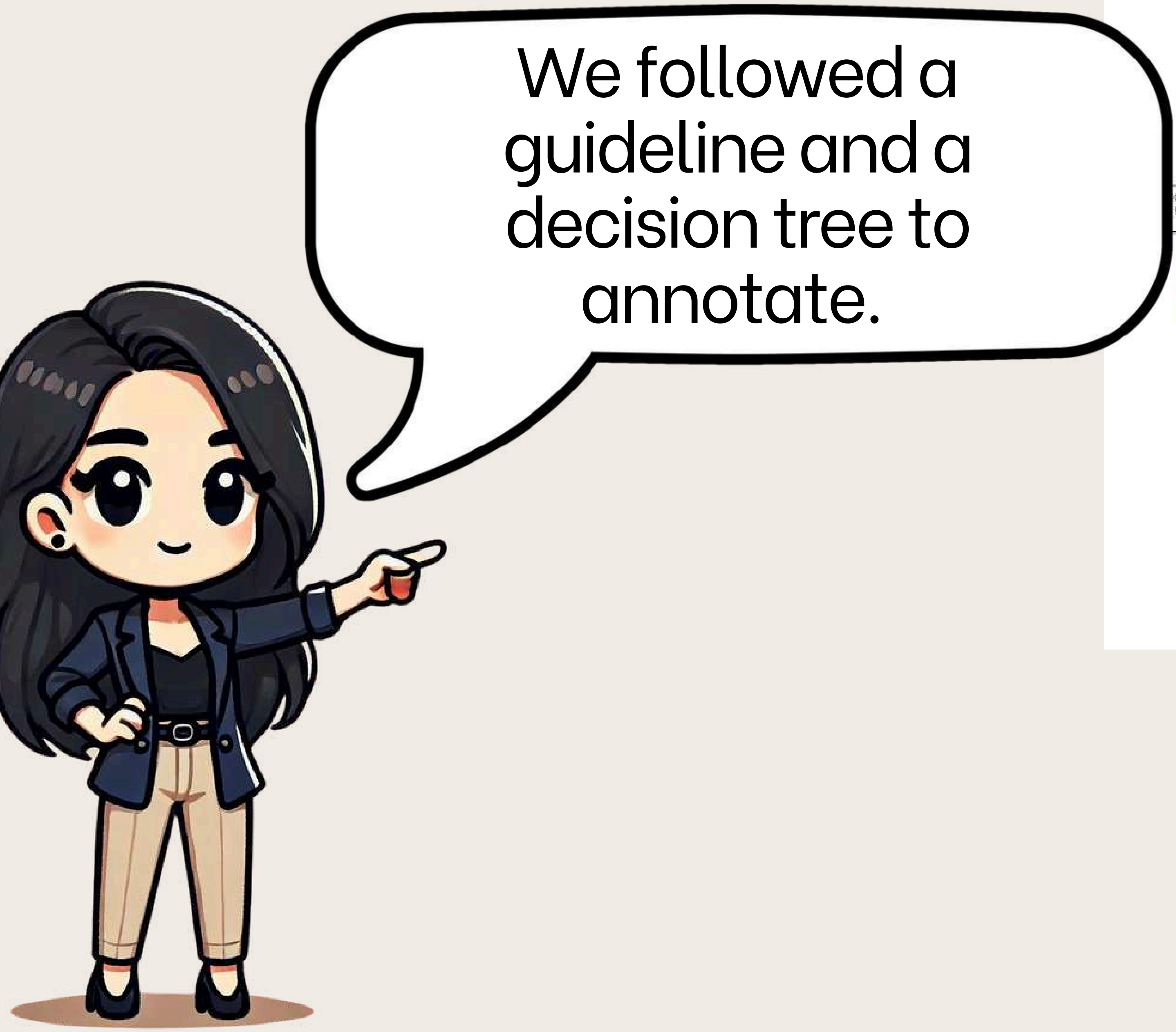
Castilho et al. (2021)

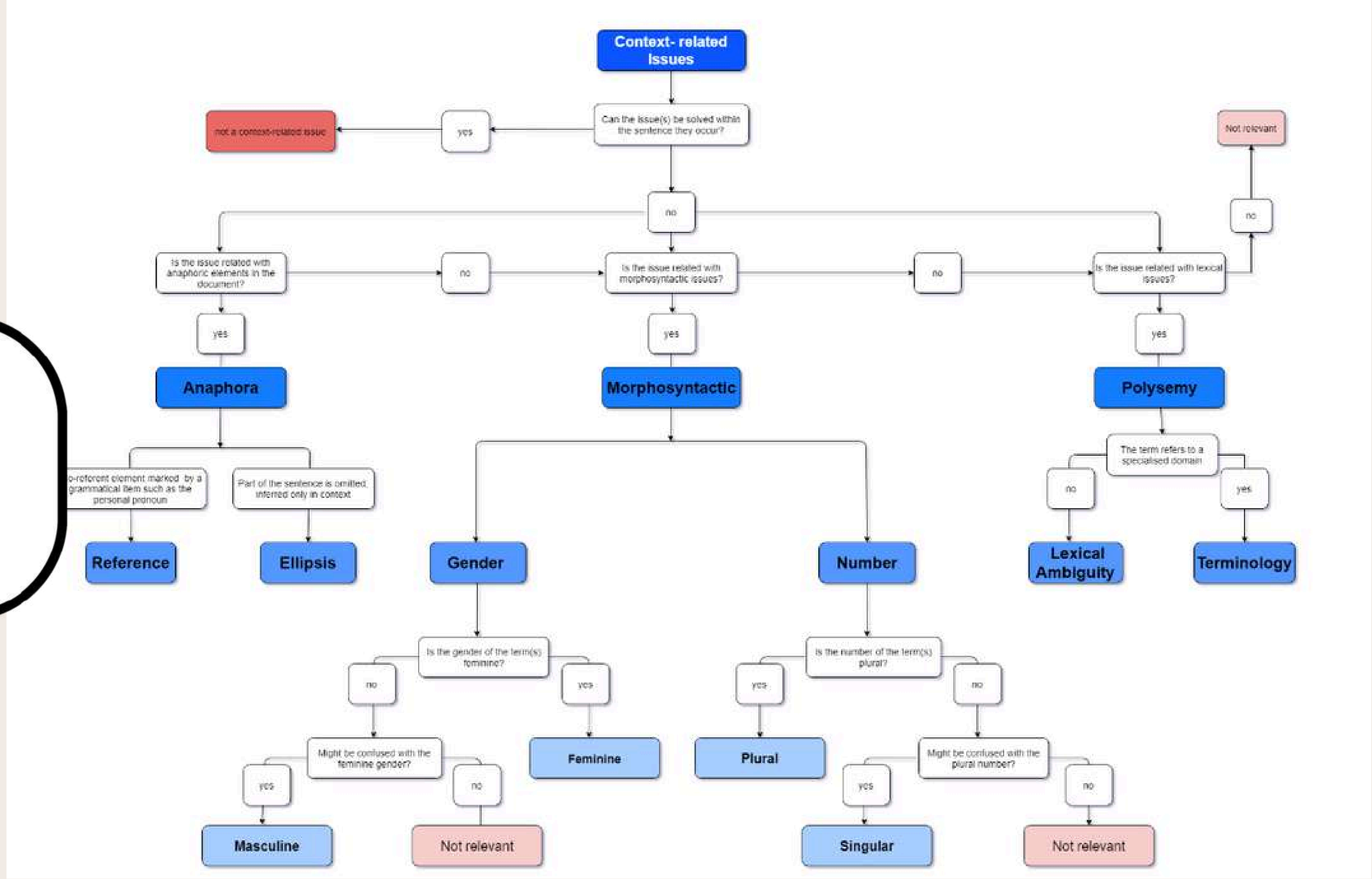


Methodology for the Annotation

- Three annotators looked for issues of: Gender, Number, Ellipsis, Reference, Lexical Ambiguity and Terminology.
- Issues were tagged in English into Brazilian Portuguese.
- Different MT systems were used to check for issues that would go unnoticed.

Test Suites - DELA Corpus

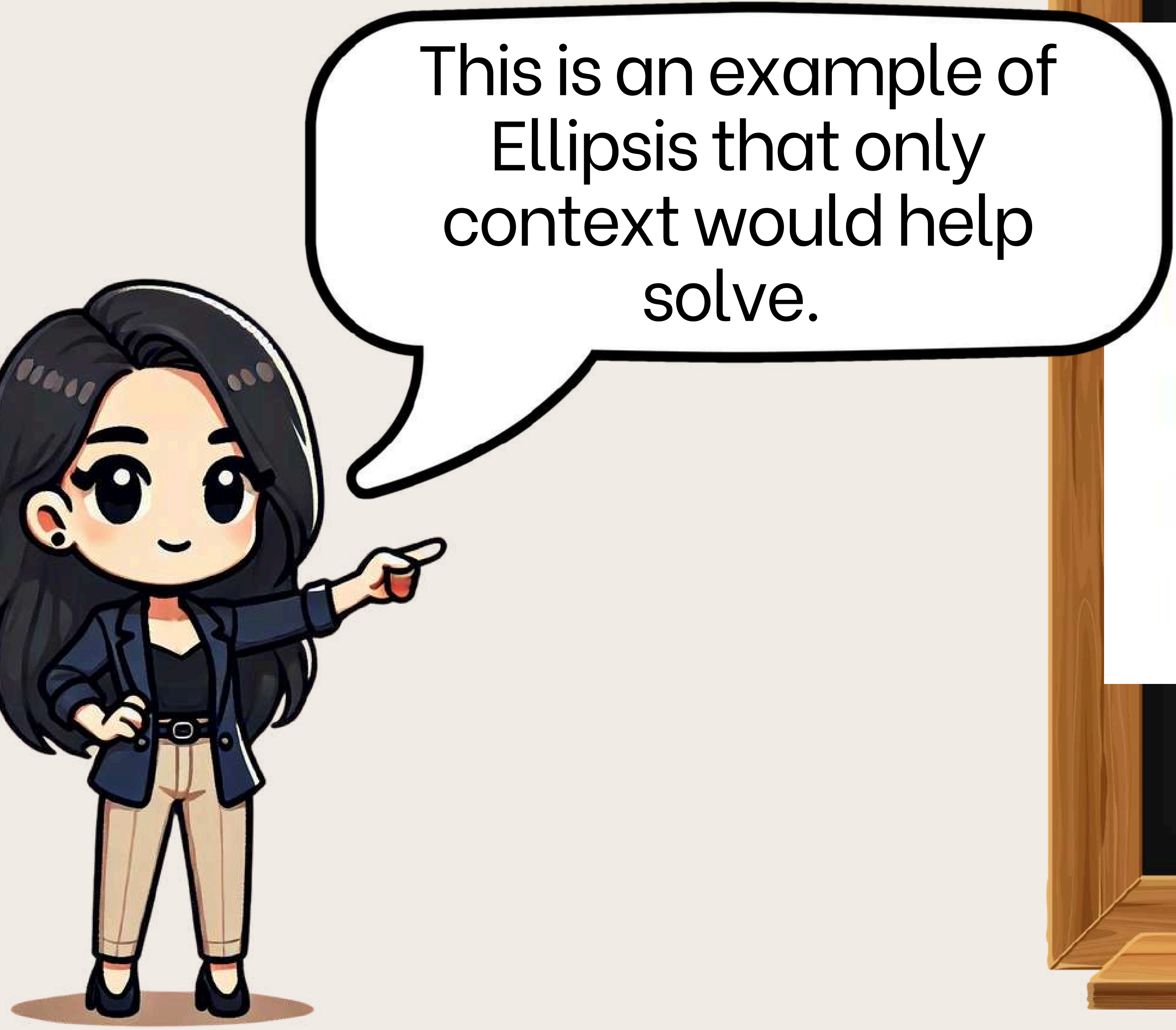




Scan the QR code for the full decision tree.

Test Suites – DELA Corpus

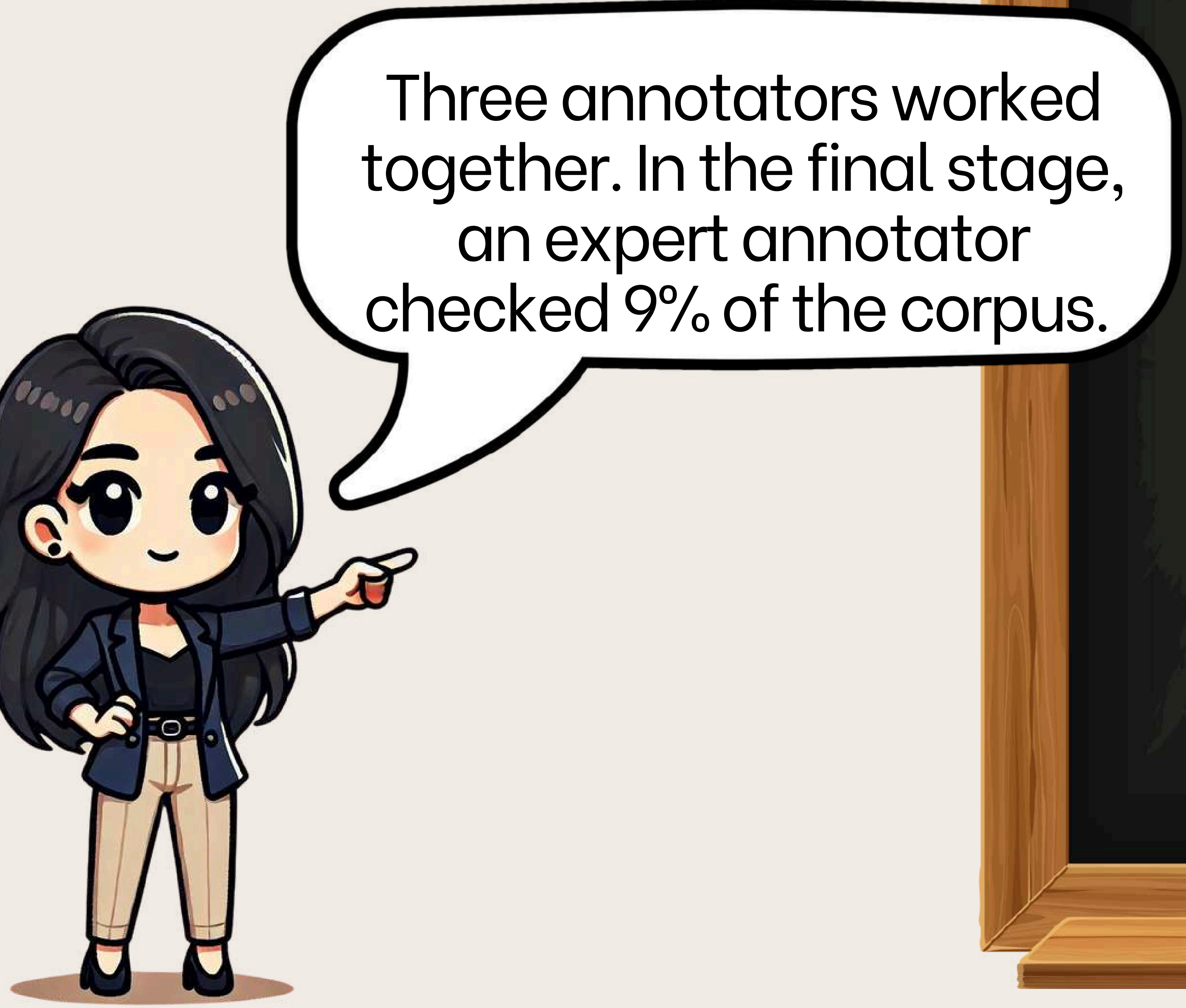
Castilho et al. (2021)



C) *In my laughter, I bellied out a "YES, I do!!"*
ellipsis –> do = think
lexical ambiguity –> do = make (incorrect) *vs* think (correct)
Sim, eu faço! (Yes, I make, incorrect) *vs*
Sim, eu **acho!** (Yes, I "think", correct)

Test Suites - DELA Corpus

Castilho et al. (2021)



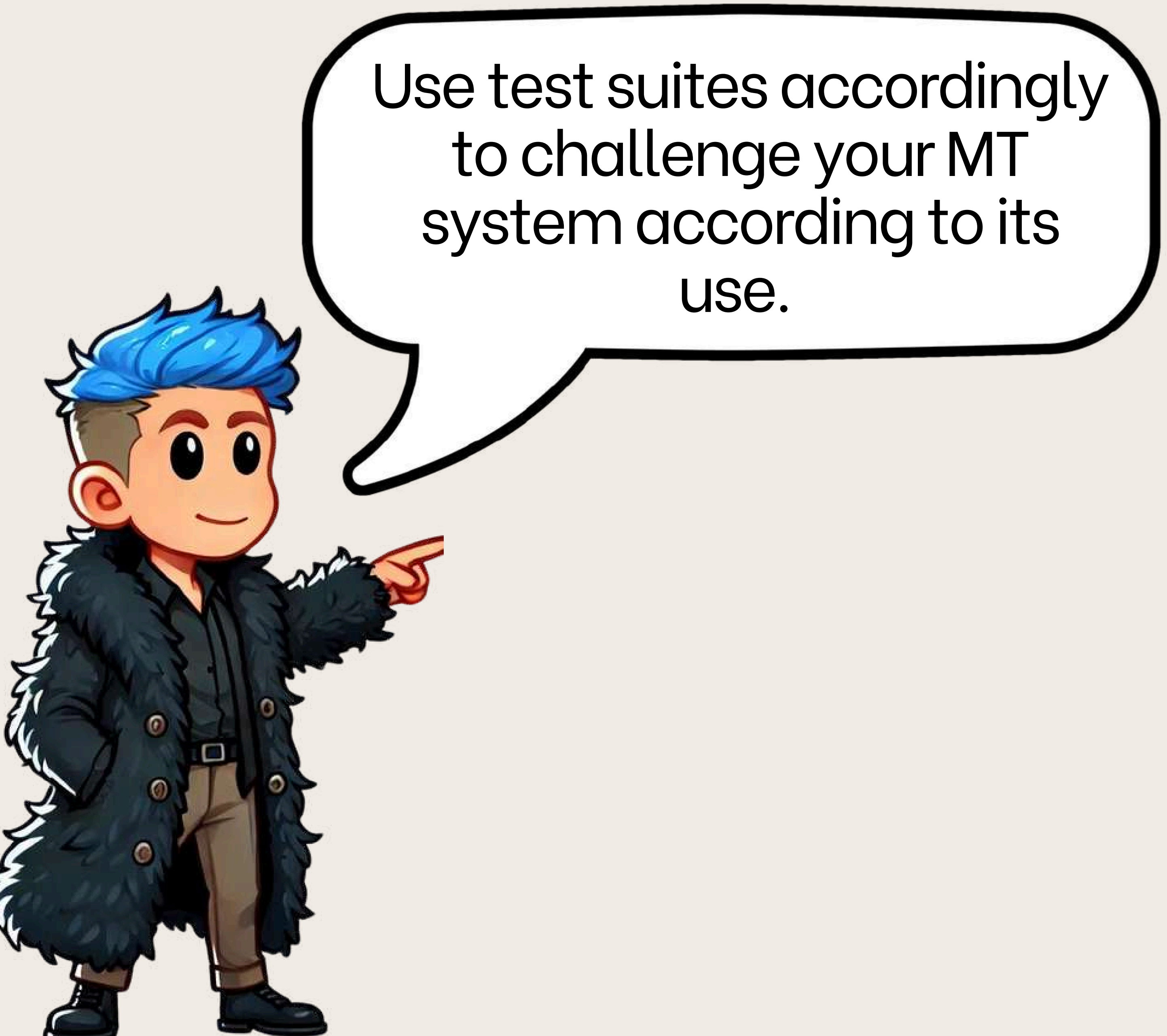
Profile of the Annotators and Agreement

- Annotators had backgrounds in linguistics, translation and computational linguistics.
- Disagreements were discussed and resolved, if it occurred.
- An additional expert annotator was involved at the final stage.
- IAA was calculated (Cohen’s Kappa) and agreement was 0.61, a substantial agreement was reached.

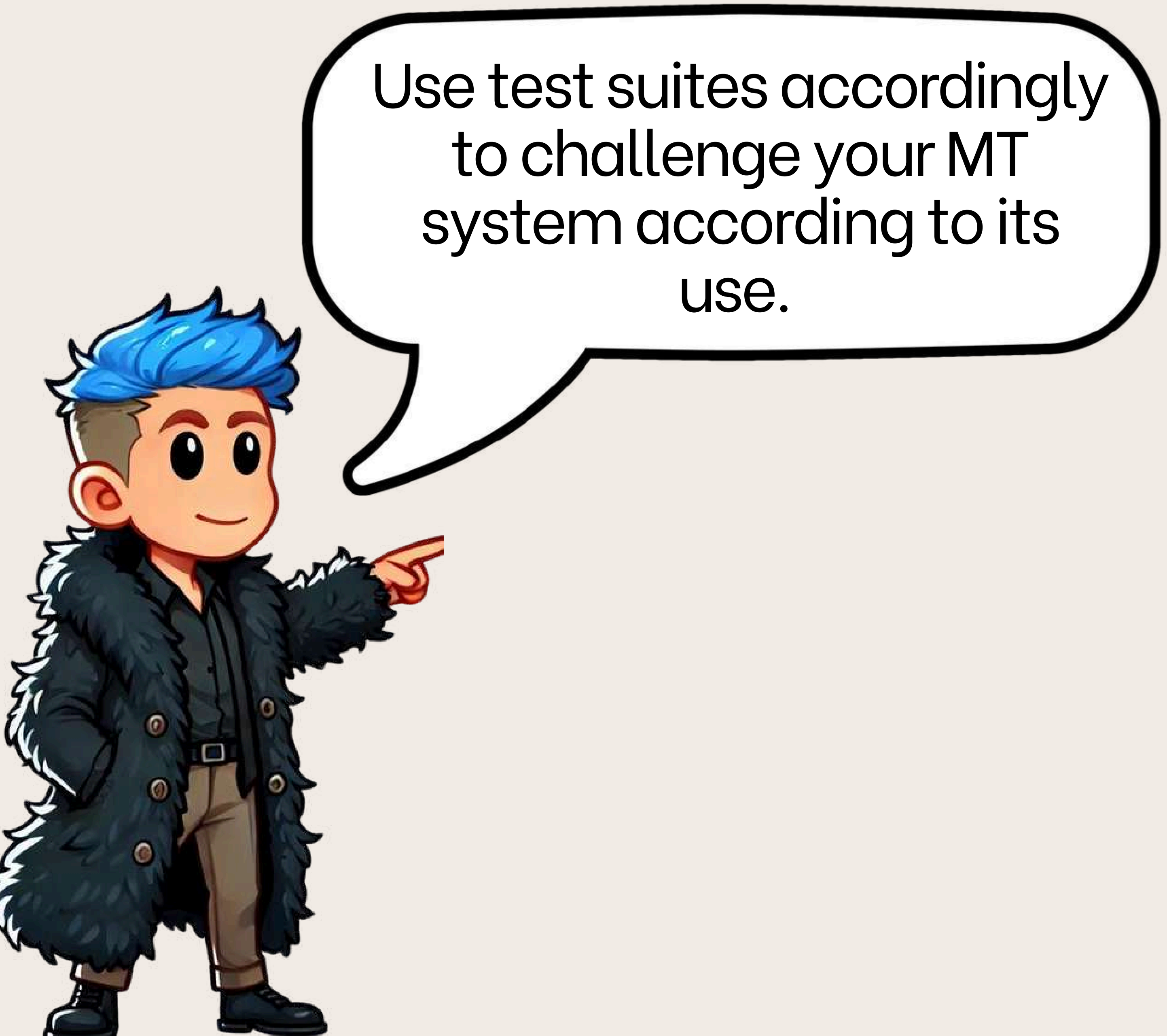
Test suites

Rachel Bawden, Rico Sennrich, Alexandra Birch, and Barry Haddow. 2018. Evaluating discourse phenomena in neural machine translation. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1304–1313, New Orleans, Louisiana. Association for Computational Linguistics.

Liane Guillou, Christian Hardmeier, Ekaterina Lapshinova-Koltunski, and Sharid Loáiciga. 2018. A pronoun test suite evaluation of the English–German MT systems at WMT 2018. In *Proceedings of the Third Conference on Machine Translation: Shared Task Papers*, pages 570–577, Belgium, Brussels. Association for Computational Linguistics.



Test suites



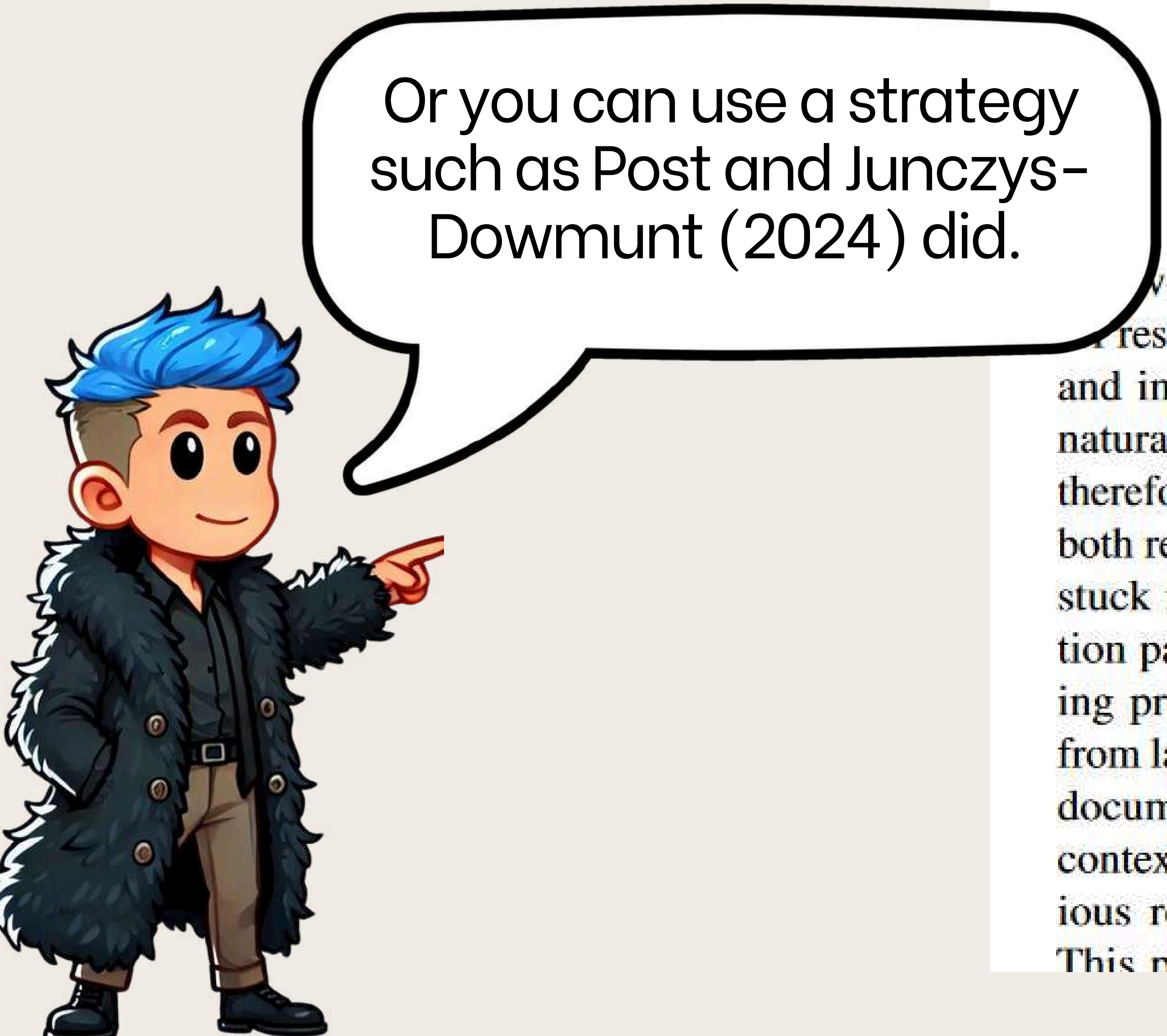
Mathias Müller, Annette Rios, Elena Voita, and Rico Sennrich. 2018. A large-scale test set for the evaluation of context-aware pronoun translation in neural machine translation. In *Proceedings of the Third Conference on Machine Translation: Research Papers*, pages 61–72, Brussels, Belgium. Association for Computational Linguistics.

Elena Voita, Rico Sennrich, and Ivan Titov. 2019. When a good translation is wrong in context: Context-aware machine translation improves on deixis, ellipsis, and lexical cohesion. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1198–1212, Florence, Italy. Association for Computational Linguistics.

Test suites

Escaping the sentence-level paradigm in machine translation

Matt Post and **Marcin Junczys-Dowmunt**
Microsoft
Redmond, Washington
{mattpost,marcinjd}@microsoft.com



Abstract

well-known that document context is vital resolving a range of translation ambiguities, and in fact the document setting is the most natural setting for nearly all translation. It is therefore unfortunate that machine translation—both research and production—largely remains stuck in a decades-old sentence-level translation paradigm. It is also an increasingly glaring problem in light of competitive pressure from large language models, which are natively document-based. Much work in document-context machine translation exists, but for various reasons has been unable to catch hold. This paper suggests a path out of this rut by

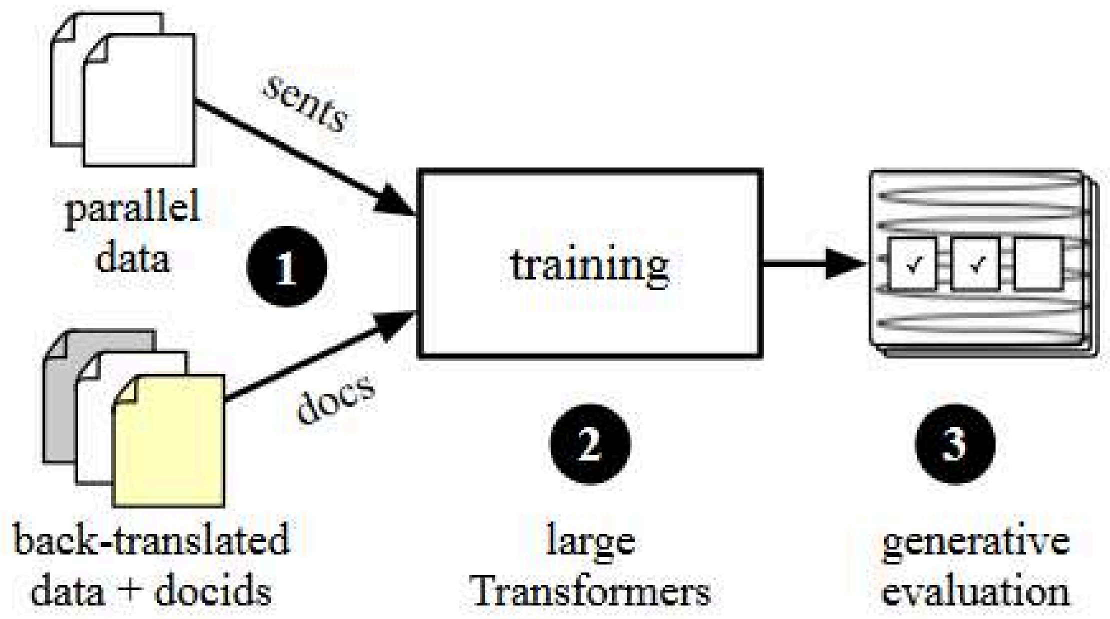


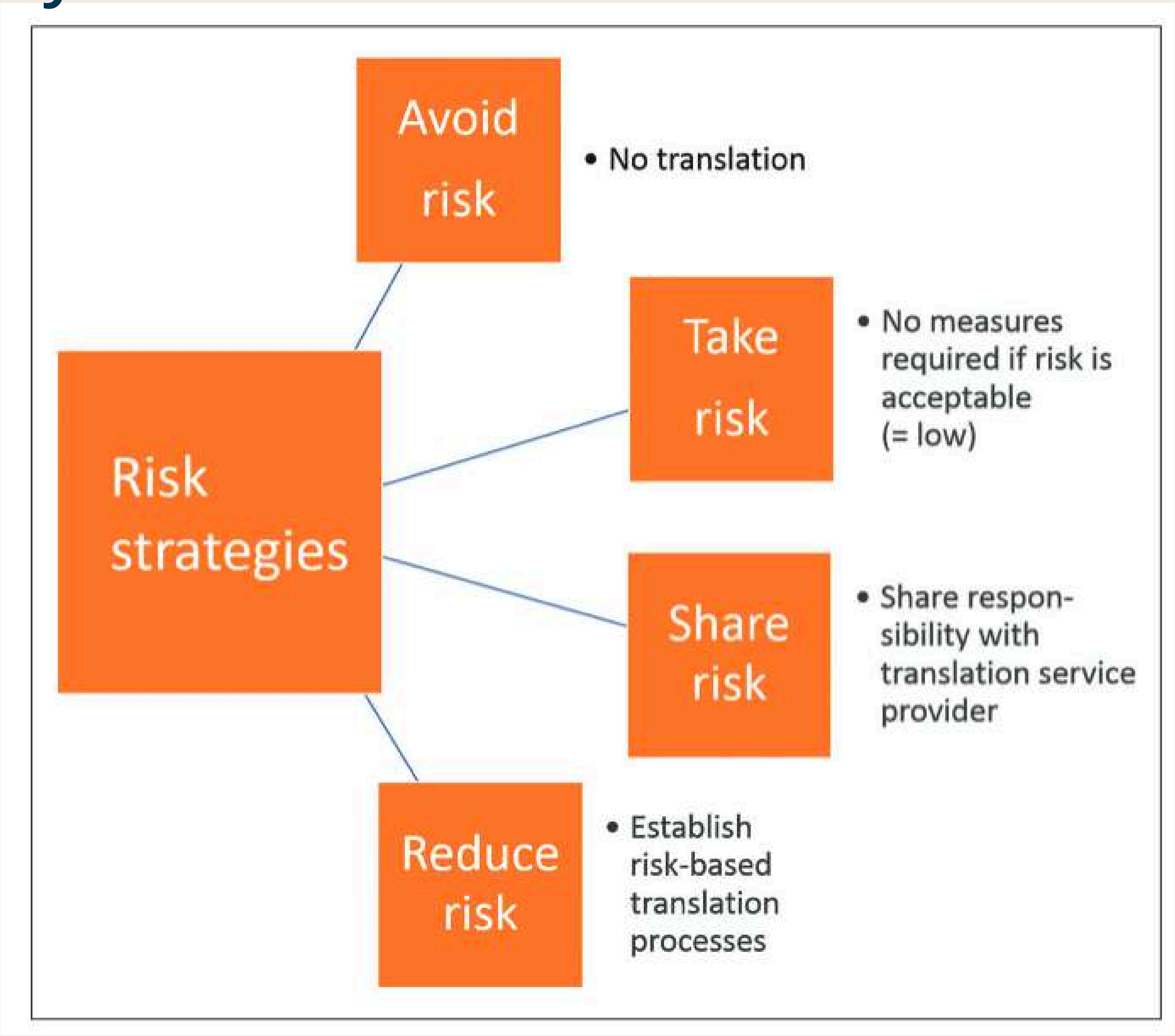
Figure 1: Escaping the rut of sentence-level translation: (1) source documents from trustworthy data only, (2) feed them into large-capacity standard Transformer models, and (3) use test sets that evaluate a model's generative ability.

Why do we need risk assessment?

ISO 31000:2018-02

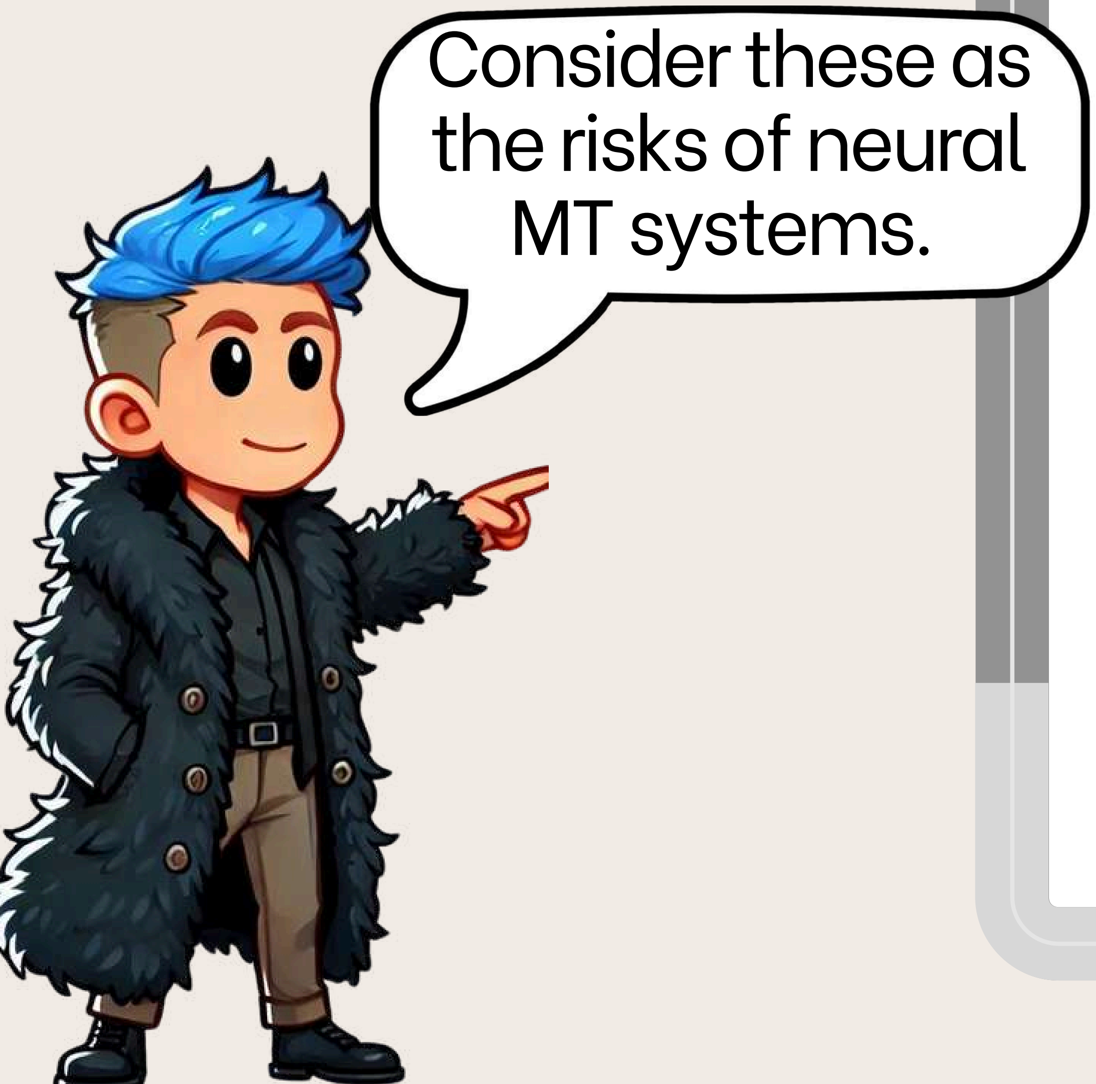
- According to this international standard, the whole organisation is responsible for risk management.
- Canfora and Ottmann (2020) suggest integrating risk management measures to translation.
- It is worth knowing how risk management processes happen in translation when considering how your MT system can be used to prevent risks.

Why do we need risk assessment?



Ottmann and Canfora (2020)

Why do we need risk assessment?



Risks

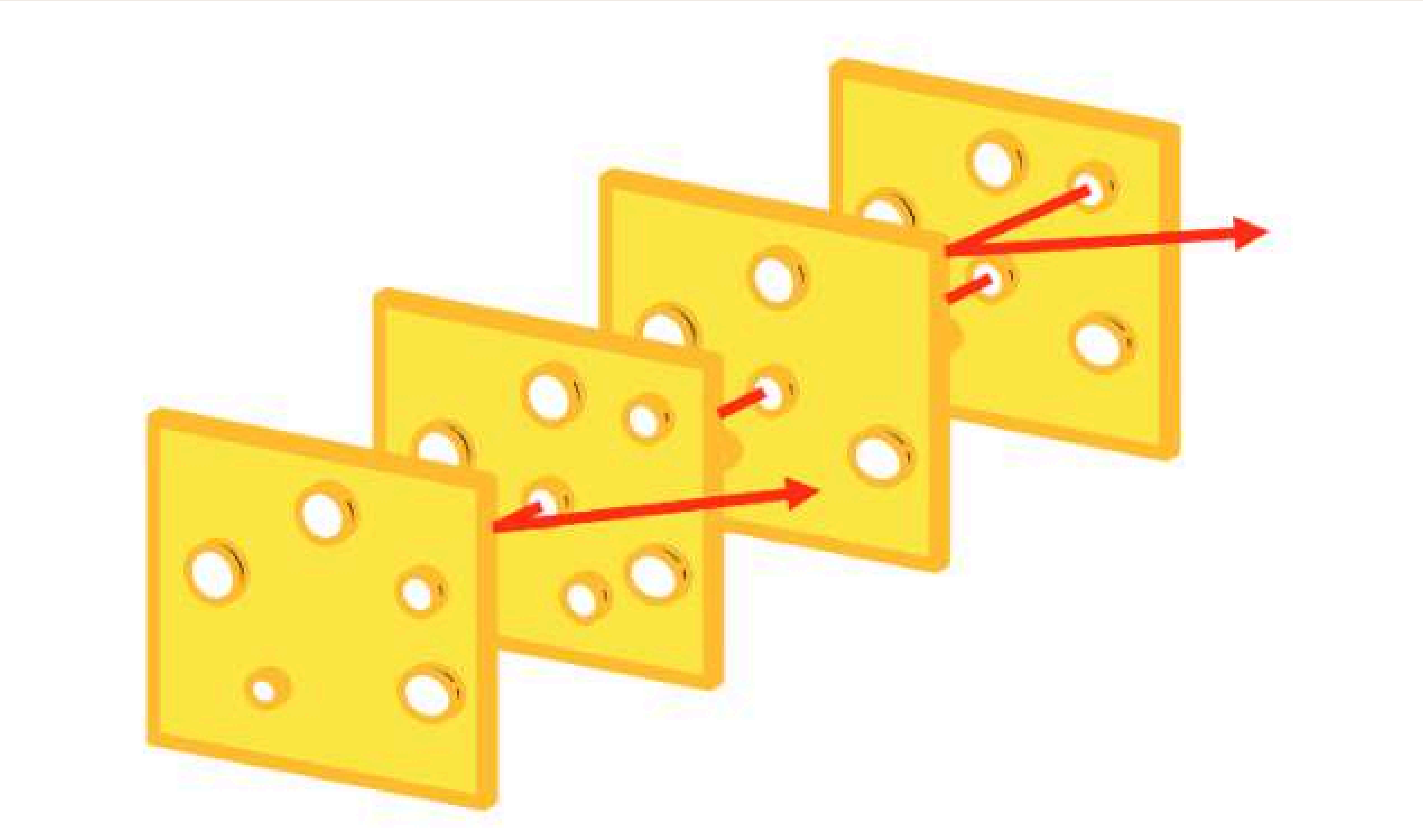
- **Translation Errors**: NMT systems, while improving, are still prone to errors. (Koehn & Knowles, 2017).
- **Liability Challenges:** Establishing accountability for errors made by NMT systems remains complex. There is no clear legal framework for assigning liability in cases where NMT output leads to issues (Moorkens & Lewis, 2020).
- **Data Privacy Risks:** The risk of personal or confidential data being processed by free online NMT engines continues to pose significant concerns, as these services may lack proper data handling or protection measures. (Slator, 2017)

Why do we need risk assessment?

Translation Risks (Canfora and Ottman, 2018)

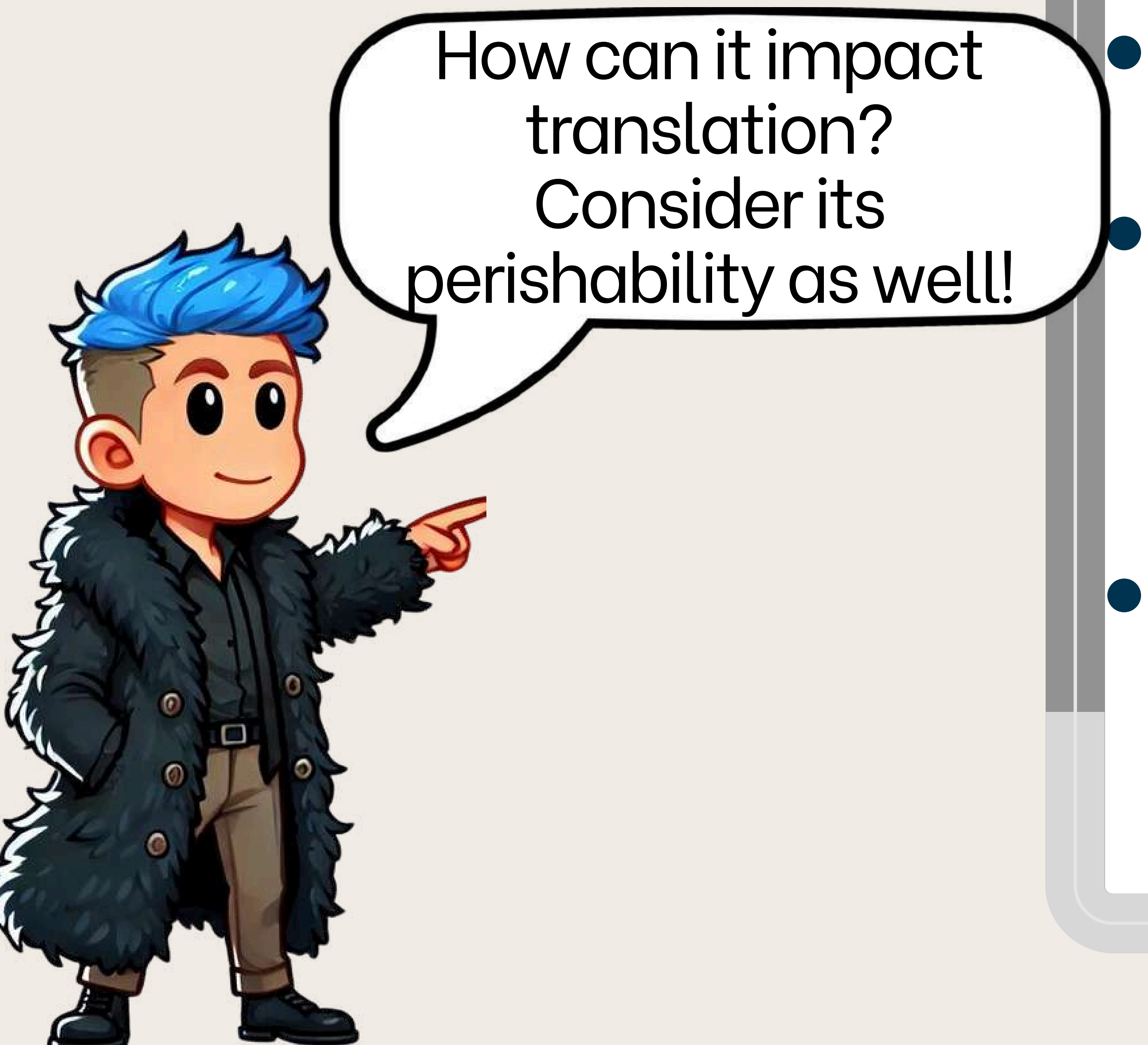
- Injury or death.
- Legal consequences
- Loss of reputation
- Financial damage
- Damage to property
- Communication impaired or impossible

Why do we need risk assessment?



Swiss cheese model (Reason 1990)

Why do we need risk assessment?



Factors for severity of risks(Canfora and Ottman, 2018)

- Circulation: Translations with a large number of copies.
- Number of languages into which a document is translated
- Use of Translation Memories: Either faulty segments in a database being used to multiply the exposure of errors.

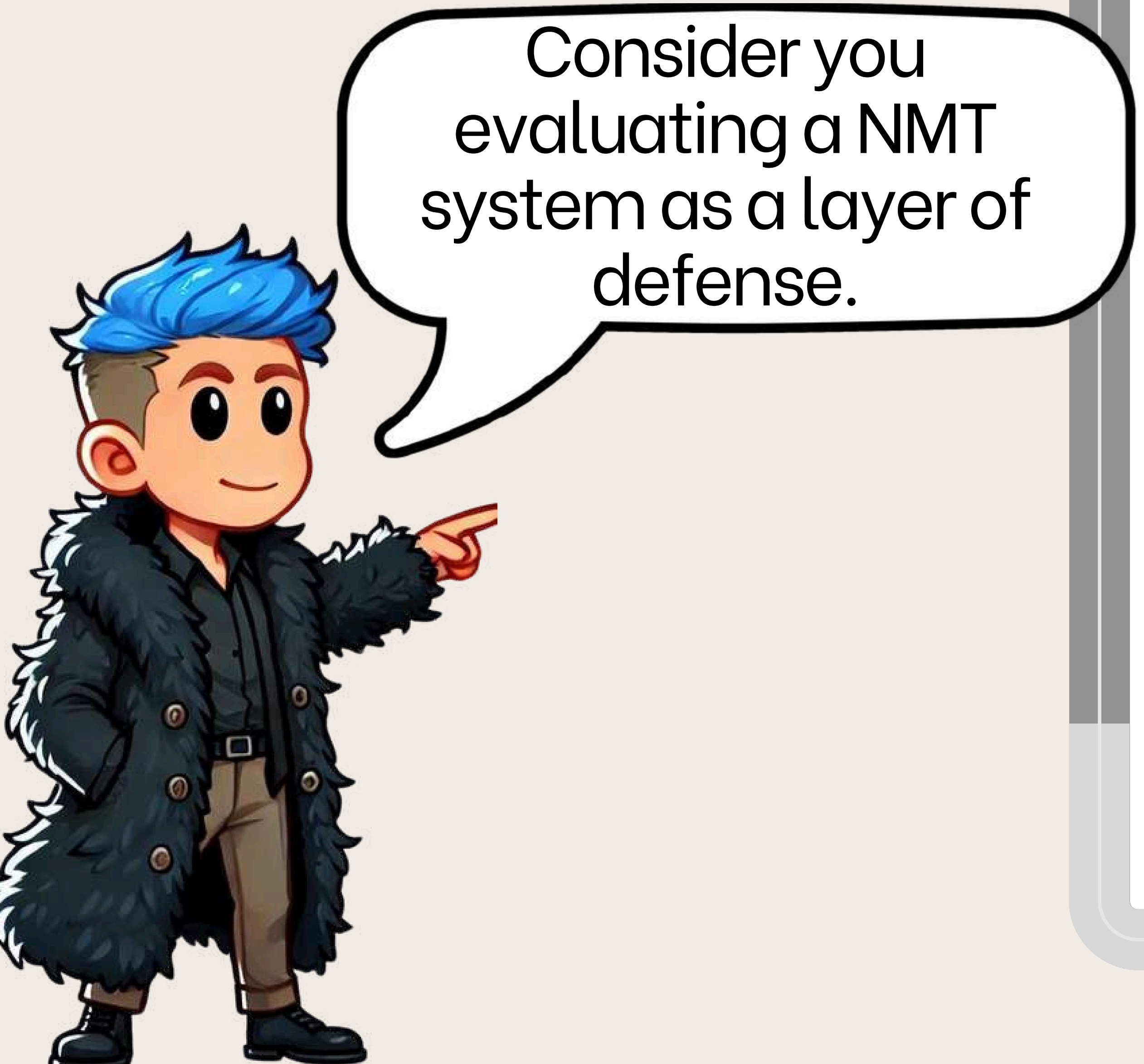
- If you train your MT system with corpora that contains errors, it will increase the likelihood of risks when using your system. **Always test it through evaluation!!!**

Why do we need risk assessment?

Factors to reduce risk (Canfora and Ottman, 2018)

- Identification of near misses.
- Reporting of identified near misses and collecting them in a database.
- Identify the root cause of factors that lead to the errors. Faulty corpora? Problems in human evaluation? Only using automatic evaluation?
- Determine how can it be fixed based on the analysis.
- Disseminate the information to everyone involved!

Why do we need risk assessment?



Layers of defense(Canfora and Ottman, 2018)

- Canfora and Ottman (2018) mention layers of defense in translation processes. But we can think about layers of defense in MT development through evaluation.
- What measures of human evaluation can you use to identify your MT systems is producing critical errors?
- Consider the specialisation of your translator acting as an expert to identify what can be a problem.

In this lecture you were able to...

Understand how to design the evaluation of MT systems considering different steps.

Understand the importance of factors such as the type of evaluator, user interface and their experience with evaluation platforms.

Understand what risk management entails and what are its consequences when not implemented.

Thank you!
See you next class.
Questions?

Send an e-mail to
joo.cavalheirocamargo2@
mail.dcu.ie