

INTRODUCING A NEW GENAI-ASSISTED L2 WRITING CORPUS

A replicable methodology for collecting L2 writing product and process data

Jennifer Thewissen, Anna Zaffinelli, Shiyi Guo, Pauline Jadoulle, Nina Vandermeulen, Nathan Vandeweerd

September 19th 2026

Learner Corpus Research Conference, Charles University

jennifer.thewissen@uantwerpen.be, anna.zaffinelli@ru.nl

INTRODUCING A NEW GENAI-ASSISTED VS. UNAIDED CORPUS

BACKGROUND: THE THREAT OF GENAI (1)

Chat GPT Use in Fall Quarter 2022 Finals

If you used ChatGPT to assist with your final fall quarter assignments and/or exams, to what extent did you use it?

Brainstorming, outlining, and forming ideas

59.2%

Answered multiple-choice questions with the help of ChatGPT

29.1%

Submitted written material from ChatGPT with edits

7.3%

Submitted written material from ChatGPT without edits

5.5%



Source: [Fizz Post](#) • In an anonymous poll uploaded on Fizz, 4,497 Stanford students answered whether or not they used ChatGPT to help with their Fall final assignments.
Graphic: MARK ALLEN CU/The Stanford Daily

(cf. [Stanford Daily](#))

Het Parool 

Plus

**Door AI is scriptie als
toetsmiddel in onderwijs niet
langer houdbaar: 'We geven
diploma's aan chatbots'**

(cf. [Het Parool](#))

INTRODUCING A NEW GENAI-ASSISTED VS. UNAIDED CORPUS

BACKGROUND: THE THREAT OF GENAI (2)

“

*As a teacher, I noticed that there's a discrepancy. I read an read an article written by a student, and I thought that it that it was really perfect. Then he or she presented, and and **there was a knowledge gap between what I had had assumed from the report and the presentation** he or he or she gave.*

“

*And we see more and more that students are using generative AI. I think 80–90% uses it in some way. And I And I think **that's also a future that we're working towards, towards**, that it is implemented in our work, but **we still still want students to be able to think critically** themselves themselves and **that asks a change in education**.*

“

*We face some pretty significant issues in the program that I teach [...] specifically in the linguistics area of modern languages [...] because **we're looking to test aptitudes and proficiency** in certain respects, and **that has become almost impossible**.*

Zaffinelli et al. (forthcoming)

We are entering a new writing ecology

raising fundamental questions about:

Authorship

Authenticity

Language development

Task performance

INTRODUCING GENAISIS

RATIONALE: WHY GenAIsis?

GenAIsis

GenAI

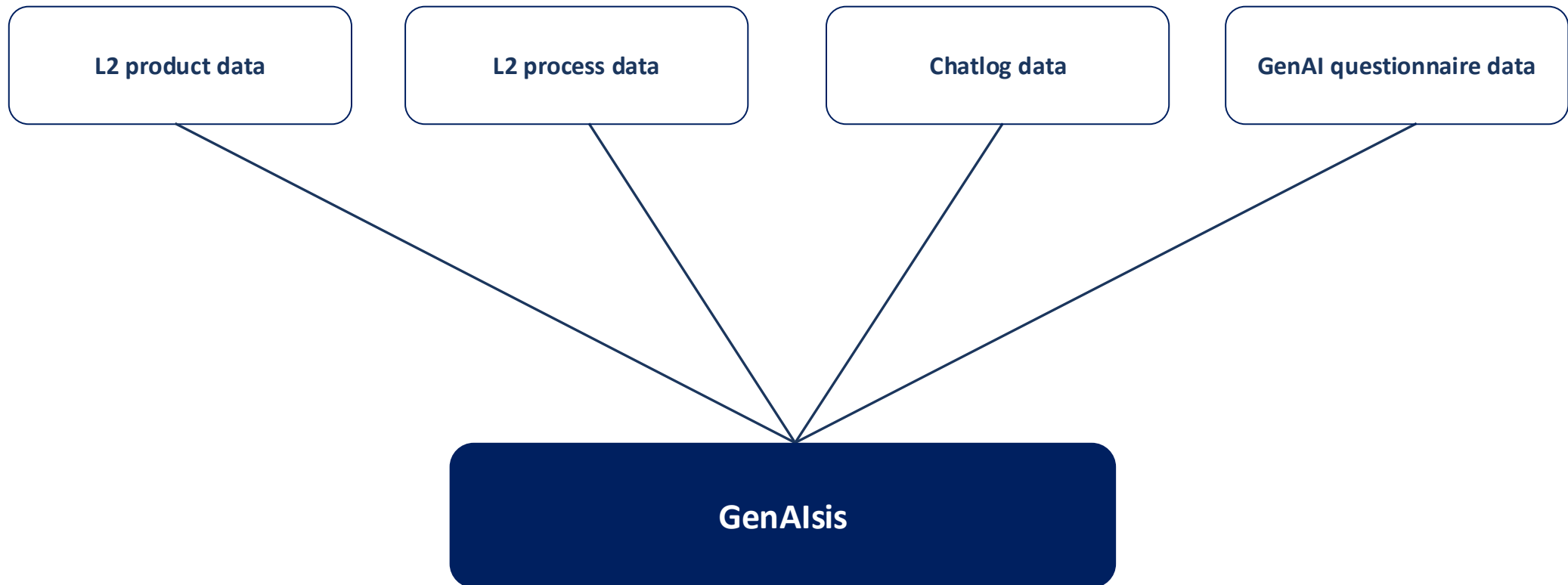
+

genesis

Evoking both “origin” and “transition” , i.e., the emergence of a new mode of interactive, GenAI-assisted writing

INTRODUCING GENAISIS

MULTIPLE LENSES ON ONE DATASET



BUT BEFORE WE COULD EVEN COLLECT ANYTHING..

INTRODUCING GENAISIS

NEW ETHICAL CHALLENGES TO ADDRESS

TEACHING DEPENDENCY

OUR QUESTION

How do we put students at ease using GenAI if the study is also part of a teaching component?

CONSENT

OUR QUESTION

Can we 'force' students to use GenAI?

GENAI INTERACTION

OUR QUESTION

How do we let students interact with GenAI safely?

PROCESS DATA

OUR QUESTION

How do we collect process data safely while students interact with GenAI?

L2 PRODUCT DATA

INTRODUCING GENAIS

PRODUCT DATA (1)

PARTICIPANTS

COHORT

1st-year BA students

PROGRAMME

Language and Literature

COURSE

English Proficiency 2 (semester II)

LEVEL ENTERING UNIVERSITY

B2 (expected)

SAMPLE

100 tested, 81 consented to share final texts

WRITING TASKS

PERSUASIVE ESSAY

Essay on English-medium instruction (EMI) in the Flanders

Source material provided

NARRATIVE ESSAY

Share a personal narrative about your experience with language.

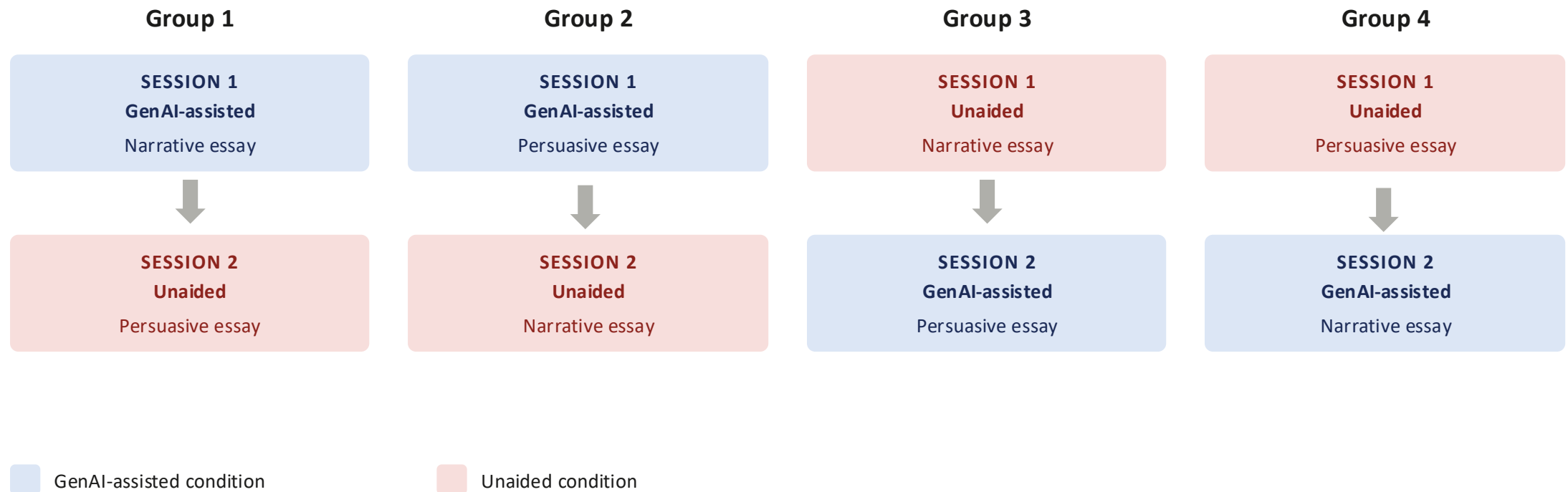
1 hour, 400 words minimum

INTRODUCING GENAISIS

PRODUCT DATA (2)

EACH PARTICIPANT COMPLETES BOTH CONDITIONS AND TASK TYPES, IN OPPOSITE PAIRINGS

They are randomly assigned to one of four pairings:



INTRODUCING GENAISIS

PRODUCT DATA (3)

SESSION 1

- 1 **Randomization**
Condition (GenAI-assisted / unaided) × task (persuasive / narrative)
- 2 **LexTALE vocabulary test**
- 3 **Writing task**
1 hour, word limit ≥ 400 words, Inputlog / OBS recorded
- 4 **GenAI questionnaire**

1 WEEK
LATER

SESSION 2

- 1 **Opposite condition & task**
- 2 **Writing task**
1 hour, word limit ≥ 400 words, Inputlog / OBS recorded
- 3 **Language background questionnaire**
- 4 **Language background questionnaire**

INTRODUCING GENAISIS

PRODUCT DATA (4)

Condition	Persuasive task	Narrative task	Total
Unaided scripts (assigned)	42 <i>17,753</i>	39 <i>18,315</i>	81 <i>36,068</i>
GenAI-assigned scripts (actual use)*	22 <i>9,603</i>	26 <i>11,290</i>	48 <i>20,893</i>
GenAI-assigned scripts (no use)**	17 <i>6,909</i>	15 <i>7,341</i>	32 <i>14,250</i>
Total scripts	81 <i>34,265</i>	80 <i>36,946</i>	161 <i>71,211</i>

* Texts with verified GenAI use based on chatlog evidence.

** Texts assigned to the GenAI condition but for which the participants did not use GenAI.

Figures in italics indicate token count.

L2 PROCESS DATA

INTRODUCING GENAISIS

L2 PROCESS DATA

Beyond the final product: **136** Inputlog (Leijten & van Waes, 2013) and **20** screen recordings available.

INPUTLOG

Enables research on

- Quantitative analyses:
 - Fluency and pausing measures
 - Revision measures
 - Source usage

OBS STUDIO

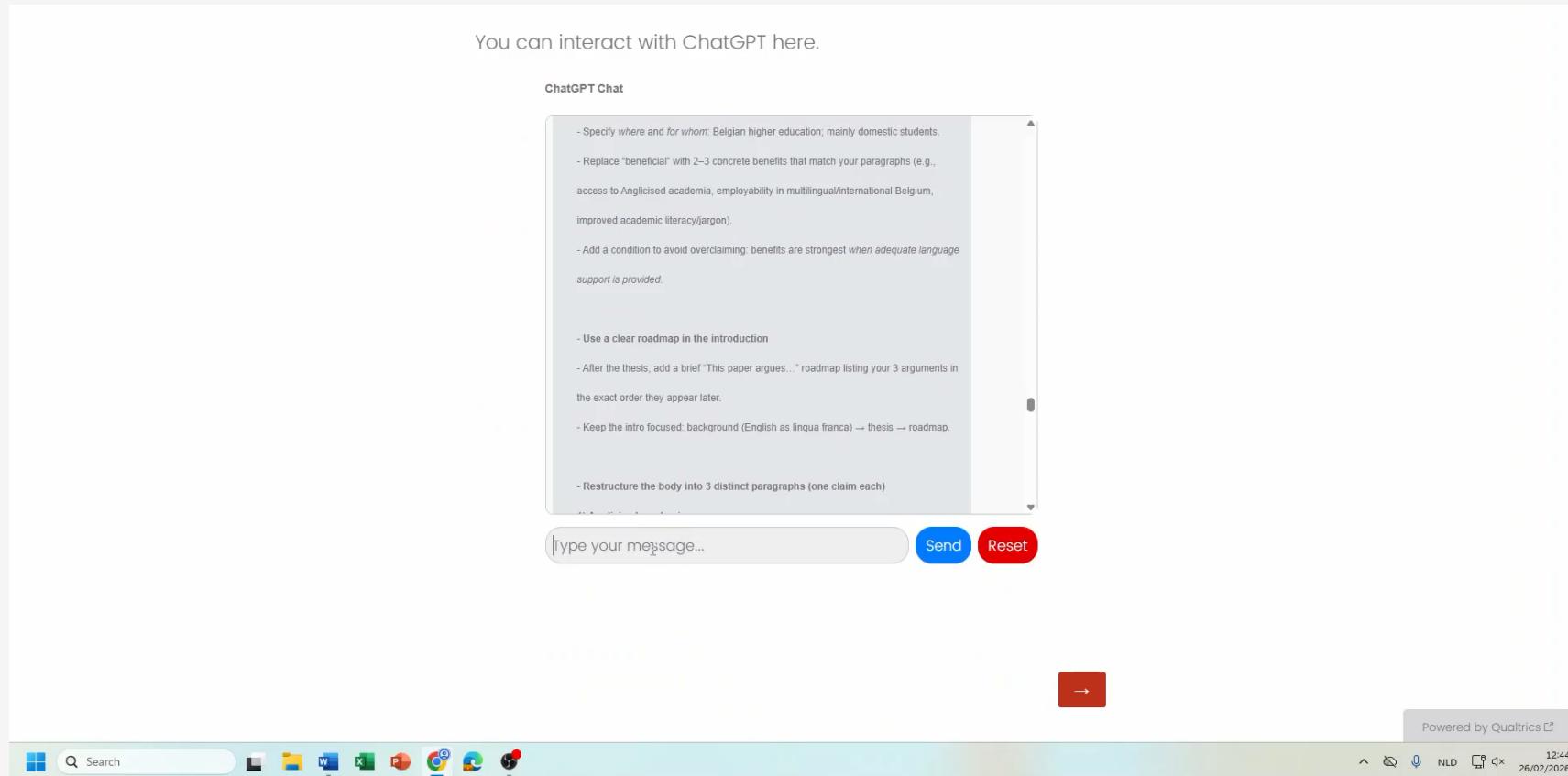
Enables research on

- Qualitative analyses:
 - Source interaction
 - Retained vs. rejected linguistic/content help from GenAI
- Possibility of merging the recordings with Inputlog timestamps

CHATLOG DATA

INTRODUCING GENAISIS

CHATLOGS (1)



Screen recording

INTRODUCING GENAISIS

CHATLOGS (2): A SAFE AND CONTROLLED GENAI ENVIRONMENT

- Participants interact with GenAI within the Qualtrics survey through a dedicated Openrouter **API key** (cf., Ballot et al., 2026)
- ChatGPT v.5.2, temperature = 1.0, data training settings = disabled



- **Safe access** to LLM
- **Comparability** and **replicability across participants**
- **Chatlog conversation history** is available

ChatGPT Chat

Waiting for Response.

Hi

2 user what should I definitely include in a narrative essay about my experience with language?

0 user Okay, I have to write a text of 400 words. I have to talk about a true life story about my experience with languages. I want to talk about my brother who moves to America when I was 4 years old. That's when I started learning English and how I developed my love for languages

2 user how would one translate 'Algemeen Nederlands' in Dutch?

INTRODUCING GENAISIS

CHATLOGS (3)

0 user Are there any spelling mistakes in this sentence?
"Living in this modern day and age, it is fair to say that every individual comes in contact with language almost constantly throughout their life. "

0 assistant There are no spelling mistakes in that sentence.
A small style note: the common phrase is "**modern day and age**" or "**modern age**"—using both "day" and "age" together can sound a bit redundant, but it isn't a spelling error.

0 user what does 'decree' mean?

0 assistant "Decree" means an **official order or decision**, usually made by a **government authority** or a **judge/court**, that must be followed.
Examples:

- **Government:** "The king issued a decree banning travel." (an official rule)
- **Court:** "The divorce decree was finalized." (a court's official judgment)

As a verb, **to decree** means **to officially order** something: "The court decreed that the property be divided."

0 user outline intro body argument 1: academia is anglicised argument 2: Belgian students who enter higher education already have a good understanding of english -> ofte second or third language required on the job market+ belgium is a multilingual state -> languages are important argument 3 : it boosts academic jargon and academic literacy conclusion
Make an outline for a short paper f about 500 words,

0 assistant **Outline for a ~500-word paper
(English in Belgian higher education)**

Title (working)

English in Belgian Higher Education: Pragmatic Benefits in an Anglicised Academic World

LEARNER & TASK METADATA

INTRODUCING GENAISIS

LEARNER METADATA

- **Learner metadata**

Based on the core metadata scheme for learner corpora (Paquot et al., 2024)

- o Language background
- o Learners' exposure to English L2

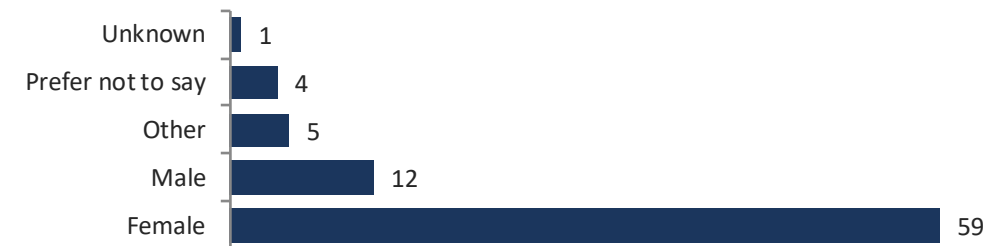
- **LexTALE score**

- **GenAI questionnaire**

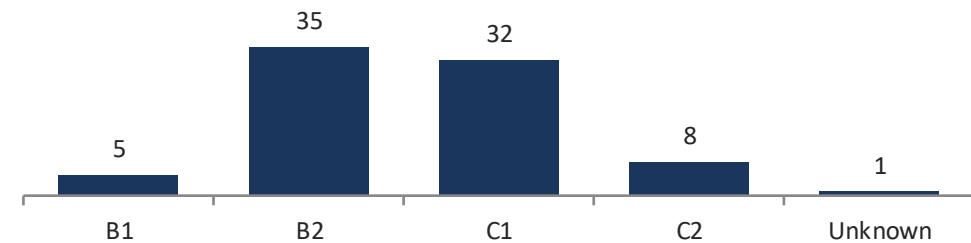
- o GenAI acceptance
- o Perceived extent of GenAI usage
- o Perceived type of GenAI usage

Participant profile ($n = 81$)

Gender



Self-assessed English proficiency



Age 17–39 ($M = 20$, $SD = 2.88$)

LexTALE test 64.6–100 / 100 ($M = 87.2$, $SD = 8.65$)

INTRODUCING GENAISIS

GENAI QUESTIONNAIRE (1)

“

*I believe **one can learn the skills to write better texts with more depth than a machine can do.** Everything the machine can do, you can do yourself, or with the help of other people.*

“

*Because **I'm too scared that using GenAI will lead to my texts lacking personality** and I prefer having a text that is absolutely mine than to one that does not feel like it is.*

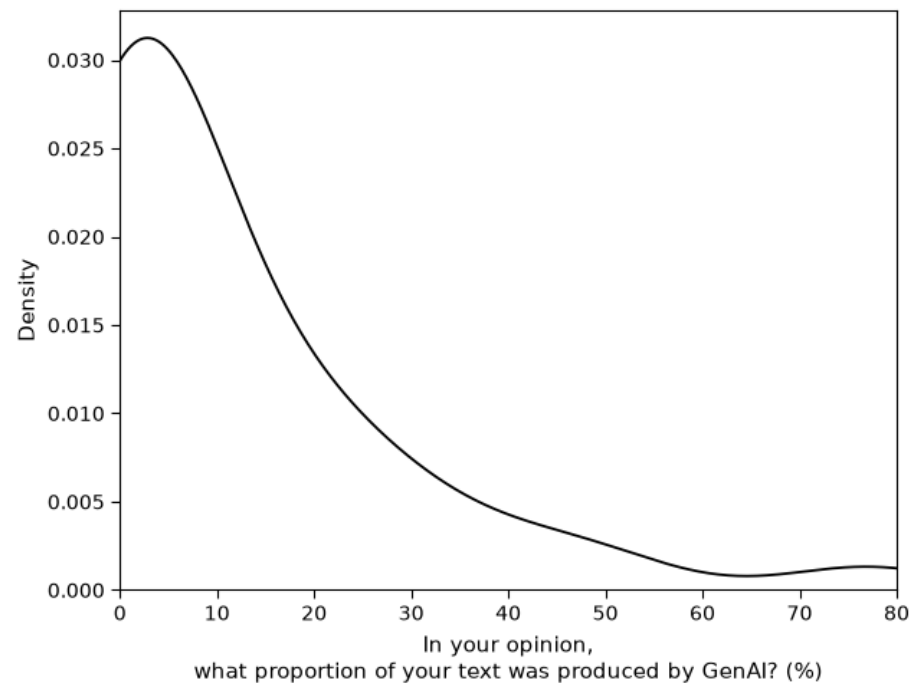
“

*Because **I think it is important to develop the skill myself,** instead of relying on GenAI.*

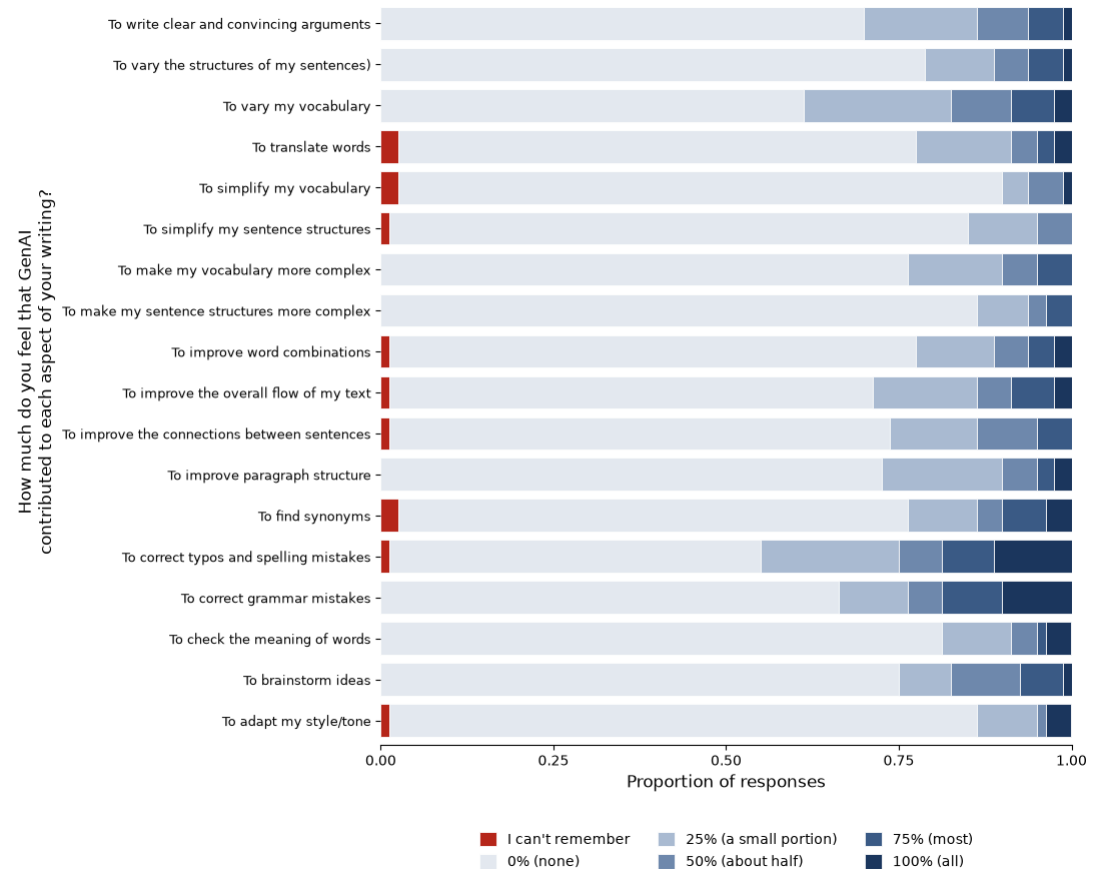
INTRODUCING GENAISIS

GENAI QUESTIONNAIRE (2)

PERCEIVED EXTENT OF GENAI USAGE

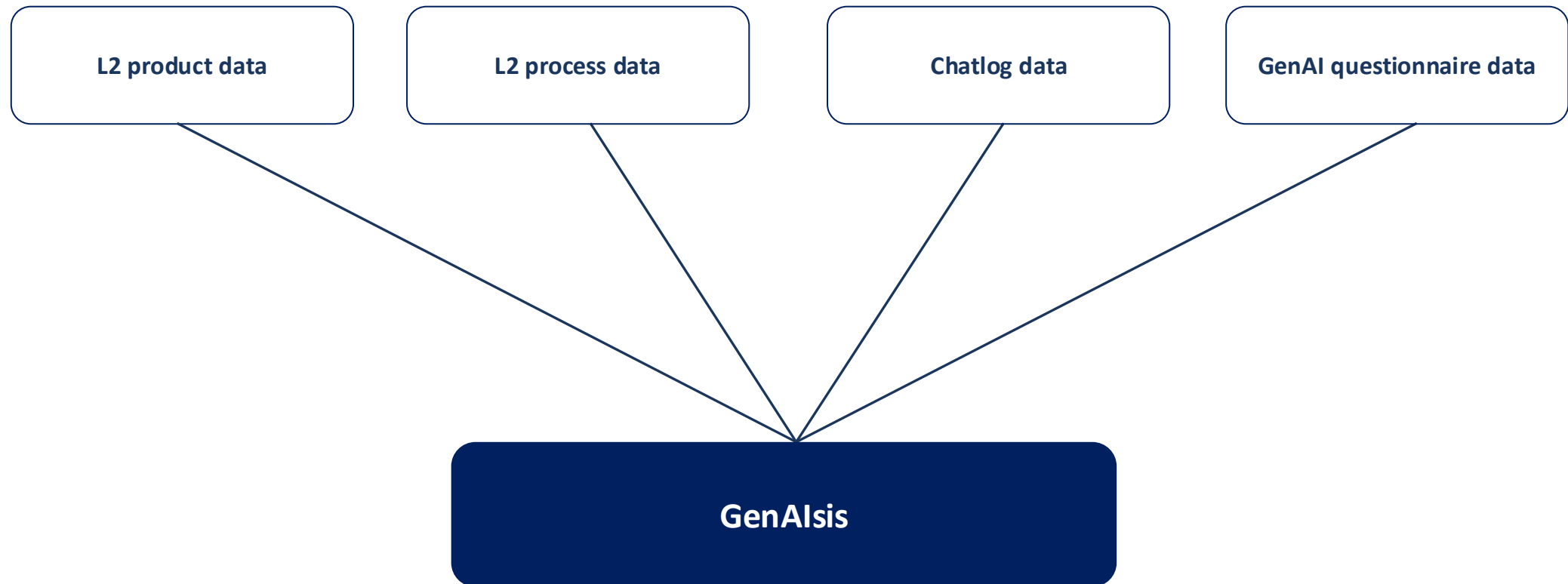


PERCEIVED TYPE OF GENAI USAGE



INTRODUCING GENAISIS

RECAP



INTRODUCING GENAISIS

ETHICAL ISSUES AND BEST PRACTICES

TEACHING DEPENDENCY

OUR QUESTION

How do we put students at ease using GenAI if the study is also part of a teaching component?

BEST PRACTICE

Carefully maneuvering teaching and research responsibilities throughout the data collection.

CONSENT

OUR QUESTION

Can we 'force' students to use GenAI?

BEST PRACTICE

Leave students the option to 'opt out' from the GenAI condition.

GENAI INTERACTION

OUR QUESTION

How do we let students interact with GenAI safely?

BEST PRACTICE

Use an API key, embed the LLM into a safe survey environment, and ensure consistent use across participants.

PROCESS DATA

OUR QUESTION

How do we collect process data safely while students interact with GenAI?

BEST PRACTICE

Make sure students never enter passwords or emails (again, use an API key!).

INTRODUCING GENAISIS

STAY UPDATED!

INTERESTED IN THIS CORPUS?

Expected release: Spring 2027

Follow for more updates on: genaisis.cls.ru.nl or reach out to genaisis.corpus@ru.nl

INTRODUCING GENAISIS

REFERENCES

- Ballot, P., van de Sande, Y., Schraffenberger, H. *et al.* Chatbots reduce health-related conspiracy beliefs not because of but despite being perceived as AI. *Sci Rep* 16, 25692 (2026). <https://doi.org/10.1038/s41598-026-66242-5>
- Leijten, M., & Van Waes, L. (2013). Keystroke Logging in Writing Research. *Written Communication*, 30(3), 358–392. <https://doi.org/10.1177/0741088313491692>
- Lemhöfer, K., & Broersma, M. (2011). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, 44(2), 325–343. <https://doi.org/10.3758/s13428-011-0146-0>
- Paquot, M., König, A., Stemle, E. W., & Frey, J.-C. (2024). The Core Metadata Schema for Learner Corpora (LC-meta). *International Journal of Learner Corpus Research*, 10(2), 280–300. Portico. <https://doi.org/10.1075/ijlcr.24010.paq>

