Assessing the effectiveness of ChatGPT-3.5 and ChatGPT-4o in simplifying Italian institutional texts

Claudia Gigliotti (Università di Firenze) & *Mariachiara Pascucci* (Università di Pisa/Universität Basel)

claudia.gigliotti(at)unifi.it, mariachiara.pascucci(at)phd.unipi.it

Abstract

This research aims to describe the performance of ChatGPT-3.5 and ChatGPT-4o in the task of Automatic Text Simplification (ATS) in Italian institutional texts. The aim is to analyse the linguistic differences between the original texts compared to their simplified rewritings by ChatGPT, and the impact of these differences on non-expert users' experience. A dataset of six short texts was compiled to be rewritten using a zero-shot instructional prompt. The methodological approach combined quantitative linguistic analyses, manual analysis and human judgment to assess the effectiveness of the simplification. For the quantitative linguistic analysis, an additional comparison was made between ChatGPT's rewritings and human revisions, used as an external benchmark to better contextualize the AI's simplification strategies. The study provides new insights into the linguistic structure of administrative-bureaucratic texts by examining readability parameters and collecting subjective assessments of comprehension and perceived comprehensibility. It also aims to contribute to the growing body of research on text simplification methods and the role of large language models (LLMs) in enhancing accessibility to complex institutional discourse.

Keywords

Large Language Models, ChatGPT, bureaucratic and professional texts, linguistic simplification, human evaluation

1 Introduction

This study investigates the effectiveness of ChatGPT in performing Automatic Text Simplification (ATS) (Shardlow 2014; Saggion 2017; Al-Thanyyan and Azmi 2021), with a focus on Italian institutional texts. Specifically, the study aims to (1) identify the syntactic and lexical transformations introduced during the simplification process and evaluate their impact on structural features that influence readability; and (2) assess how automatic simplification affects judgments among non-expert users. The aim of the study is to investigate the interplay between readability, comprehension[1] and comprehensibility (a more comprehensive theoretical discussion is provided in Section 2). To this end, a dataset of six short institutional-administrative texts was compiled and simplified using a zero-shot instructional prompt. The linguistic analysis involved extracting and examining key features from both the original and simplified texts. This provided quantitative data to assess readability at the lexical and syntactic levels, focusing on features that are

[1] In this paper, the terms *comprehension*/*understanding* and *comprehensibility*/*understandability* are employed interchangeably. The preference for one term over another reflects linguistic variation rather than a conceptual distinction. More precise definitions will be provided in the following sections.

AILing

Claudia Gigliotti & Mariachiara Pascucci.
Assessing the effectiveness of ChatGPT-3.5 and ChatGPT-4o in simplifying Italian institutional and administrative texts.
Special Issue: *AI-Driven Linguistic Studies*, Vol. 2. No.2
DOI: 10.62408/ai-ling.v2i2.18
ISSN: 2943-0070

AI-Linguistica. Linguistic Studies on AI-Generated Texts and Discourses

commonly used to measure the effects of simplification (Fiorentino and Ganfi 2024). A survey inspired by the work of De Mauro and Vedovelli (1999), was carried out to investigate whether different versions of the same text influence the readers' comprehension and their judgment of subjective comprehensibility (Friedrich and Heise 2025). Prior to the survey, a manual examination of the rewritten texts was conducted to assess the fidelity of information transmission. The survey's experimental design was developed drawing on principles from eye-tracking methodology (Conklin, Pellicer-Sánchez and Carrol 2018; Godfroid 2020). This study is closely linked to Tavosanis (2025),[2] as we used six of the eight texts used by Tavosanis (2025) and the same prompt. In addition, the same rewritings of the author were used for the ChatGPT-3.5 model, while the ChatGPT-4o model was used to produce new rewritings updated in August 2024 and queried via chat.

The investigation has been conducted in different stages, one author focused on the quantitative linguistic analysis of both the original texts and their rewritten versions, while the other author dealt with quality of rewritings and human evaluation.[3] Although we recognize that the aspects we examine alone cannot fully account for a text's communicative effectiveness, our goal is to present some initial findings that may serve as a useful starting point for future research.

2 Theoretical Framework

Institutional texts play a crucial role in enabling citizens to participate in public life. A substantial body of research has examined the features of institutional language (Raso 2005; Viale 2008; Gualdo and Telve 2011; Lubello 2014, 2017; Vellutino 2018; Cortelazzo 2021; Piemontese 2023). The label *italiano istituzionale* is broad and multifaceted. In this context, we refer specifically to administrative texts, setting aside – at least for now, and leaving to other studies – the other side of institutional discourse: legal-normative texts, which pursue different communicative goals and exhibit distinct linguistic traits. Throughout this work, references to "institutional texts" should therefore be understood as referring exclusively to administrative texts. The main features of administrative language are described in Cortelazzo (2021), among others. These texts are often criticized for their excessive complexity, as evidenced by the numerous governmental initiatives and guidelines aimed at improving the drafting of administrative documents (Codice di stile 1993; Fioritto 1997; Cortelazzo and Pellegrino 2002, 2003; Franceschini and Gigli 2003; ITTIG/Accademia della Crusca 2011). Enhancing clarity is a key objective across various research domains, including institutional communication, and it is also one of the abilities currently being evaluated in Large Language Models such as ChatGPT. The application of AI-

[2] The author used the same texts employed for the present work to conduct a human evaluation to assess the effectiveness of the simplification process by ChatGPT-3.5. Tavosanis (2025) rated both AI and human-simplified rewritings based on five criteria: (1) accuracy; (2) linguistic correctness; (3) clarity; (4) improvement; (5) information preservation.

[3] Sections 3-4 were edited by Mariachiara Pascucci, Sections 5-7 are the work of Claudia Gigliotti. The authors gratefully acknowledge James S. Hanlon from ELTC staff (University of Sheffield) for his assistance and availability in serving as a native English language reviewer.

based automatic text simplification (ATS) techniques to bureaucratic and administrative texts is an emerging area of research focused on simplifying complex technical language (Cherubini et al. 2023; Paci et al. 2024). As Tavosanis (2024) notes, there are currently no widely accepted evaluation frameworks for generated texts and, more generally, for textual clarity. Although extensive literature provides guidance on clear writing (Cortelazzo 2021; Fiorentino and Ganfi 2024), a set of commonly accepted criteria for assessing textual clarity has not yet been established.

It is worth noting that the obstacles that make a text unclear can be of various kinds. Administrative language is often considered problematic due to the way institutional texts are typically written; however, writing difficulties inevitably affect reading, which in turn impacts comprehension. As stated by Piemontese (1996: 109) "la chiarezza è prodotta dall'azione simultanea di leggibilità e comprensibilità" ['Clarity is produced by the simultaneous action of readability and comprehensibility'; our transl.[4]]. Readability and comprehensibility (*leggibilità* and *comprensibilità* in Italian) are two terms sometimes used interchangeably; however, it is useful to distinguish between the two levels (Piemontese 1996:105). The comprehensibility of a text does not coincide with its linguistic readability which, instead, refers to automatic measures that allow us to measure it through specific readability formulas (Benjamin 2012; Collins-Thompson 2014; Vajjala 2021). In this study (in line with Piemontese's point of view), readability is the dimension assessed through the linguistic analysis of selected parameters, and it refers to objective and quantifiable aspects of the text, so-called surface obstacles. Conversely, comprehension involves the readers' ability to construct a mental representation appropriate to the content and communicative function of the text (Kintsch 1988, 1998; Mayer 2014; Schnotz 2014). Although comprehension can be influenced by linguistic factors such as word frequency, syntax complexity, and text cohesion (McNamara et al. 2014; Reed and Kershaw-Herrera 2016), it also depends on reader characteristics such as prior knowledge, vocabulary, reading goals, interest, and working memory capacity (Friedrich and Heise 2025). Therefore, comprehensibility pertains to qualitative aspects – deep obstacles – which are often linked to the reader. As reported in previous studies (Friedrich and Heise 2025), comprehensibility concerns six features that make a text more comprehensible to a certain reader, and they are: (1) the difficulty of word; (2) the difficulty of sentence; (3) the effort needed for reorganizations; (4) clarity of representation; (5) variety of language use; and (6) subjective comprehensibility. The last one refers to a global judgment of the text's comprehensibility for the readers and it depends on how well they think they understood the text. While controlling the linguistic aspects of a text can facilitate the reader's performance, the reference literature seems to converge towards the idea that subjective judgments are the only way to assess the ease of processing (Reber and Greifeneder 2017; Friedrich and Heise 2025).

The present study is structured in two phases: the first analyses institutional texts' readability by comparing the original versions with the rewritings through a quantitative linguistic analysis of surface features; the second analyses the quality of the rewritings' transmitted information and focuses on evaluating differences in

[4] All translations provided in this work are ours.

terms of comprehension and comprehensibility through subjective judgments by non-expert readers. This dual focus allows us to investigate how textual simplification affects not only the structural complexity of administrative texts, but also their perceived clarity and ease of understanding.

3 Methodology

3.1 Dataset

For the purposes of this study, a small dataset of six short texts was compiled to represent key features of Italian institutional-administrative language. Four of these texts were drawn from the CITRIN-LA corpus (*Corpus Italiano Testi Regolativi Informativi Normativi – Lingua Amministrativa*). It is a self-assembled and continuously expanding corpus of Italian institutional and administrative documents, compiled by the authors of this study. It currently contains over one million words and is organized into three subcorpora, based on the taxonomy proposed by Sabatini (2012), which distinguishes between informative, normative, and regulative texts. CITRIN-LA is a private corpus, and it is not accessible online now. The texts analysed in this study are drawn from the subcorpus of regulative texts, which includes documents outlining procedures, protocols, and operational regulations. These texts are intended both for internal communication between offices and departments, and for external communication directed to the public. The selected texts consist of full paragraphs taken from official ministerial guidelines and address topics such as researcher mobility, infrastructure security, and the distribution of public funds. The selected guidelines were produced between 2018 and 2024. Two additional texts, drawn with minor modifications from Cassese's *Codice di Stile* (1993), were included as exemplary and comparable samples due to their similarity in text type and communicative intent. Each text ranges in length from approximately 200 to 500 words. Table 1 presents the original texts, each identified by a code, along with a brief description of its content and the source from which it was taken.

Table 1: Selection of the original texts for the construction of the small dataset.

Code	Topic	Source
CASS-1	*Bando comunale – Vacanze persone anziane.* 'Municipal announcement – Holidays for seniors.'	*Codice di Stile*
CASS-4	*Delibera di un Consiglio Circoscrizionale.* 'Resolution of a District Council.'	*Codice di Stile*
MOB-1	*Guida operativa per i beneficiari azione i.2 "attrazione e mobilità dei ricercatori".* 'Operational guide for beneficiaries of action i.2 "attraction and mobility of researchers".'	CITRIN-LA
PONTI-1	*Linee guida per la classificazione e gestione del rischio, la valutazione della sicurezza ed il monitoraggio dei ponti esistenti.* 'Guidelines for risk classification and management, safety assessment and monitoring of existing bridges.'	CITRIN-LA
PRIN-4 PRIN-5	*Linee guida per la rendicontazione destinate ai soggetti attuatori degli interventi del PNRR Italia di cui il Ministero dell'Università e della Ricerca è amministrazione titolare.* 'Guidelines for reporting intended for the implementers of the interventions of the PNRR Italia of which the Ministry of University and Research is the titular administration'	CITRIN-LA

To ensure the conditions required for human evaluation, the texts were selected based on the criterion of content unfamiliarity – that is, the topics fall outside the participants' social background and prior knowledge. This approach aimed to ensure that comprehension and perceived comprehensibility would rely solely on textual content information provided within the text itself, rather than on prior experience or subject-matter expertise. From each text, we selected portions of text that clearly provided practical instructions to citizens. The six original texts, the six ChatGPT-3.5 rewritings and the six ChatGPT-4o rewritings returned a dataset of 18 items.[5] They were organized into three different experimental lists based on the type of text, and each list corresponds to a distinct experimental condition as follows in Table 2.

Table 2: Lists and number of items per condition.

Experimental lists	Conditions (text type)	Items
List 1	*Original texts*	6 items
List 2	*ChatGPT-3.5 rewritings*	6 items
List 3	*ChatGPT-4o rewritings*	6 items

3.2 Prompt and human rewritings

The prompt employed for the rewritings was intended to simplify the text in a general manner, without focusing on specific or well-defined linguistic features. Given that the selected texts are addressed to ordinary citizens with no expertise, the intention was to simulate the experience of a non-expert user and to observe the outcomes that a generic, non-specialized prompt might generate.

[5] From now on, *items* will mean all the elements of the experimental material, therefore the original texts and their respective rewritings.

The ATS was guided by the aim of preserving content integrity while enhancing linguistic clarity. It employed a so-called *zero-shot instructional prompt* (Efrat and Levy 2020; Mishra et al. 2022), with which the user provides explicit instructions to guide the LLM's response. Unlike general or open-ended questions, instructional prompts are specific, directing the LLM not just on what information is needed but also on how it should be presented. The prompt was:

(1) *Puoi semplificare la forma linguistica del seguente testo amministrativo-burocratico pur mantenendo tutti i dettagli del contenuto? Voglio che il testo prodotto sia dettagliato e lungo tanto quanto il testo da semplificare che è qui tra virgolette* "[testo originale]".
'Can you simplify the linguistic form of the following administrative-bureaucratic text while maintaining all the details of its content? I want the text produced to be as detailed and long as the text to be simplified that is here in quotation marks "[original text]"'.

Despite its simplicity, the prompt proved suitable for addressing a fundamental requirement of the evaluation as it generated rewritings that – as much as possible[6] – preserved a similar size to the originals in terms of word count, as shown in Table 3.

Table 3: Text length (in words).

Code	Original	ChatGPT-3.5	ChatGPT-4o
CASS-1	329	271	249
CASS-4	298	277	257
MOB-1	366	308	303
PONTI-1	575	467	409
PRIN-4	241	222	207
PRIN-5	292	255	221

For the readability text we decided to compare the results with the one of human rewritings to have a comparison with other rewriting methodologies. The rewritings were made by Mariachiara Pascucci following guidelines such as *Guida* (ITTIG/Accademia della Crusca) and the recommendations of Cortelazzo (2021).

Human rewritings were included in the linguistic analysis as an external benchmark to compare AI simplification against real-world rewriting strategies. However, due to methodological constraints, human rewritings were not included in the comprehension survey. Future research will address this limitation by incorporating a broader set of rewritten versions in the experimental phase.

[6] Based on observations from the prompt testing phase, we found that when ChatGPT is instructed to simplify a text it tends to reduce the overall word count automatically. In this study, although the reduction in average was approximately between 10%-20%, the selected prompt was considered the most appropriate for ensuring the desired experimental conditions.

4 Linguistic quantitative analysis

The linguistic data were extracted by the software READ-IT (Dell'Orletta et al. 2011). The texts were analysed following five key linguistic parameters selected from the literature on so-called clear writing (Piemontese 1996) and on clarity in administrative language (Cortelazzo 2021; Fiorentino and Ganfi 2024). The selected parameters are listed in Table 4.

Table 4: Selected parameters.

(1) Average Sentence Length (in tokens)
(2) Percentage of Subordinate Clauses
(3) Average Number of Clauses Per Sentence
(4) Average Number of Words Per Clauses
(5) Percentage of Basic Vocabulary

A first parameter useful for describing texts from a syntactic perspective is *average sentence length* (1). The literature on administrative language and clear writing consistently emphasizes that shorter sentences contribute to greater textual accessibility (Piemontese 1996). For the second parameter, Fioritto (1997) and Piemontese (1996) recommend limiting *subordinate clauses* (2) to improve clarity: the parameter helps to determine the degree of syntactic complexity, since it is generally assumed that the greater the number of subordinates, the greater the difficulty of comprehension. The *average number of clauses per sentence* (3) is a metric, highlighted by Korzen (2022), that is often used to measure text complexity. Similarly, the fourth parameter – the *average number of words per clause* (4) – is also linked to readability. In this case, fewer words per clause usually indicate a simpler and more accessible text. The *use of words belonging to the basic vocabulary*[7] (*vocabolario di base*) (5) is widely recognized to enhance text accessibility (De Mauro 1980; De Mauro and Chiari 2016). For this reason, it was evaluated in the present analysis to assess its impact on readability: the higher the percentage of lemmas from the basic vocabulary in a text, the more comprehensible it is expected to be. These parameters help assess how accessible the language is to the "general public". According to the cited literature, improvements in readability are generally associated with decreases in parameters (1) – (4) and an increase in parameter (5). The analysis was carried out on each of the three sets of short texts, both collectively and individually. This approach made it possible to identify general patterns in language use across the dataset, while also enabling a more detailed examination of the specific features of each text. The results of the aggregated data extraction are presented in Table 5. Further details on the individual texts can be found in Appendix 2.

[7] The Italian *vocabolario di base* ('basic vocabulary'), first introduced in De Mauro (1980) and most recently updated in De Mauro and Chiari (2016), brings together two categories of words into a unified set: (1) those most frequently used in the language at a given historical moment, as identified through frequency dictionaries; and (2) words that, while less frequent in actual usage, are nonetheless perceived by speakers as being equally – or even more – accessible than the most common terms. The basic vocabulary includes approximately 7,000 lemmas, grouped into three bands: fundamental, high-frequency, and high-availability words. All are considered part of the core lexicon expected to be known by speakers with a basic level of education.

Table 5: Results of the extracted linguistic data for each text type.

Parameters	Original	ChatGPT-3.5	ChatGPT-4o	Human
(1) Average Sentence Length (in tokens)	43,4	29,9	23,1	23,7
(2) Percentage of Subordinate Clauses	35,9%	37,4%	43%	35,1%
(3) Average Number of Clauses Per Sentence	3,8	2,3	2,2	2,5
(4) Average Number of Words Per Clause	11,5	10,7	10,6	9,6
(5) Percentage of Basic Vocabulary	59,7%	61,40%	62,0%	65,1%

The results show that ChatGPT-4o rewritings tend to outperform ChatGPT-3.5 in most of the parameters selected for this study. Both models reduce the average sentence length (1) and increase basic vocabulary usage (5), though ChatGPT-4o achieves greater simplification overall. When comparing AI-generated rewritings to human ones, ChatGPT-3.5 surpasses human ones in only a few cases, mainly in parameters (1) and (5). Human rewritings outperform ChatGPT-4o for three parameters: percentage of subordinate clauses (2), average number of words per clause (3), and percentage of basic vocabulary lemmas (5).

4.1 Average sentence length

A closer examination of the five selected parameters shows that average sentence length, measured in tokens (1), decreases in the ChatGPT-3.5 rewritings compared to the original texts, though the reduction is generally less marked than in ChatGPT-4o. For instance, in PRIN-4, ChatGPT-3.5 reduces the average sentence length from 33.6 to 31.4 tokens, and in PONTI-1 from 63.2 to 47.1 tokens. ChatGPT-4o consistently produces shorter sentences than both the original texts and the ChatGPT-3.5 versions, as shown in Table 6 and Table 7. However, the human rewritings display the most consistent and substantial reductions overall (up to -57.3% in PONTI-1 and -44.1% in MOB-1), apart from CASS-1, where the change is less significant. This suggests that, although the models are capable of effective simplification, human strategies remain more consistent and effective in reducing sentence length.

Table 6: Average sentence length (in tokens) across original and rewritings.

Text	Originals	ChatGPT-3.5	ChatGPT-4o	Human
CASS-1	32,5	12,6	14,1	25,5
CASS-4	33,4	30,9	26,0	21,7
MOB-1	45,6	38,7	38,6	25,5
PONTI-1	63,2	47,1	29,2	27,0
PRIN-4	33,6	31,4	30,2	23,8
PRIN-5	56,3	50,3	30,0	25,0

Table 7: Percentage variation in average sentence length (tokens).

Text	**ChatGPT-3.5** Δ%	**ChatGPT-4o** Δ%	**Human** Δ%
CASS-1	-62.2%	-56.6%	-21.5%
CASS-4	-7.5%	-22.2%	-35.0%
MOB-1	-15.1%	-15.3%	-44.1%
PONTI-1	-25.5%	-53.8%	-57.3%
PRIN-4	-6.5%	-10.1%	-29.2%
PRIN-5	-10.7%	-46.7%	-55.6%

This trend is well illustrated by the following example from PONTI-1, where the sentence length decreases progressively across the different versions of the same text passage: from 90 tokens (80 words) in the original, to 70 tokens (63 words) in the ChatGPT-3.5 rewrite, to 48 (43 words) in ChatGPT-4o, and 41 (36 words) in the human rewriting.

(2a) *Originale: Qualora le indagini svolte evidenzino fenomeni e/o difetti rilevanti, quali importanti stati di corrosione o rotture, anche parziali, dei cavi da precompressione, o nel caso in cui si ritenga che le ispezioni speciali non siano sufficienti a definire con adeguato grado di affidabilità lo stato di conservazione generale dell'opera ed il quadro completo dei difetti, occorre procedere a valutazioni di sicurezza approfondite, previste dal Livello 4 dell'approccio multilivello, assumendo come riferimento la parte delle presenti linee guida ad esse dedicate.*
'Original: If the investigations carried out reveal significant phenomena and/or defects, such as major states of corrosion or even partial breakage of the prestressing cables, or in the case where it is deemed that the special inspections are not sufficient to define with an adequate degree of reliability the general state of preservation of the structure and the complete picture of the defects, it is necessary to proceed with in-depth safety assessments, as provided for by Level 4 of the multi-level approach, taking as a reference the part of these guidelines dedicated to them.'

(2b) *ChatGPT-3.5: Se le ispezioni evidenziano fenomeni o difetti significativi, come gravi stati di corrosione o rotture parziali dei cavi di precompressione, o se si ritiene che le ispezioni speciali non siano sufficienti a valutare adeguatamente lo stato generale della struttura e i difetti presenti, è necessario eseguire valutazioni di sicurezza più approfondite, come previsto dal Livello 4 dell'approccio multilivello, seguendo le linee guida appropriate.*
'ChatGPT-3.5: If the inspections highlight significant phenomena or defects, such as severe states of corrosion or partial breakages of the prestressing cables, or if it is considered that the special inspections are not sufficient to adequately assess the general condition of the structure and the defects present, it is necessary to carry out more in-depth safety assessments, as provided by Level 4 of the multi-level approach, following the appropriate guidelines.'

(2c) *ChatGPT-4o: Se queste indagini rivelano gravi difetti, come importante corrosione o rottura dei cavi di precompressione, o se le ispezioni speciali non definiscono adeguatamente le condizioni generali del ponte, si devono fare valutazioni di sicurezza avanzate di Livello 4, seguendo le linee guida specifiche.*
'ChatGPT-4o: If these investigations reveal serious defects, such as significant corrosion or breakage of the prestressing cables, or if the special inspections do not adequately define the general condition of the bridge, advanced Level 4 safety assessments must be carried out, following the specific guidelines.'

(2d) *Riscrittura umana: Se le indagini svolte evidenziano fenomeni e/o difetti rilevanti, o nel caso in cui si ritenga che le ispezioni speciali non siano sufficienti, occorre procedere a valutazioni di sicurezza approfondite, previste dal Livello 4 dell'approccio multilivello.*
'Human rewrite: If the investigations carried out highlight significant phenomena and/or defects, or in the case where it is considered that the special inspections are not sufficient, it is necessary to proceed with in-depth safety assessments, as provided by Level 4 of the multi-level approach.'

These examples suggest that sentence shortening – though varying in degree – emerges as a shared simplification strategy across all rewriting methods. The human version shows the most substantial reduction in this case. Similarly, ChatGPT-4o demonstrates the ability to reduce sentence length without compromising the core message, indicating a more context-sensitive simplification process. In contrast, ChatGPT-3.5 adopts a more mechanical approach, often simplifying through linear shortening.

4.2 Subordinate clauses

While a general tendency can be observed – namely, that ChatGPT-3.5 tends to maintain values close to the original texts, and both ChatGPT-4o and human rewriting often increase the use of subordination – the variation across datasets is substantial, as shown in Table 8 and Table 9. This pattern suggests that, contrary to standard simplification guidelines, both ChatGPT and human rewriters may increase subordinate clauses to improve clarity.

Table 8: Percentage of subordinate clauses across originals and rewritings.

Text	Originals	ChatGPT-3.5	ChatGPT-4o	Human
CASS-1	16,7%	11,1%	30,8%	41,2%
CASS-4	42,3%	44,4%	54,5%	61,2%
MOB-1	47,1%	46,2%	42,9%	41,5%
PONTI-1	41,2%	41,2%	50%	38,5%
PRIN-4	27,3%	29,9%	42,9%	37,5%
PRIN-5	22,2%	45,5%	25,5%	10%

Table 9: percentage variation in the proportion of subordinate clauses.

Text	**ChatGPT-3.5** Δ%	**ChatGPT-4o** Δ%	**Human** Δ%
CASS-1	-33,5%	84,4%	146,7%
CASS-4	5,0%	28,8%	44,7%
MOB-1	-1,9%	-8,9%	-11,9%
PONTI-1	0%	21,4%	-6,6%
PRIN-4	9,5%	57,1%	37,4%
PRIN-5	105%	14,9%	-55%

The percentage differences reveal distinct rewriting strategies across models. ChatGPT-3.5 tends to maintain levels of subordination close to the originals, showing only minor variations in most cases. ChatGPT-4o, on the other hand, more frequently increases the use of subordinate clauses suggesting a tendency to restructure sentences for improved logical flow or cohesion. Human rewritings exhibit the most variable behavior: in texts like CASS-4 and CASS-1, subordination increases substantially, while in PRIN-5 it decreases (–54.9%). These patterns highlight the flexibility and adaptiveness of human interventions and suggest that ChatGPT-4o is more likely than ChatGPT-3.5 to adopt structural modifications aimed at enhancing textual clarity through subordination, as we can see in the following example extracted from PRIN-4. In this text, the percentage of subordinate clauses rises from 27.3% in the original to 42.9% in the ChatGPT-4o version. This happens partly due to restructuring choices that improve textual clarity:

(3a) *Originale: Secondo le disposizioni impartite dal MEF, il Soggetto Attuatore/Beneficiario, a fronte di esito positivo delle verifiche condotte dalla Direzione Generale del MUR sulla documentazione trasmessa attraverso la piattaforma informatica "PRIN", è tenuto a questo punto a generare il Rendiconto di Progetto direttamente sul sistema MEF ReGiS.*
'Original: According to the provisions issued by the MEF, the Implementing Body/Beneficiary, upon a positive outcome of the checks carried out by the Directorate General of the MUR on the documentation submitted via the "PRIN" platform, is required at this stage to generate the Project Report directly in the MEF ReGiS system.'

(3b) *ChatGPT-4o: Secondo le regole del Ministero dell'Economia e delle Finanze (MEF), il soggetto responsabile del progetto, dopo aver ricevuto l'ok dalla Direzione Generale del Ministero dell'Università e della Ricerca (MUR) sulla documentazione inviata tramite la piattaforma 'PRIN', deve creare il Rendiconto di Progetto nel sistema MEF ReGiS.*
'ChatGPT-4o: According to the rules of the Ministry of Economy and Finance (MEF), the project manager, after receiving approval from the Directorate General of the Ministry of University and Research (MUR) on the documentation submitted through the 'PRIN' platform, must create the Project Report in the MEF ReGiS system.'

Here, the temporal relation expressed in the original ("a fronte di esito positivo…") is rendered as a temporal subordinate clause ("dopo aver ricevuto…"), which enhances the explicitness of the logical relation while at the same time increasing the number of subordinate clauses. A similar shift occurs at the end of the text:

(4a) *Originale: Nel caso di richieste di integrazioni o chiarimenti la procedura di contraddittorio avviene per il tramite della DG competente.*
'Original: In the case of requests for additions or clarifications, the adversarial procedure takes place through the competent Directorate General.'

(4b) *ChatGPT-4o: Se ci sono richieste di integrazioni o chiarimenti, queste avvengono tramite la Direzione Generale competente.*
'ChatGPT-4o: If there are requests for additions or clarifications, these are handled by the competent Directorate General.'

Overall, these results indicate that the increase in subordinate clauses often reflects strategic choices aimed at clarifying logical relations, as seen in PRIN-4, where subordination replaces less explicit prepositional or nominal structures. This suggests that a higher rate of subordination may, in certain contexts, enhance readability rather than hinder it.

This pattern sheds light on ChatGPT's evolving simplification strategies. While ChatGPT-3.5 generally maintains subordination levels close to the original texts, ChatGPT-4o occasionally increases them. This suggests that newer models are not simply reducing complexity mechanically; rather, they appear to apply more nuanced transformations aimed at improving clarity. These choices reveal a shift from rigid simplification toward more context-sensitive editing, a tendency that aligns them more closely with human rewriting practices.

A more fine-grained qualitative analysis of the types and functions of subordinate clauses would be required to fully understand how subordination is handled across different rewriting strategies – namely, human, ChatGPT-3.5, and ChatGPT-4o. Although such an investigation exceeds the scope of the present study, it offers a promising avenue for future research.

4.3 Clauses per sentence

While the general trend shows that ChatGPT-3.5 and ChatGPT-4o tend to reduce the number of clauses per sentence – thus simplifying sentence structures as seen in PONTI-1 (from 5.8 in the original to 4.6 and 3.3, respectively) – this pattern is not entirely uniform across datasets. In PRIN-4 and PRIN-5, for example, the average number of clauses per sentence slightly increases in the ChatGPT rewritings, as shown in Table 10 and 11. Human rewritings show a more varied pattern. In some cases, such as CASS-4 and PONTI-1, they result in a significantly lower number of clauses per sentence compared to both the original and AI versions. In other cases, like PRIN-4 and MOB-1, the number of clauses remains relatively high, though still below the original.

Table 10: Average number of clauses per sentence across originals and rewritings.

Text	Originals	ChatGPT-3.5	ChatGPT-4o	Human
CASS-1	2,6	1,1	1,0	3,0
CASS-4	3,7	3,3	3,0	1,7
MOB-1	4,4	3,1	3,2	3,0
PONTI-1	5,8	4,6	3,3	2,8
PRIN-4	3,6	4,0	3,7	3,0
PRIN-5	2,3	3,1	2,4	1,7

Table 11: percentage variation of clauses per sentence.

Text	ChatGPT-3.5 Δ%	ChatGPT-4o Δ%	Human Δ%
CASS-1	-57.69%	-61.54%	15.38%
CASS-4	-10.81%	-18.92%	-54.05%
MOB-1	-29.55%	-27.27%	-31.82%
PONTI-1	-20.69%	-43.1%	-51.72%
PRIN-4	11.11%	2.78%	-16.67%
PRIN-5	34.78%	4.35%	-26.09%

As already discussed, the parameter of average number of clauses per sentence is often used to measure text complexity. However, an increase in this parameter does not necessarily correspond to reduced clarity in our sample. In several cases, it reflects the transformation of nominal constructions into more explicit verbal forms, which may improve textual transparency rather than hinder it, as we can see from the following Example from PRIN-4:

(5a) *Originale: Unitamente al Rendiconto, il Soggetto Attuatore dovrà altresì confermare di aver svolto i controlli sopra richiamati mediante l'inserimento di appositi flag e caricare anche su ReGiS la documentazione a comprova già fornita al MUR […].*
'Original: Together with the Report, the Implementing Body shall also confirm that it has carried out the aforementioned checks by means of the insertion of appropriate flags and shall also upload to ReGiS the supporting documentation already provided to the MUR […].'

(5b) *ChatGPT-3.5: Insieme al Rendiconto, il Soggetto Attuatore dovrà anche confermare di aver effettuato i controlli sopra menzionati inserendo appositi flag e caricando su ReGiS la documentazione già fornita al MUR […].*
'ChatGPT 3.5: Along with the Report, the Implementing Body must also confirm that it has performed the above-mentioned checks by inserting appropriate flags and uploading to ReGiS the documentation already provided to the MUR […].'

Here, the nominalization "inserimento" is replaced by the verbal form "inserendo," increasing the number of clauses by making the action explicit and shifting from a nominal to a verbal construction.

These findings suggest that both ChatGPT's models can reduce clause density when appropriate, but do not apply this strategy consistently across texts. ChatGPT-3.5 occasionally increases the number of clauses per sentence – likely due to shifts from compact nominal constructions to more explicit verbal forms. This indicates a form of simplification that prioritizes semantic transparency rather than purely reducing structural complexity. ChatGPT-4o, by contrast, tends to apply a more controlled and coherent approach. Human rewritings display a more varied pattern, likely reflecting a flexible adaptation to each text's communicative context. This suggests that while clause reduction can support simplification, it is not always a necessary condition for improved clarity.

4.4 Words per clauses

In general, ChatGPT-4o tends to produce more concise clauses than the original texts, as shown in Table 12 and Table 13. Notable reductions are observed in PONTI-1 (from 10.9 to 8.8), PRIN-4 (from 10.8 to 8.1), and PRIN-5 (from 24.1 to 12.4), the latter showing the most significant simplification. ChatGPT-3.5 also shows consistent reductions in several cases, though the effect is less pronounced. However, exceptions emerge. In CASS-1, ChatGPT-4o increases the number of words per clauses (from 12.3 to 15.7). Similarly, MOB-1 shows a modest increase with ChatGPT-3.5 (from 10.4 to 12.4). Human rewritings tend to produce the shortest clauses overall, particularly in MOB-1 (8.5) and CASS-4 (7.7). The percentage variations in words per clause reveal that ChatGPT-4o generally produces more concise clauses than the original texts, particularly in PRIN-5, where the reduction reaches nearly 49%. Conversely, ChatGPT-3.5 exhibits a less consistent pattern, as seen in MOB-1 (+19.2%). Human rewritings, on the other hand, show the most stable and substantial reductions across nearly all texts, indicating a more systematic approach. These findings indicate that ChatGPT models are generally effective at simplifying individual clause units, but not always consistently.

Table 12: Words per clauses across originals and rewritings.

Text	**Originals**	**ChatGPT-3.5**	**ChatGPT-4o**	**Human**
CASS-1	12,3	11,1	15,7	12,7
CASS-4	9,1	9,4	8,7	7,7
MOB-1	10,4	12,4	11,1	8,5
PONTI-1	10,9	10,2	8,8	9,8
PRIN-4	10,8	7,8	8,1	7,9
PRIN-5	24,1	15,9	12,4	14,3

Table 13: Percentage variation in words for clause.

Text	**ChatGPT-3.5** Δ%	**ChatGPT-4o** Δ%	**Human** Δ%
CASS-1	-9,76%	27,6%	3,25%
CASS-4	3,30%	-4,40%	-15,38%
MOB-1	19,23%	6,37%	-18,27%
PONTI-1	-6,43%	-19,27%	-10,09%
PRIN-4	-27,78%	-25,00%	-26,85%
PRIN-5	-34,02%	-48,55%	-40,66%

4.5 Lemmas belonging to the basic vocabulary

The percentage of lemmas belonging to the basic vocabulary generally increases in the ChatGPT-3.5 rewritings. For example, in PRIN-4, this value rises from 61.8% to 68.9%. ChatGPT-4o further increases the use of basic vocabulary in nearly all cases, achieving the highest scores across datasets, as shown in Table 14 and Table 15.

Table 14: percentage of lemmas belonging to the basic vocabulary across originals and rewritings.

Text	**Originals**	**ChatGPT-3.5**	**ChatGPT-4o**	**Human**
CASS-1	61,9%	63,2%	64,6%	75,8%
CASS-4	77,1%	79,9%	79,1%	61,4%
MOB-1	66,5%	66,9%	66,3%	75,8%
PONTI-1	65,2%	65,2%	72,3%	67,4%
PRIN-4	61,8%	68,9%	73,4%	66,7%
PRIN-5	54,2%	53,4%	51,0%	67,6%

Table 15: percentage variation in lemmas belonging to the basic vocabulary.

Text	**ChatGPT-3.5** Δ%	**ChatGPT-4o** Δ%	**Human** Δ%
CASS-1	2,1%	4,36%	22,46%
CASS-4	3,63%	2,59%	-20,36%
MOB-1	0,6%	-0,3%	13,98%
PONTI-1	0%	10,89%	3,37%
PRIN-4	11,49%	18,77%	7,93%
PRIN-5	-1,48%	-5,9%	24,72%

These lexical modifications are immediately noticeable. Archaic or overly technical expressions – including adverbs such as *ivi* and *altresì* – are frequently replaced by ChatGPT with simpler and more accessible equivalents. This tendency can be observed in the following example from PRIN-4:

(6a) *Originale: Unitamente al Rendiconto, il Soggetto Attuatore dovrà* ***altresì*** *confermare di aver svolto i controlli sopra richiamati mediante l'inserimento di appositi flag e caricare anche su ReGiS la documentazione a comprova già fornita al MUR,* ***ivi*** *compresa l'Attestazione sull'esecuzione delle attività di autocontrollo amministrativo-contabile.*
'Original: Together with the Report, the Implementing Body shall also confirm that it has carried out the aforementioned checks by inserting the appropriate flags and shall also upload to ReGiS the supporting documentation already provided to the MUR, including the Certificate of execution of the administrative-accounting self-monitoring activities.'

(6b) *ChatGPT-3.5: Insieme al Rendiconto, il Soggetto Attuatore dovrà **anche** confermare di aver effettuato i controlli sopra menzionati inserendo appositi flag e caricando su ReGiS la documentazione già fornita al MUR, compresa l'Attestazione sull'esecuzione delle attività di autocontrollo amministrativo-contabile.*
'ChatGPT-3.5: Along with the Report, the Implementing Body must also confirm that it has performed the above-mentioned checks by inserting appropriate flags and uploading to ReGiS the documentation already provided to the MUR, including the Certificate of execution of the administrative-accounting self-monitoring activities.'

(6c) *ChatGPT-4o: Insieme al Rendiconto, il responsabile del progetto deve **anche** confermare di aver effettuato i controlli richiesti inserendo appositi flag e caricare in ReGiS la stessa documentazione inviata al MUR, compresa l'Attestazione delle attività di autocontrollo amministrativo e contabile.*
'ChatGPT-4o: Along with the Report, the project manager must also confirm that the required checks have been carried out by inserting the appropriate flags and uploading to ReGiS the same documentation sent to the MUR, including the Certificate of administrative and accounting self-monitoring activities.'

Overall, the increased use of basic vocabulary – particularly in the ChatGPT-4o rewritings – suggests a consistent tendency toward lexical simplification aimed at improving accessibility.

5 Evaluation survey

The survey was designed to focus on readers' global judgments on texts. Specifically, it considered comprehension as the ability to form a mental representation appropriate to the content of the text, and perceived comprehensibility as the judgment of how well the reader believes they have understood the text (Friedrich and Heise 2025). In line with the purpose of the survey, each item has undergone a linguistic adaptation to focus on comprehension and perceived comprehensibility.

As stated by Palermo (2013: 27), "*i significati potenziali di un testo vengono attualizzati, e acquistano un senso univoco, all'interno di uno specifico contesto*" ['The potential meanings of a text are actualized, and acquire an univocal sense, within a specific context']. To reduce the decontextualization effect of extracted passages, brief introductory sentences – no longer than 35 words – were added before each text. These introductions were intended solely to provide a minimal contextual frame necessary for the interpretation and to resolve possible ambiguities (e.g., clarify acronyms). The introduction was not part of the comprehension task itself. Participants were asked to focus their evaluation on the main text only. To minimise content-related effects, explicit references to laws or regulations were removed from the texts. Legal-juridical content requires specific knowledge and skills to be fully understood. This choice was made to limit the

influence of prior legal knowledge on comprehension performance. Care was taken to preserve the overall coherence and informational integrity of the texts despite these modifications. Particular attention was paid to balancing the length of the items. As a result, each item was reduced to approximately 62-165 words, and – when including the introductory context – each final item ranged between 90 and 190 words. Edits were applied uniformly to both the original and rewritten versions to maintain consistency in length and content. An example of linguistic adaptation is provided in Appendix 3. Overall, the simplified rewritings aimed to preserve the integrity of the original content. Nevertheless, during the manual analysis of the AI-generated rewritings, we identified three recurrent patterns: (I) loss of information; (II) lack of information; and (III) recomputation of incorrect information. The *loss of information* refers to cases in which an entire piece of information is omitted in the rewriting process. The *lack of information* involves the partial omission of content: the core information remains, but the missing elements may generate ambiguity for the reader. The *recomputation of incorrect information* occurs when the rewritten text loses accuracy, and the inaccurate transformation of information can lead the reader to a wrong interpretation. Such phenomena depend on the nature of the LLM and reflect how, operating through statistical approximation, it may rework segments of the original input until the output becomes inaccurate. Below is an example of the original text and its rewritings that carry these 'simplification errors':[8]

(7a) *Originale: […] Il Consiglio Circoscrizionale risolve, per motivi espressi in narrativa, di fissare i seguenti criteri al fine di consentire all'ufficio Tecnico circoscrizionale di essere autorizzato ad installare direttamente ovvero di autorizzare i cittadini ad installare barriere metalliche senza dover sottoporre le richieste al parere della Commissione circoscrizionale del Traffico.*
'Original: […] The District Council resolves, for reasons expressed in the narrative, to establish the following criteria in order to allow the District Technical Office to be authorized to directly install or to authorize citizens to install metal barriers without having to submit the requests to the opinion of the District Traffic Commission.'

(7b) *ChatGPT-3.5: […] Il Consiglio Circoscrizionale, in virtù delle ragioni esposte, decide di stabilire i seguenti criteri per consentire all'Ufficio Tecnico circoscrizionale di autorizzare direttamente l'installazione dei parapedonali o di concedere ai cittadini l'autorizzazione per farlo senza dover ricorrere al parere della Commissione del Traffico.*
'ChatGPT-3.5: '[…] The District Council, by virtue of the reasons set out, decides to establish the following criteria to allow the District Technical Office to directly authorize the installation of pedestrian barriers or to grant citizens authorization to do so without having to resort to the opinion of the Traffic Commission.'

[8] Full texts are presented in Appendix 3.

(7c) *ChatGPT-4o: […] Il Consiglio Circoscrizionale decide, per le ragioni sopra esposte, di stabilire i seguenti criteri per permettere all'Ufficio Tecnico circoscrizionale di autorizzare direttamente o ai cittadini di installare barriere senza chiedere sempre il parere della Commissione del Traffico.*
'ChatGPT-4o: '[…] The District Council decides, for the reasons stated above, to establish the following criteria to allow the District Technical Office to directly authorize or the citizens to install barriers without always asking the opinion of the Traffic Commission.'

(I) *Loss of information*: in the original text, it is specified that a request must be submitted to install metal barriers and both rewritings omit this requirement.
(II) *Lack of information*: in the ChatGPT-4o rewrite the non-repetition of the verb "to authorize" [the citizens] can generate ambiguity.
(III) *Recomputation of incorrect information*: in both ChatGPT's rewritings, the original reference to the 'District Traffic Commission' was simplified to 'Traffic Commission', and the action of submitting a formal request was inaccurately represented as an 'opinion'.

These 'simplification errors' were not corrected to observe their potential effects on the survey's results. The rewritten items were screened to ensure they remained answerable within the comprehension task. While ensuring that the correct response remained logically accessible despite the textual inaccuracies, their inclusion aimed to assess whether such formal deviations might influence participants' subjective judgments of global comprehensibility.

6 Comprehension and comprehensibility task

The comprehension task was based on typical comprehension errors identified in administrative texts, as documented in the *Guida alla redazione degli atti amministrativi* (ITTIG/Accademia della Crusca 2011). These errors fell into the following categories:

- *Misinterpretation of vocabulary,* due to polysemy or technical meanings of words.
- *Misinterpretation of the sequence,* also called *timeline error,* due to incorrect hierarchization of linguistic information.
- *Association of roles or characteristics of the subjects involved,* resulting from difficulty in identifying and distinguishing agents and actions in complex linguistic constructions.

Participants were required to choose a correct answer; below is an example illustrating an error type for incorrect association of roles or characteristics of the subjects involved – it is related to the previous 'simplification errors' examples, fully presented in Appendix 3. The example illustrates how, despite the preservation

of the identified linguistic phenomena in the rewritings, the correct answer to the comprehension task remained logically accessible to the reader:

(8) *Scegli la risposta corretta*:
- *Il Consiglio Circoscrizionale consente ai cittadini di installare i "parapedonali" per impedire il parcheggio alle auto.*
- *L'Ufficio Tecnico circoscrizionale consente al Consiglio Circoscrizionale di autorizzare i cittadini a installare i "parapedonali" per impedire il parcheggio alle auto.*
- *Il Consiglio Circoscrizionale consente all'Ufficio Tecnico circoscrizionale di autorizzare i cittadini a installare i "parapedonali" per impedire il parcheggio alle auto.*

'Choose the correct answer:
- The District Council allows citizens to install "pedestrian barriers" to prevent cars from parking.
- The District Technical Office allows the District Council to authorize citizens to install "pedestrian barriers" to prevent cars from parking.
- The District Council allows the District Technical Office to authorize citizens to install "pedestrian barriers" to prevent cars from parking.'

Each item had three randomized answer options, with only one correct answer. Each correct answer was awarded 1 point, for a maximum possible score of 6 points per participant. Afterwards, participants were asked to evaluate the perceived comprehensibility of the text using a Likert scale ranging from 0 (*Per niente comprensibile*, 'not at all comprehensible') to 5 (*Molto comprensibile*, 'completely comprehensible') as shown in Appendix 3. This procedure was repeated for each item. Upon completing the main test, participants were asked to respond to a set of sociodemographic questions, including gender, age, education level, geographical origin, and employment status.

The full set of items was divided into three surveys, each created using a different Google Form. This was intended to keep the evaluation duration manageable for participants. Each survey consisted of six different items: two original texts, two rewritten by ChatGPT-3.5, and two rewritten by ChatGPT-4o. Items order and response options were randomized to minimize response order bias. To ensure data quality each survey included two control questions: one multiple-choice instruction-check question (9) after the second item, to verify understanding of the task instructions; one open attention-check question (10) presented after the fourth item, to assess participants' engagement with the task. Participants who failed these controls were excluded from the final analysis.

The control questions are illustrated in the following examples:

(9) *Leggi tutte le risposte e poi seleziona l'opzione che dice "Non selezionare questa risposta*".
- *Non rispondere a questa domanda, passa alla successiva.*
- *Seleziona tutte le opzioni come risposte a questa domanda.*
- *Non selezionare questa risposta.*

'Read all the answers and then select the option that says, "Do not select this answer".
- Don't answer this question, go to the next one.
- Select all options as answers to this question.
- Do not select this answer.'

(10) *Quale mese viene dopo Aprile?*
'Which month comes after April?

Items and control questions in each survey are presented in Table 16.

Table 16: Items and control questions in each survey.

	Survey 1	**Survey 2**	**Survey 3**
Items	CASS-1 Original	CASS-1 ChatGPT-3.5	CASS-1 ChatGPT-4o
	CASS-4 Original	CASS-4 ChatGPT-3.5	CASS-4 ChatGPT-4o
	Instruction-check question		
	MOB-1 ChatGPT-3.5	MOB-1 ChatGPT-4o	MOB-1 Original
	PONT-1 ChatGPT-3.5	PONT-1 ChatGPT-4o	PONT-1 Original
	Attention-check question		
	PRIN-4 ChatGPT-4o	PRIN-4 Original	PRIN-4 ChatGPT-3.5
	PRIN-5 ChatGPT-4o	PRIN-5 Original	PRIN-5 ChatGPT-3.5

7 Survey results

The aim of the survey was to obtain judgments from ordinary citizens who do not have specific experience in legal-juridical documents. During recruitment, the survey targeted Italian native-speaking, adults aged 18 and over who were not employed in professions that require specific legal training (e.g., lawyers, notaries).

From a theoretical perspective, experimental material must be appropriate for the participants, just as participants must be appropriately matched to the material under investigation. While we acknowledge the importance of gathering assessments from individuals with low levels of education, in this case, a clearly defined target group based on educational attainment was necessary to guarantee the validity of the survey.[9] The minimum educational requirement was set at

[9] The identification of the target group was informed not only by theoretical considerations, but also by the application of established metrics, such as the readability index. In Italian, readability index is linked to educational level: the more difficult a text, the higher the level of education required for it to be accessible. Institutional texts fall within the domain of technical texts and, as such, presuppose a high degree of formal education. In addition to the data gained from quantitative analysis, the READ-IT tool was also used to calculate the GULPEASE readability index – a readability metric suitable for Italian – for each item text (Dell'Orletta et al. 2011). The results

completion of high school (equivalent to the *diploma di scuola superiore* in the Italian education system) and it was chosen based on text readability.

Participants were recruited online, and participation was entirely voluntary. Although participants were informed of the academic purpose of the research, no financial compensation was offered. All responses were collected anonymously. Exclusion criteria were based on self-reported data collected through demographic questions.

To facilitate recruitment, a single link access was created to a GitHub repository that randomly assigned participants to one of the three surveys. This method enabled automatic balancing across conditions and helped to distribute the different surveys evenly across social media. A total of 96 participants completed the online survey, with 32 participants completing each of the three survey versions, as shown in Table 17.

Table 17: Surveys distribution to groups.

Groups	**Participants**	**Surveys**
Group 1	32	Survey 1
Group 2	32	Survey 2
Group 3	32	Survey 3

While the sample reflects a range of adult Italian speakers, it is not fully representative of the general population in terms of age, education, or employment status.[10] These limitations are acknowledged and discussed in relation to the scope and generalizability of the findings. The composition of the sample by age,[11] level of education and employment status are presented in Tables 18. In Table 19 the data allow a more detailed comparison across age and education level compared to employment status.

showed an overall average GULPEASE score of 37,6 that indicates a high level of text difficulty. According to Lucisano and Piemontese (1986; 1988), this value indicates a high level of text difficulty: such texts are generally perceived as already "difficult" by readers with a high school diploma and can be considered "very difficult" or even "almost incomprehensible" for readers with lower educational qualifications.

[10] The lack of participants over the age of 60 is primarily attributable to the recruitment method, which relied on dissemination of the survey via social media. On the contrary, using social media as a tool to recruit participants explains the high number in the 18-28 age group.

[11] Since the lack of participants over the age of 60, the subdivision by age was carried out based on ISTAT (2016) generation classification.

Table 18: Participants' screening.

Age	**Participants**	**%**
Over 44	19	19,8%
29-43	32	33,3%
18-28	45	46,9%
Total	**96**	**100%**
Education level	**Participants**	**%**
Master's Degree and higher	47	48,9%
Bachelor's Degree	28	29,2%
High school	21	21,9%
Total	**96**	**100%**
Employment status	**Participants**	**%**
Employed	66	68,75%
Unemployed	30	31,25%
Total	**96**	**100%**

Table 19: Sample's distribution across age and education level compared to employment status.

Age: Education level	**#employed**	**#unemployed**	**#**	**%**
Over 44: Master's Degree and higher	8	2	10	10,42%
Over 44: Bachelor's Degree	1	0	1	1,04%
Over 44: High school	4	4	8	8,33%
29-43: Master's Degree and higher	16	2	18	18,75%
29-43: Bachelor's Degree	7	3	10	10,4%
29-43: High school	2	2	4	4,17%
18-28: Master's Degree and higher	15	3	18	18,75%
18-28: Bachelor's Degree	9	9	18	18,75%
18-28: High school	4	5	9	9,37%
Total	**66**	**30**	**96**	**100%**

The average comprehension scores, ranging from a minimum of 0 to a maximum of 6 points, reflect participants' overall performance on the comprehension task. The proportion of correct answers relative to the total number of questions for each item represents the response accuracy percentage. Perceived comprehensibility judgment was measured using a Likert scale from 0 to 5; the average perceived comprehensibility represents the participants' subjective judgments.

Table 20 reports the response accuracy percentage and the average perceived comprehensibility ratings, aggregated by text type. The results distinguish among original texts, and those rewritten with ChatGPT-3.5 and ChatGPT-4o. These aggregate results are then followed by item-by-item data, allowing a more detailed comparison of each text version's performance in terms of both response accuracy percentage and subjective perceived comprehensibility.

In Table 21 the sample was subdivided, and the results for both comprehension scores and perceived comprehensibility ratings are presented according to participants' age and education level. For each group, the data distinguish between employed status (abbreviated as 'E.'), unemployed (abbreviated as 'U.'), and their total average (indicated as 'Average').

Table 20: Texts type' response accuracy and average perceived comprehensibility, followed by item-level analysis.

Text type	**% response accuracy**	**Average perceived comprehensibility**
Original	74%	2,9
ChatGPT-3.5 rewritings	76%	3,1
ChatGPT-4o rewritings	78%	3,3
Items	**% response accuracy**	**Average perceived comprehensibility**
CASS-1 Original	86%	3,6
CASS-1 ChatGPT-3.5	77%	4,1
CASS-1 ChatGPT-4o	86%	4,3
CASS-4 Original	77%	3,2
CASS-4 ChatGPT-3.5	77%	3,1
CASS-4 ChatGPT-4o	73%	3,2
MOB-1 Original	68%	2,5
MOB-1 ChatGPT-3.5	72%	3,1
MOB-1 ChatGPT-4o	82%	3,4
PONT-1 Original	81%	3,1
PONT-1 ChatGPT-3.5	63%	2,7
PONT-1 ChatGPT-4o	77%	3,4
PRIN-4 Original	81%	2,2
PRIN-4 ChatGPT-3.5	91%	2,9
PRIN-4 ChatGPT-4o	77%	2,6
PRIN-5 Original	77%	2,6
PRIN-5 ChatGPT-3.5	91%	3
PRIN-5 ChatGPT-4o	91%	2,6

Table 21: Average comprehension score and average perceived comprehensibility across age, education level and employment status.

	Comprehension			**Perceived comprehensibility**		
Age: Education level	E.	U.	Average	E.	U.	Average
Over 44: Master's Degree and higher	3,1	4,5	3,8	4,1	2,8	3,4
Over 44: Bachelor's Degree	3	Ø	3	2,8	Ø	2,8
Over 44: High school	4	2,8	3,4	3,4	3,2	3,3
29-43: Master's Degree and higher	4,3	6	5,1	3,1	3	3,1
29-43: Bachelor's Degree	4,7	5,5	5,1	2,9	2	2,5
29-43: High school	6	5,5	5,7	1,8	3,6	2,7
18-28: Master's Degree and higher	4,9	5	4,9	3	3,5	3,2
18-28: Bachelor's Degree	4,4	4,5	4,4	2,6	2,9	2,8
18-28: High school	5	5,2	5,1	3,5	3,3	3,4

We are aware that, since it was not possible to control the experimental conditions, the high scores observed in the comprehension evaluation may be partly attributed to participants having the opportunity to reread the texts before answering. Even

though the results should be considered indicative only and not statistically significant, some trends were nonetheless observed.

7.1 Comprehension

When the data are aggregated by text type, response accuracy remains relatively stable and does not reveal significant differences. Item-level analysis indicates that the rewritten versions do not consistently outperform the originals in terms of comprehension. For example, the original of PONT-1 achieved better response accuracy than its rewritten counterparts. Although the statistical differences are minor, texts rewritten with ChatGPT-4o generally obtained better accuracy scores compared to those generated with ChatGPT-3.5. The only exceptions are CASS-4 and PRIN-4, where the rewritings generated by ChatGPT-3.5 outperformed those produced by ChatGPT-4o. Additionally, in CASS-1, the original version and the ChatGPT-4o rewrite yielded identical response accuracy rates. Regarding performance, better average comprehension scores were recorded among the younger and middle-aged groups, suggesting that age influences cognitive abilities. Educational level does not appear to play a determining role in comprehension performance. Although the selection of texts was based on the unfamiliarity of their content, the participants' performance may have been influenced by the familiarity of linguistic structures characteristic of administrative language. For instance, employed individuals may be more frequently exposed to institutional texts, particularly those working in sectors involving bureaucratic procedures, thus enhancing their comprehension. Conversely, unemployed individuals may encounter administrative documents in the context of job-seeking activities or welfare procedures, which may also contribute to their performance.

Since we do not have information on participants' experience with such texts or whether younger participants reread the texts more or less frequently than older ones, we cannot attribute their better performance solely to the possibility of rereading or to familiarity. To draw reliable conclusions in this regard, additional data would be required – specifically, information on the average number of re-readings per age group and current or previous familiarity with these texts. Regarding comprehension tasks, further investigations will be needed. These aspects will be explored in future studies.

7.2 Perceived Comprehensibility

When data are aggregated by text type, the average perceived comprehensibility ratings vary between the texts rewritten with ChatGPT-3.5 and ChatGPT-4o and their original counterparts. On average, the rewritings produced by ChatGPT-4o are generally judged more positively than the original texts. The 4o rewritings receive slightly higher comprehensibility ratings than those generated by ChatGPT-3.5. The only cases in which a ChatGPT-3.5 outperformed the 4o versions in terms of perceived comprehensibility are PRIN-5 and PRIN-4. A few notable exceptions emerged, such as CASS-4 and PRIN-5, in which the original texts were rated on par with the ChatGPT-4o versions. CASS-4's original is also rated better than the 3.5 version. Although the possibility of rereading the texts may have influenced the

survey comprehension task outcomes, the subjective judgments on average perceived comprehensibility remain generally medium-low. This result confirms the perceived "difficult" readability of the average GULPEASE score assessed with READ-IT. No significant differences emerged in the average perceived comprehensibility ratings across groups: age, level of education and employment status do not influence the perceived comprehensibility. It can therefore be assumed that even a potential familiarity with administrative language does not necessarily enhance the reader's perception of the texts as more comprehensible.

These results suggest that the rewriting process, while slightly improving perceived comprehensibility in some cases, does not produce a systematic change. This outcome may be partly attributable to the quality of the rewritings generated by the LLM and/or the generic, non-specialized prompt. Different results may emerge using other LLMs or by employing specialized prompts and more sophisticated prompting techniques; these aspects are left for future research.

8 Conclusions

The aim of this preliminary study was to analyse the linguistic differences between the original texts compared to their simplified rewritings by ChatGPT, and to explore how AI-generated simplified texts are processed by readers.

The quantitative analysis reveals that both ChatGPT-3.5 and ChatGPT-4o tend to improve several surface-level linguistic parameters commonly associated with textual readability. However, these improvements are not uniformly consistent across all parameters or texts. Importantly, model choices do not always align with conventional simplification guidelines: in several cases, unexpected increases in subordination or clausal restructuring appear to serve functional purposes in clarifying logical relations. As such, the present findings would benefit from being complemented by qualitative investigations into syntactic strategies and communicative adequacy – dimensions that go beyond what surface metrics can reveal.

A closer examination shows a recurring pattern of content-related issues in AI-generated rewritings. We observed three recurring phenomena such as (I) loss of information; (II) lack of information; and (III) recomputation of incorrect information. These were deliberately retained in some of the rewritten texts to investigate their effect on comprehension and perception. Importantly, these patterns did not prevent participants from correctly identifying correct answers, nor did they compromise the coherence of the texts. Nevertheless, a gap remains between actual comprehension and perceived comprehensibility. Although comprehension scores were generally medium-high – likely aided by rereading opportunities and participants' familiarity with administrative language – subjective judgments of perceived comprehensibility are generally medium-low. Overall, the texts are evaluated in a relatively similar way, without generating substantial differences. This suggests that while participants were able to reconstruct a mental representation appropriate to the content of the text, their subjective global comprehensibility judgment and their assessed ease of processing remained limited. These findings point to a possible impact of the quality of the rewritings on users' global judgment of comprehensibility and cognitive effort. We

recognize the influence of individual differences among participants and acknowledge that the variables considered here are not sufficient to draw definitive conclusions. The observed trends, while indicative, are not statistically significant and should be interpreted with caution. To gain more reliable insights, further research will be conducted under controlled conditions – specifically, experiments that limit rereading – to better isolate comprehension gaps. Crucially, the study underscores a broader risk inherent in simplification processes: prioritizing accessibility – particularly from the perspective of the average citizen – can inadvertently lead to the loss or distortion of essential information. Improving textual readability requires more than formal linguistic adjustments; it demands careful management of content integrity and coherence.

In conclusion, the current limitations of ATS point to the need for more sophisticated prompting strategies, alongside the development of robust evaluation frameworks for assessing both the clarity and informational accuracy of simplified texts. Only through such refinement, automatic simplification technologies can become truly reliable tools for enhancing accessibility without compromising content.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this contribution.

References

Al-Thanyyan, Suha & Azmi, Aqil. 2021. Automated text simplification: a survey. *ACM Computing Surveys* 54(2). 1-36. https://doi.org/10.1145/3442695

Benjamin, Rebekah George. 2012. Reconstructing readability: recent developments and recommendations in the analysis of text difficulty. *Educational Psychology Review* 24. 63-88. https://doi.org/10.1007/s10648-011-9181-8

Cherubini, Manola & Romano, Francesco & Bolioli, Andrea & De Francesco, Nazareno & Benedetto, Irene. 2023. La summarization di testi giuridici: una sperimentazione con GPT-3. *Rivista italiana di informatica e diritto* 5(1). 191-204. https://doi.org/10.32091/RIID0103

Codice di stile = AA.VV. 1993. *Codice di stile delle comunicazioni scritte ad uso delle amministrazioni pubbliche proposta e materiali di studio.* Roma: Istituto Poligrafico e Zecca dello Stato.

Collins-Thompson, Kevyn. 2014. Computational assessment of text readability: a survey of current and future research. *International Journal of Applied Linguistics* 165(2). 97-135. https://doi.org/10.1075/itl.165.2.01col

Conklin, Kathy & Pellicer-Sánchez, Ana & Carrol, Gareth. 2018. *Eye-tracking. A Guide for Applied Linguistics Research*. Cambridge: Cambridge University Press.

Cortelazzo, Michele A. & Pellegrino, Federica. 2002. *30 regole per scrivere testi amministrativi chiari.* Università di Padova, http://www.maldura.unipd.it/buro/ (last accessed on 18/12/2024).

Cortelazzo, Michele A. & Pellegrino, Federica. 2003. *Guida alla scrittura istituzionale*. Bari: Laterza.

Cortelazzo, Michele. 2021. *Il linguaggio amministrativo: principi e pratiche di modernizzazione*. Roma: Carocci.

De Mauro, Tullio & Chiari, Isabella. 2016. *Nuovo vocabolario di base della lingua italiana*. https://www.internazionale.it/opinione/tullio-de-mauro/2016/12/23/il-nuovo-vocabolario-di-base-della-lingua-italiana (last accessed on 09/01/2025).

De Mauro, Tullio. 1980. *Guida all'uso delle parole*. Italia: Editori Riuniti.

De Mauro, Tullio & Vedovelli, Massimo (eds). 1999. *Dante, il gendarme e la bolletta. La comunicazione pubblica in Italia e la nuova bolletta Enel*. Bari: Laterza.

Dell'Orletta, Felice & Montemagni, Simonetta & Venturi, Giulia. 2011. READ-IT: assessing readability of Italian with a view to text simplification. In *Proceedings of the second workshop on speech and language processing for assistive technologies* (Edinburgh, July 30, 2011), 73-83. Stroudsburg: Association for Computational Linguistics.

Efrat, Avia & Levy, Omer. 2020. The Turking Test: Can Language Models understand instructions? *arXiv*. https://doi.org/10.48550/arXiv.2010.11982

Fiorentino, Giuliana & Ganfi, Vittorio. 2024. Parametri per semplificare l'italiano istituzionale: revisione della letteratura. *Italiano LinguaDue* 16(1). 220-237. https://doi.org/10.54103/2037-3597/23835

Fioritto, Alfredo (ed.). 1997. *Manuale di stile. Strumenti per semplificare il linguaggio delle amministrazioni pubbliche*. Presidenza del Consiglio dei Ministri Dipartimento della Funzione Pubblica. Bologna: Il Mulino.

Franceschini, Fabrizio & Gigli, Sara (eds). 2003. *Manuale di scrittura amministrativa*. Dipartimento di Studi Italianistici (Università di Pisa) & Agenzia delle Entrate.

Friedrich, Marcus C. G. & Heise, Elke. 2025. The influence of comprehensibility on interest and comprehension. *Zeitschrift für pädagogische Psychologie* 39(1-2). 139-152. https://doi.org/10.1024/1010-0652/a000349

Godfroid, Aline. 2020. *Eye Tracking in Second Language Acquisition and Bilingualism. A Research Synthesis and Methodological Guide*. New York: Routledge Taylor & Francis Group.

Gualdo, Riccardo & Telve, Stefano. 2011. *Linguaggi specialistici dell'italiano*. Roma: Carocci.

ISTAT. 2016. *Classificazione delle generazioni*. Istituto Nazionale di Statistica. https://www.istat.it/it/files//2011/01/Generazioni-nota.pdf (last accessed on 09/01/2025).

ITTIG & Accademia della Crusca. 2011. *Guida alla redazione degli atti amministrativi. Regole e suggerimenti*. Firenze.http://hdl.handle.net/2158/540886.

Kintsch, Walter. 1988. The role of knowledge in discourse processing: a construction-integration model. *Psychological Review* 85(2). 363-394. https://doi.org/10.1037/0033-295X.95.2.163.

Kintsch, Walter. 1998. *Comprehension. A Paradigm for Cognition*. New York: Cambridge University Press.

Korzen, Iørn. 2022. Cosa ci rivelano i corpora sulla complessità testuale dell'italiano? In Cresti, Emanuela & Moneglia, Massimo (eds), *Corpora e Studi Linguistici. Atti del LIV Congresso Internazionale di Studi della Società di Linguistica Italiana* (Online, September 8-10, 2021), 354-365. Officinaventuno. https://doi.org/10.17469/O2106SLI000023

Lubello, Sergio. 2014. *Il linguaggio burocratico*. Roma: Carocci.

Lubello, Sergio. 2017. *La lingua del diritto e dell'amministrazione*. Bologna: Il Mulino.

Lucisano, Pietro & Piemontese, Maria Emanuela. 1986. Leggibilità dei testi e comprensione della lettura. *Linguaggi* III, 28-38.

Lucisano, Pietro & Piemontese, Maria Emanuela. 1988. GULPEASE: una formula per la predizione della difficoltà dei testi in lingua italiana. *Scuola e città* XXIX(1), 110-124.

Mayer, Richard E. 2014. Cognitive theory of multimedia learning. In Mayer, Richard E. (ed.), *The Cambridge Handbook of Multimedia Learning* (2nd ed.), 43-71. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.005

McNamara, Danielle S. & Graesser, Arthur C. & McCarthy, Philip M. & Cai, Zhiqiang. 2014. *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511894664

Mishra, Swaroop & Khashabi, Daniel & Baral, Chitta & Choi, Yejin & Hajishirzi, Hannaneh. 2022. Reframing Instructional Prompts to GPTk's Language. *arXiv*. https://doi.org/10.48550/arXiv.2109.07830

Paci, Walter & Gregori, Lorenzo & Acerborni, Giovanni & Panunzi, Alessandro & Perugini, Maria Roberta. 2024. Exploring ChatGPT to simplify Italian bureaucratic and professional texts. *AI-Linguistica* (1)1. 1-23. https://doi.org/10.62408/ai-ling.v1i1.13

Palermo, Massimo. 2013. *Linguistica testuale dell'italiano*. Bologna: Il Mulino.

Piemontese, Maria Emanuela (ed.). 2023. *Il dovere costituzionale di farsi capire. A trent'anni dal Codice di stile.* Roma: Carocci.

Piemontese, Maria Emanuela. 1996. *Capire e frasi capire: teorie e tecniche della scrittura controllata.* Napoli: Tecnodid.

Raso, Tommaso. 2005. *La scrittura burocratica. La lingua e l'organizzazione del testo*. Roma: Carocci.

Reber, Rolf & Greifeneder, Rainer. 2017. Processing fluency in education: how metacognitive feelings shape learning, belief formation, and affect. *Educational Psychologist* 52(2). 84-103. https://doi.org/10.1080/00461520.2016.1258173.

Reed, Debora K. & Kershaw-Herrera, Sarah. 2016. An examination of text complexity as characterized by readability and cohesion. *The Journal of Experimental Education* 84(1). 75-97. https://doi.org/10.1080/00220973.2014.963214

Sabatini, Francesco. 2012. Tra grammatica e testi. *L'italiano nel mondo moderno*, vol. 3, 183-210. Napoli: Liguori.

Saggion, Horacio. 2017. Automatic text simplification. In Hirst, Graeme (ed.), *Synthesis Lectures on Human-Language Technologies*. Toronto: Morgan & Claypool Publishers. https://doi.org/10.1007/978-3-031-02166-4

Schnotz, Wolfgang. 2014. Integrated model of text and picture comprehension. In Mayer, Richard E. (ed.), *Cambridge Handbook of Multimedia Learning* (2nd ed.), 72-103. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.006

Shardlow, Matthew. 2014. A survey of automated text simplification. *International Journal of Advanced Computer Science and Applications* 4(1). 58-70. https://doi.org/10.14569/SpecialIssue.2014.040109

Tavosanis, Mirko. 2024. Valutare la qualità dei testi generati in lingua italiana. *AI-Linguistica* 1(1). 1-24. https://doi.org/10.62408/ai-ling.v1i1.14

Tavosanis, Mirko. 2025. Valutare il miglioramento della chiarezza eseguito da intelligenze artificiali generative. In Fiorentino, Giuliana & Cioffi, Alessandro & Simonelli, Maria Ausilia (eds), *Amministrazione attiva: semplicità e chiarezza per la comunicazione amministrativa*. Firenze: Franco Cesati.

Vellutino, Daniela. 2018. *L'italiano istituzionale per la comunicazione pubblica*. Bologna: Il Mulino.

Viale, Matteo. 2008. *Studi e ricerche sul linguaggio amministrativo.* Padova: CLEUP.

Vajjala, Sowmya. 2021. Trends, limitations and open challenges in automatic readability assessment research. *arXiv*. https://doi.org/10.48550/arXiv.2105.00973

Appendix 1– Example: CASS-4.

Testo originale completo: *Considerato che numerosi cittadini avanzano richieste di concessione di installazione dei cosiddetti "parapedonali" al fine di impedire la sosta, sempre più frequente, delle autovetture sui marciapiedi;*
considerato che tale "sosta selvaggia" sui marciapiedi impedisce il normale transito dei pedoni, restringe, fino a renderlo impossibile, l'accesso ai passi carrabili e causa l'immissione di agenti inquinanti nelle abitazioni situate nei piani seminterrati o rialzati;
considerato che il Servizio Tecnico circoscrizionale ha predisposto in merito una relazione che individua criteri di carattere generale, affinché il Servizio Tecnico possa, previo parere del Comando del II Gruppo VV.UU., autorizzare l'installazione delle barriere metalliche, concordate sia nel numero che nella forma estetica, senza dover sottoporre le richieste, ogni volta, al parere della Commissione circ.le Traffico;
visto il parere favorevole espresso dal Il gruppo VV.UU.:
visto il parere favorevole espresso dalla Commissione circ.le Traffico nella seduta del 29.1.2021;
Il Consiglio Circoscrizionale risolve, per motivi espressi in narrativa, di fissare i seguenti criteri al fine di consentire all'ufficio Tecnico circ.le di essere autorizzato ad installare direttamente ovvero di autorizzare i cittadini ad installare barriere metalliche senza dover sottoporre le richieste al parere della Commissione circ.le Traffico:

1) *non potranno essere autorizzati più di due o tre parapedonali (secondo l'ampiezza del marciapiede e quindi la necessità) sui due lati dei passi carrabili;*
2) *i parapedonali dovranno essere installati lungo i marciapiedi fronteggianti l'uscita delle scuole (per una lunghezza ipotizzabile in 20 metri circa, secondo le necessità e le disponibilità economiche);*
3) *i parapedonali dovranno essere installati lungo i marciapiedi in corrispondenza degli incroci al fine di impedire la sosta delle autovetture con grave limitazione della visibilità;*
4) *i parapedonali dovranno essere installati lungo i marciapiedi in corrispondenza delle Ambasciate che, in genere per motivi di sicurezza, ne fanno richiesta.*

'Original full text: Considering that numerous citizens are making requests for the installation of so-called "pedestrian barriers" in order to prevent the increasingly frequent parking of cars on sidewalks;
considering that this "wild parking" on the sidewalks prevents the normal transit of pedestrians, restricts access to driveways to the point of making it impossible and causes the release of pollutants into homes located in the basement or mezzanine floors;
considering that the district Technical Service has prepared a report on the matter that identifies general criteria, so that the Technical Service can, after consulting the Command of the II Group VV.UU, authorize the installation of metal barriers,

agreed upon both in number and aesthetic form, without having to submit the requests, each time, to the opinion of the Traffic Circular Commission;
having seen the favorable opinion expressed by the VV.UU. group:
having seen the favorable opinion expressed by the Circ. Traffic Commission in the session of 29.1.2021;
The District Council resolves, for reasons expressed in the narrative, to establish the following criteria in order to allow the district Technical office to be authorized to directly install or to authorize citizens to install metal barriers without having to submit requests to the opinion of the district Traffic Commission:

1) no more than two or three pedestrian barriers may be authorized (depending on the width of the sidewalk and therefore the need) on the two sides of the driveways;
2) the pedestrian barriers must be installed along the sidewalks facing the exits of the schools (for a length estimated at approximately 20 meters, according to needs and financial availability);
3) the pedestrian barriers must be installed along the sidewalks at intersections in order to prevent parking of cars with serious limitation of visibility;
4) the pedestrian barriers must be installed along the sidewalks at the Embassies which, generally for security reasons, request them.'

Source: adaptation from *Codice di stile*, 1993: 78.

Riscrittura ChatGPT-3.5: *Vista la crescente richiesta da parte dei cittadini di installare i cosiddetti 'parapedonali' per contrastare il problema sempre più diffuso della sosta selvaggia delle auto sui marciapiedi;*
considerato l'impedimento al normale transito dei pedoni e l'ostacolo all'accesso ai passi carrabili causato da questa pratica, oltre all'inquinamento atmosferico che si verifica nelle abitazioni situate ai piani seminterrati o rialzati;
dato che il Servizio Tecnico circoscrizionale ha redatto una relazione per stabilire criteri generali per l'autorizzazione dell'installazione delle barriere metalliche, evitando così di dover sottoporre ogni singola richiesta al parere della Commissione circoscrizionale del Traffico;
tenuto conto del parere favorevole del II Gruppo VV.UU. e della Commissione circoscrizionale del Traffico espresso nella seduta del 29.1.2021;
Il Consiglio Circoscrizionale, in virtù delle ragioni esposte, decide di stabilire i seguenti criteri per consentire all'Ufficio Tecnico circoscrizionale di autorizzare direttamente l'installazione dei parapedonali o di concedere ai cittadini l'autorizzazione per farlo senza dover ricorrere al parere della Commissione circoscrizionale del Traffico:

1) Sarà consentita l'installazione di massimo due o tre parapedonali su entrambi i lati dei passi carrabili, a seconda dello spazio disponibile e delle esigenze;
2) I parapedonali dovranno essere posizionati lungo i marciapiedi di fronte alle uscite delle scuole, con una lunghezza approssimativa di 20 metri, a seconda delle necessità e delle risorse finanziarie disponibili;
3) Sarà obbligatorio installare i parapedonali nei pressi degli incroci per impedire la sosta delle auto, che altrimenti potrebbe ostacolare gravemente la visibilità;

4) *Sarà richiesto l'installazione di parapedonali lungo i marciapiedi in corrispondenza delle Ambasciate, su richiesta di queste ultime per ragioni di sicurezza.*

'ChatGPT-3.5 rewrite: 'Given the growing demand from citizens to install the so-called 'pedestrian barriers' to combat the increasingly widespread problem of illegal parking of cars on sidewalks;
considering the impediment to the normal transit of pedestrians and the obstacle to access to driveways caused by this practice, in addition to the air pollution that occurs in homes located on basement or mezzanine floors;
given that the District Technical Service has drawn up a report to establish general criteria for the authorization of the installation of metal barriers, thus avoiding having to submit each individual request to the opinion of the District Traffic Commission;
taking into account the favorable opinion of the II Group VV.UU. and the District Traffic Commission expressed in the session of 29.1.2021;
The District Council, by virtue of the reasons set out, decides to establish the following criteria to allow the District Technical Office to directly authorize the installation of pedestrian barriers or to grant citizens authorization to do so without having to resort to the opinion of the District Traffic Commission:

1) The installation of a maximum of two or three pedestrian barriers will be permitted on both sides of driveways, depending on the available space and needs;
2) The pedestrian barriers must be positioned along the sidewalks in front of school exits, with an approximate length of 20 meters, depending on needs and available financial resources;
3) It will be mandatory to install pedestrian barriers near intersections to prevent parking of cars, which otherwise could seriously hinder visibility;
4) The installation of pedestrian barriers along the sidewalks in correspondence with Embassies will be required, at the request of the latter for security reasons.'

Riscrittura ChatGPT-4o*: Considerato che molti cittadini chiedono di poter installare barriere pedonali per impedire alle auto di parcheggiare sui marciapiedi, situazione sempre più frequente;*
considerato che questo parcheggio sui marciapiedi ostacola il passaggio dei pedoni, rende difficile o impossibile l'accesso ai passi carrai e causa l'ingresso di agenti inquinanti nelle case vicine situate al piano terra o rialzato;
considerato che il Servizio Tecnico della circoscrizione ha preparato un documento che stabilisce criteri generali per l'installazione delle barriere, affinché il Servizio Tecnico possa autorizzare l'installazione, previo parere favorevole del Comando del II Gruppo della Polizia Locale, senza dover chiedere ogni volta l'approvazione della Commissione circoscrizionale Traffico;
visto il parere favorevole del II Gruppo della Polizia Locale;
visto il parere favorevole della Commissione Traffico espresso nella seduta del 29.1.2021;
Il Consiglio Circoscrizionale decide, per le ragioni sopra esposte, di stabilire i seguenti criteri per permettere all'Ufficio Tecnico circoscrizionale di autorizzare

direttamente o ai cittadini di installare barriere senza chiedere sempre il parere della Commissione Traffico:

1) *si possono autorizzare al massimo due o tre barriere pedonali (a seconda della larghezza del marciapiede) sui due lati dei passi carrai;*
2) *le barriere devono essere installate sui marciapiedi di fronte alle uscite delle scuole, per una lunghezza di circa 20 metri, in base alle esigenze e alle disponibilità economiche;*
3) *le barriere devono essere installate ai bordi dei marciapiedi vicino agli incroci per impedire alle auto di parcheggiare e limitare la visibilità;*
4) *le barriere devono essere installate sui marciapiedi davanti alle Ambasciate che ne fanno richiesta per motivi di sicurezza.*

'ChatGPT-4o rewrite: Considering that many citizens are asking to be able to install pedestrian barriers to prevent cars from parking on sidewalks, an increasingly frequent situation;
considering that this parking on sidewalks hinders the passage of pedestrians, makes access to driveways difficult or impossible and causes pollutants to enter nearby houses located on the ground floor or mezzanine;
considering that the Technical Service of the district has prepared a document that establishes general criteria for the installation of barriers, so that the Technical Service can authorize the installation, subject to the favorable opinion of the Command of the II Group of the Local Police, without having to request the approval of the district Traffic Commission each time;
having seen to the favorable opinion of the II Group of the Local Police;
having seen to the favorable opinion of the Traffic Commission expressed in the session of 29.1.2021;
The District Council decides, for the reasons stated above, to establish the following criteria to allow the District Technical Office to directly authorize or allow citizens to install barriers without always asking the opinion of the Traffic Commission:

1) a maximum of two or three pedestrian barriers (depending on the width of the sidewalk) can be authorized on both sides of driveways;
2) the barriers must be installed on the sidewalks in front of school exits, for a length of approximately 20 meters, based on needs and financial availability;
3) the barriers must be installed at the edges of the sidewalks near intersections to prevent cars from parking and limit visibility;
4) the barriers must be installed on the sidewalks in front of Embassies that request them for security reasons.'

Appendix 2 - Analysis of the five parameters for each text.

Table 1: CASS-1.

Parameters	**Originals**	**ChatGPT-3.5**	**ChatGPT-4o**	**Human**
(1) Average Sentence Length (in tokens)	32,5	12,6	14,1	25,5
(2) Percentage of Subordinate Clauses	16,7%	11,1%	30,8%	41,2%
(3) Average Number of Clauses Per Sentence	2,6	1,1	1,0	3
(4) Average Number of Words Per Clause	12,3	11,1	15,7	12,7
(5) Percentage of Basic Vocabulary	61,9%	63,2%	64,6%	75,8%

Table 2: CASS-4.

Parameters	**Originals**	**ChatGPT-3.5**	**ChatGPT-4o**	**Human**
(1) Average Sentence Length (in tokens)	33,4	30,9	26	21,7
(2) Percentage of Subordinate Clauses	42,3%	44,4%	54,5%	61,4%
(3) Average Number of Clauses Per Sentence	3,7	3,3	3	1,7
(4) Average Number of Words Per Clause	9,1	9,4	8,7	7,7
(5) Percentage of Basic Vocabulary	77,1%	79,9%	79,1%	61,4%

Table 3: MOB-1.

Parameters	**Originals**	**ChatGPT-3.5**	**ChatGPT-4o**	**Human**
(1) Average Sentence Length (in tokens)	45,6	38,7	38,6	25,5
(2) Percentage of Subordinate Clauses	47,1%	46,2%	42,9%	41,5%
(3) Average Number of Clauses Per Sentence	4,4	3,1	3,2	3
(4) Average Number of Words Per Clause	10,4	12,4	11,1	8,5
(5) Percentage of Basic Vocabulary	66,5%	66,9%	66,3%	75,8%

Table 4: PONTI-1.

Parameters	Originals	ChatGPT-3.5	ChatGPT-4o	Human
(1) Average Sentence Length (in tokens)	63,2	47,1	29,2	27
(2) Percentage of Subordinate Clauses	41,2%	41,2%	50%	38,5%
(3) Average Number of clauses Per Sentence	5,8	4,6	3,3	2,8
(4) Average Number of Words Per clause	10,9	10,2	8,8	9,8
(5) Percentage of Basic Vocabulary	65,2%	65,2%	72,3%	67,4%

Table 5: PRIN-4.

Parameters	Originals	ChatGPT-3.5	ChatGPT-4o	Human
(1) Average Sentence Length (in tokens)	33,6	31,4	30,2	23,8
(2) Percentage of Subordinate Clauses	27,3%	29,9%	42,9%	37,5%
(3) Average Number of clauses Per Sentence	3,6	4	3,7	3
(4) Average Number of Words Per clause	10,8	7,8	8,1	7,9
(5) Percentage of Basic Vocabulary	61,8%	68,9%	73,4%	66,7%

Table 6: PRIN-5.

Parameters	Originals	ChatGPT-3.5	ChatGPT-4o	Human
(1) Average Sentence Length (in tokens)	56,3	50,3	30	25
(2) Percentage of Subordinate Clauses	22,2%	45,5%	25,5%	10%
(3) Average Number of clauses Per Sentence	2,3	3,1	2,4	1,7
(4) Average Number of Words Per clause	24,1	15,9	12,4	14,3
(5) Percentage of Basic Vocabulary	54,2%	53,4%	51%	67,6%

Appendix 3 – Example: CASS-4.

Introduzione	*Il Consiglio Circoscrizionale costituisce un organismo dell'amministrazione comunale, che si occupa della gestione dei servizi di base e manutenzione degli spazi pubblici.*
Originale: Gruppo1	*Considerato che numerosi cittadini avanzano richieste di concessione di installazione dei cosiddetti "parapedonali" al fine di impedire la sosta, sempre più frequente, delle autovetture sui marciapiedi;* *considerato che tale "sosta selvaggia" sui marciapiedi impedisce il normale transito dei pedoni, restringe, fino a renderlo impossibile, l'accesso ai passi carrabili e causa l'immissione di agenti inquinanti nelle abitazioni situate nei piani seminterrati o rialzati;* *Il Consiglio Circoscrizionale risolve, per motivi espressi in narrativa, di fissare i seguenti criteri al fine di consentire all'ufficio Tecnico circoscrizionale di essere autorizzato ad installare direttamente ovvero di autorizzare i cittadini ad installare barriere metalliche senza dover sottoporre le richieste al parere della Commissione circoscrizionale del Traffico.*
ChatGPT-3.5: Gruppo2	*Vista la crescente richiesta da parte dei cittadini di installare i cosiddetti 'parapedonali' per contrastare il problema sempre più diffuso della sosta selvaggia delle auto sui marciapiedi;* *considerato l'impedimento al normale transito dei pedoni e l'ostacolo all'accesso ai passi carrabili causato da questa pratica, oltre all'inquinamento atmosferico che si verifica nelle abitazioni situate ai piani seminterrati o rialzati;* *Il Consiglio Circoscrizionale, in virtù delle ragioni esposte, decide di stabilire i seguenti criteri per consentire all'Ufficio Tecnico circoscrizionale di autorizzare direttamente l'installazione dei parapedonali o di concedere ai cittadini l'autorizzazione per farlo senza dover ricorrere al parere della Commissione del Traffico.*
ChatGPT-4o: Gruppo3	*Considerato che molti cittadini chiedono di poter installare barriere pedonali, cosiddetti "parapedonali", per impedire alle auto di parcheggiare sui marciapiedi, situazione sempre più frequente;* *considerato che questo parcheggio sui marciapiedi ostacola il passaggio dei pedoni, rende difficile o impossibile l'accesso ai passi carrai e causa l'ingresso di agenti inquinanti nelle case vicine situate al piano terra o rialzato;*

	Il Consiglio Circoscrizionale decide, per le ragioni sopra esposte, di stabilire i seguenti criteri per permettere all'Ufficio Tecnico circoscrizionale di autorizzare direttamente o ai cittadini di installare barriere senza chiedere sempre il parere della Commissione del Traffico.
Domanda di comprensione - la stessa per ogni gruppo	**Scegli la risposta corretta:** * ○ Il Consiglio Circoscrizionale consente ai cittadini di installare i "parapedonali" per impedire il parcheggio alle auto. ○ L'Ufficio Tecnico circoscrizionale consente al Consiglio Circoscrizionale di autorizzare i cittadini a installare i "parapedonali" per impedire il parcheggio alle auto. ○ Il Consiglio Circoscrizionale consente all'Ufficio Tecnico circoscrizionale di autorizzare i cittadini a installare i "parapedonali" per impedire il parcheggio alle auto.
Valutazione comprensibilità	Quanto valuti la comprensibilità di lettura del testo? * 0 1 2 3 4 5 Per niente compresibile ○ ○ ○ ○ ○ ○ Molto comprensibile

Our translation follows:

Introduction:	'The District Council is a body of the municipal administration, which deals with the management of basic services and maintenance of public spaces.'
Original: Group1	'Considering that numerous citizens are making requests for the installation of so-called "pedestrian barriers" in order to prevent the increasingly frequent parking of cars on sidewalks; considering that such "wild parking" on sidewalks prevents the normal transit of pedestrians, restricts, to the point of making it impossible, access to driveways and causes the release of polluting agents into homes located in the basement or raised floors; The District Council resolves, for reasons expressed in the narrative, to establish the following criteria in order to allow the district Technical Office to be authorized to directly install or to authorize citizens to install metal barriers without having to submit the requests to the opinion of the district Traffic Commission.'
ChatGPT-3.5: Group2	**'**Given the growing demand from citizens to install the so-called 'pedestrian barriers' to combat the increasingly widespread problem of unauthorized parking of cars on sidewalks;

	considering the impediment to the normal transit of pedestrians and the obstacle to access to driveways caused by this practice, in addition to the air pollution that occurs in homes located on basement or mezzanine floors; The District Council, by virtue of the reasons set out, decides to establish the following criteria to allow the District Technical Office to directly authorize the installation of pedestrian barriers or to grant citizens authorization to do so without having to resort to the opinion of the Traffic Commission.'
ChatGPT-4o: Group3	**'**Considering that many citizens ask to be able to install pedestrian barriers, so-called "pedestrian barriers", to prevent cars from parking on sidewalks, an increasingly frequent situation; considering that this parking on sidewalks hinders the passage of pedestrians, makes access to driveways difficult or impossible and causes pollutants to enter nearby houses located on the ground floor or mezzanine; The District Council decides, for the reasons stated above, to establish the following criteria to allow the District Technical Office to directly authorize or the citizens to install barriers without always asking the opinion of the Traffic Commission.'
Comprehension question - the same for each group	**'Choose the correct answer**: o The District Council allows citizens to install "pedestrian barriers" to prevent cars from parking. o The District Technical Office allows the District Council to authorize citizens to install "pedestrian barriers" to prevent cars from parking. o The District Council allows the District Technical Office to authorize citizens to install "pedestrian barriers" to prevent cars from parking.'
Comprehensibility evaluation	*How much do you rate the reading comprehensibility of the text?* 0 1 2 3 4 5 ○ ○ ○ ○ ○ ○ Not at all comprehensible Very comprehensible