

Automatic Text Simplification

Written communication between municipalities and citizens

Course

Master of Applied AI

Year and Term

2025-2026 Block 2

Version

1.0

© 2025 Copyright Amsterdam University of Applied Sciences

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or made public in any form or by any means, electronic, mechanical, printouts, copies, or in any way whatsoever without the prior written consent of the Amsterdam University of Applied Sciences.

1. Introduction

Clear communication is essential in the interaction between municipalities and their citizens. Official letters must be written in such a way that citizens, independent of educational background or literacy level, can understand their rights, obligations, and actions.

Achieving this goal presents challenges: letters from municipalities can often be complex, written in administrative or legal language, making them inaccessible to citizens with lower literacy levels. In the Netherlands, one in six adults (2.5 million people) struggles with reading and writing (Netherlands Court of Audit, 2016).

There are attempts to mitigate this issue, and the public sector is trying to promote inclusive and accessible communication to citizens. One example is the guideline published by the City of Amsterdam to help employees write clearly and simply in all written communication that targets audiences with limited literacy (City of Amsterdam, 2024).

However, as Kim et al. (2024) pointed out, these efforts have proven insufficient. A study (Corsius et al., 2022) that evaluated 240 texts from 70 Dutch public organisations found that the texts, which discussed crucial information about payments and healthcare, were not sufficiently understandable due to lexical complexity, vague or indirect style, and text length, among other factors. Other challenges include the target levels described in CEFR (Common European Framework of Reference for Languages). Identifying each level is not a straightforward task and requires specialised knowledge from writers.

The Municipality of Oostzaan has developed its own **schrijfwijzer** (writing guide) to ensure that official letters are written at **B1 level**. Recently, Oostzaan initiated a pilot project to rewrite a set of letters in accordance with their **schrijfwijzer**. Oostzaan is interested in collaborating with us to explore how AI tools can support their employees in creating clear and understandable communication that aligns with their writing guidelines.

As a real-world project, students can design and test AI-based solutions with direct input from a municipality, contributing to a societal challenge: making communication from municipalities more accessible.

Based on these challenges and the potential of language simplification approaches, we propose a broad research question:

How can AI technology support municipalities such as Oostzaan in rewriting or evaluating the quality of letters according to their schrijfwijzer?

2. Use Cases

Students should design and develop an application **that supports Oostzaan's employees** in rewriting, writing or evaluating the quality of letters based on their writing guidelines. The municipality is providing a set of example letters and a writing guide.

Below, you can see some examples of more research topics. Note that they are **not exhaustive**. Students are encouraged to define their own research questions in consultation with the municipality and project coaches. Possible directions include:

1. **Exploring alternatives to proprietary AI tools:**
What open-source, sustainable, or transparent language models could support the rewriting of municipal letters?
2. **Assessing the quality of AI-generated texts:**
How can the quality of AI-rewritten letters be measured and compared to human-written ones? What kinds of errors (e.g., omissions, hallucinations, tone inconsistencies) occur, and how do they affect readability and citizen understanding?
3. **Ensuring compliance with communication guidelines:**
How can AI-assisted letters be evaluated against Oostzaan's *schrijfwijzer* or plain-language principles? What metrics, checklists, or automated methods could help ensure compliance?

Students should **consider the existence of real users (municipal staff) interacting with your system**. They should explore the context of the application to understand better the problem and for whom it needs to be solved. This has a strong connection with the design learning outcomes A.

3. Design considerations

The following elements should be considered when designing and implementing a text simplification system.

Target audience

Focus on employees who evaluate or write letters, as well as citizens who receive these letters.

User Autonomy and Control

You should ensure that the application helps employees. It should not take over the entire task. Instead, it should allow human control, creativity, and autonomy.

Bias and Fairness

Simplifications targeted to specific target groups are prone to biased results and potential hallucinations or information loss. Ensure you review AI outputs to identify these problems upfront and find ways to mitigate them.

Ensuring Privacy Concerns

Letters often contain sensitive information. Even if Oostzaan provides non-privacy-sensitive examples, solutions must be designed with data protection in mind.

Maintaining Semantic Precision

You should avoid oversimplification that could lead to the loss of critical details.

Evaluating the Success of Generated Content

Unlike discriminative tasks (e.g., classification), evaluating the quality of generated content, such as simplified text, is more complex and subjective. Therefore, you should select success metrics to measure the effectiveness of your approach. Examples of success metrics are readability, consistency, coherence, and cohesion.

Support for the Dutch Language

This reading assistant is intended for use within Dutch governmental organisations. Thus, you should design a system that focuses on the Dutch language. This involves using and potentially fine-tuning models specifically on Dutch-language data to ensure accurate and contextually relevant simplifications. Your team should investigate potential sources for training, fine-tuning and validation. Table 1 shows some interesting resources provided by Kim et al., 2024 that can serve as a starting point for your research.

Table 1: List of resources (Kim et al., 2024).

Description	Size	Source	Domain	Link
Parallel corpus	1,311 sentence pairs	Manual	Government	GitHub
Parallel corpus	1,267 sentence pairs	Automatic (LLM)	unknown	Huggingface
Parallel corpus	2.87M paragraph pairs	Automatic (LLM)	Wikipedia	Huggingface
Contextualized lexical simplifications	96 sentences	Manual	Government	GitHub
Complex words and simpler alternatives	~800 words / expressions	Manual	Government	City of Amsterdam
Complex words and simpler alternatives	~130 words / expressions	Manual	Legal	City of Amsterdam
Words and frequency distributions graded on CEFR levels	17,743 words / expressions	Automatic	n/a	NT2Lex
<i>Texts graded on CEFR levels</i>	1,200 texts	Manual	Various	EDIA
<i>Dutch-English cognates and homographs</i>	~200 words	Manual	n/a	OSF

Sustainability

Consider the environmental and computational impact of the chosen models. Large language models (LLMs) are well-known for their size. They often involve hundreds of billions of parameters, which demand extensive computational resources and have a substantial environmental footprint. Thus,

balancing performance and efficiency can be achieved by selecting appropriately scaled models for the specific problem.

4. Expected outcomes

- A functional prototype that helps Oostzaan employees rewrite or check letters based on their writing guide.
- A final presentation/demo for Oostzaan and the HvA in week 10.
- A group final report with recommendations for Oostzaan and other municipalities on adopting GenAI in letter writing.

5. Key dates

- 10th of November, 09:00 – 12:00 (Design Workshop/Coaching Slot - Internal Kick-off)
- 13th of November, 14:00 – 16:00 (Tech workshop slot) - Kick-off with the Gemeente Oostzaan.
- 21st of November, 17:00 – Deadline for groups to send their specific questions to the Gemeente Oostzaan via DLO
- 1st of December, 09:30 – 11:30 (Design Slot) - Halfway meeting; students should prepare their stands to welcome the Gemeente Oostzaan guests.
- 06th of January, 17:00 (Questions) - Deadline for groups to send their specific questions to the Gemeente Oostzaan via DLO
- 29th January, 13:00 – 15:00 - Demo Day.

6. Next steps

This report aims to outline the comprehensive project scope for Block 2. The new group of students is responsible for choosing the use case and framing a more specific research question and project scope.

References

Over Gemeenten. 2025. Schrijfwijzer (accessible on DLO)

City of Amsterdam. 2024. Schrijfwijzer. Accessed: 08-November-2025.

Kim, J., Leijnen, S., & Beinborn, L. (2024). Considering Human Interaction and Variability in Automatic Text Simplification. Third Workshop on Text Simplification, Accessibility and Readability.

Mischa Corsius, Vera Lange, Yvette Linders, Henk Pander Maat, Els van der Pool, Nina Sangers, Keun Sliedrecht, Wouter Sluis-Thiescheffer, and Charlotte Swart. 2022. Monitor Begrijpelijkheid Overheidsteksten (Monitor Understandability of Government Texts). Accessed: 08-November-2025.

Netherlands Court of Audit. 2016. Aanpak van laaggeletterdheid (Tackling low literacy). Accessed: 04-November-2024.