

SEMPLE-IT

**il primo LLM per la semplificazione
dell'italiano amministrativo**

Giuliana Fiorentino, Vittorio Ganfi, Marco Russodivito

Università degli Studi del Molise

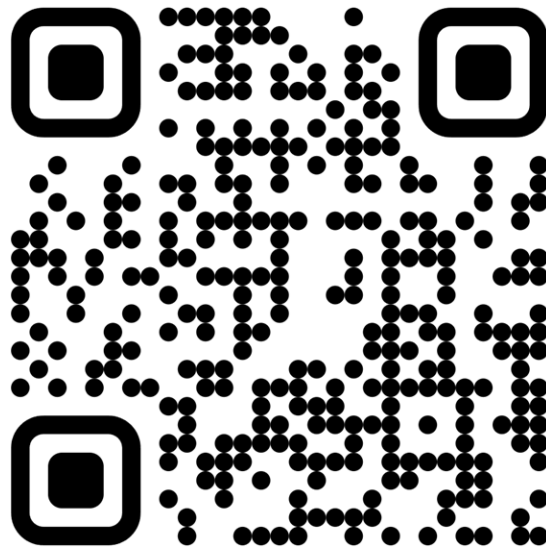


PRIN 2020

VerbACxSS: verbi analitici, complessità, verbi sintetici, e semplificazione. Per l'accessibilità.

Giugno 2022 – Maggio 2025

<https://verbacxss.it>

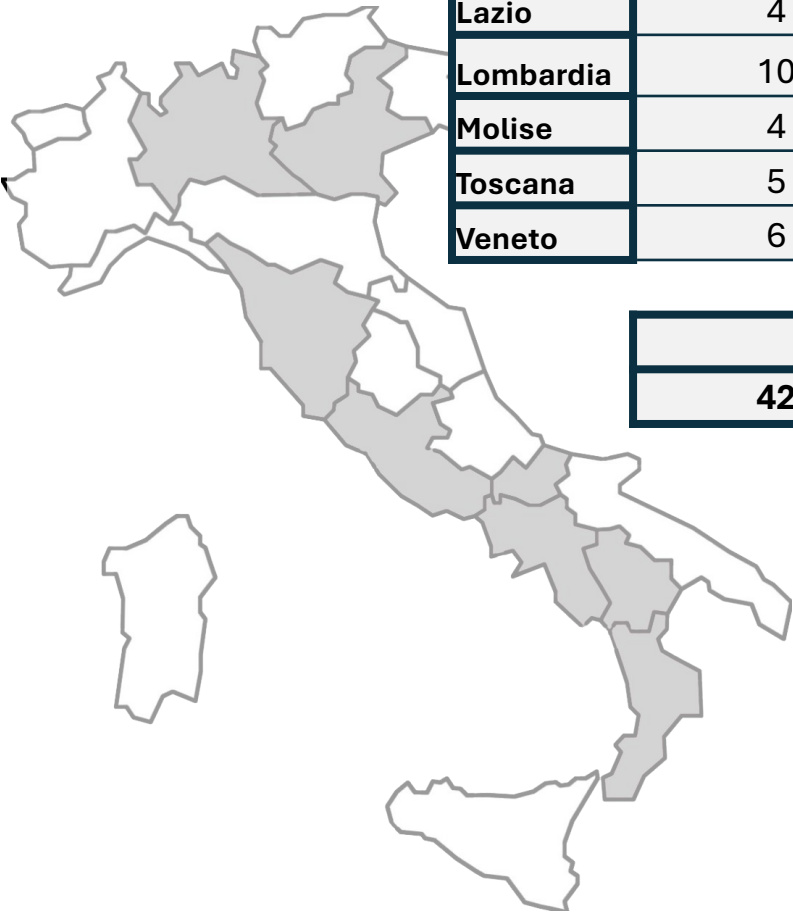


AI vs Umano

Russodivito, Ganfi, Fiorentino, Oliveto (2024) “AI vs. Human Effectiveness of LLMs in Simplifying Italian Administrative Documents” in F. Dell'Orletta, A. Lenci, S. Montemagni, & R. Sprugnoli (eds.) Proceedings of the Tenth Italian Conference on Computational Linguistics (CLiC-it 2024). CEUR, 1-12



Corpus Italst

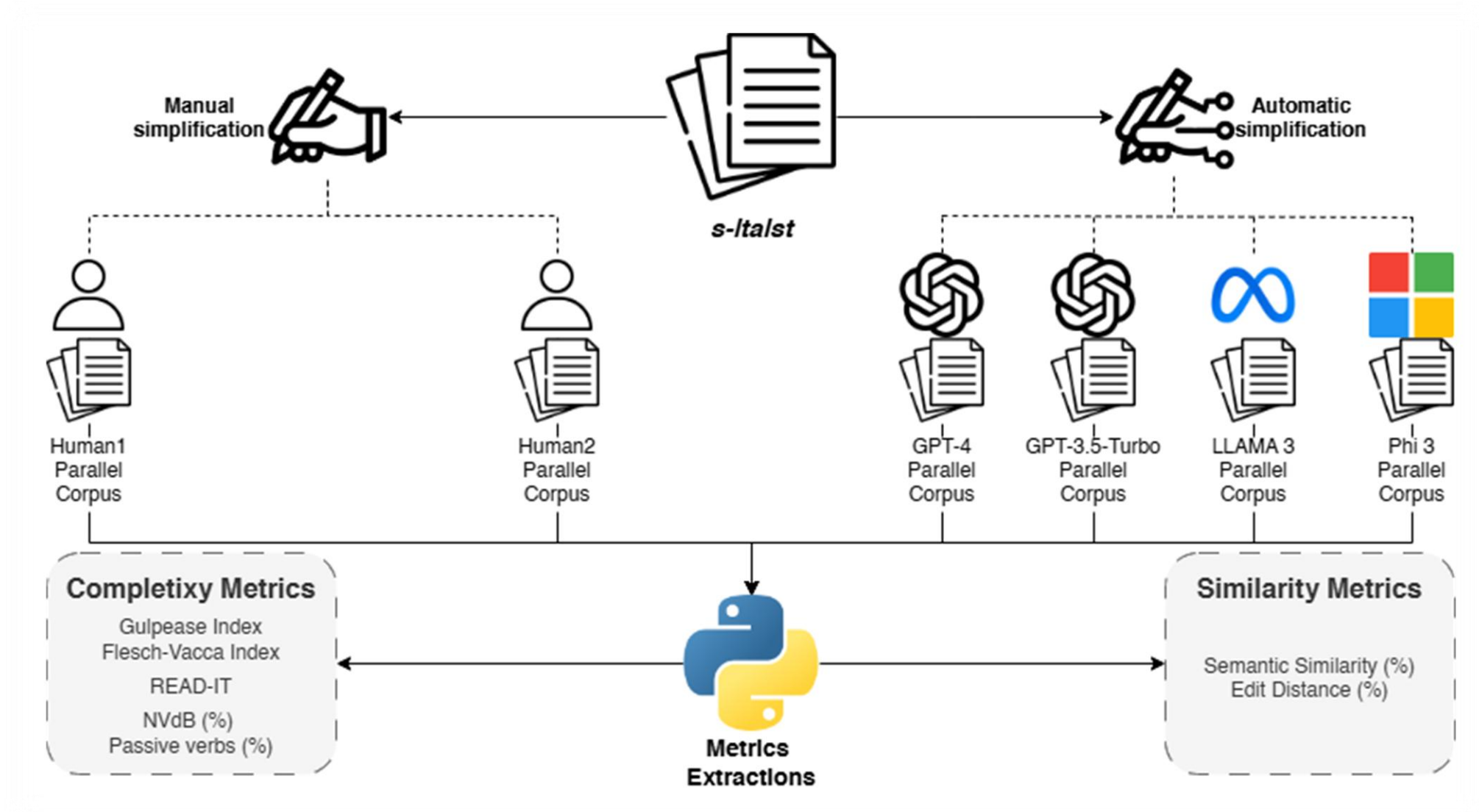


	Rifiuti		Sanità		Servizi pubblici		
	Carta dei servizi	Bandi di gara	Atti generali di prog.	Accreditamento	Carta dei servizi	Razion. partecip. pub.	
Basilicata	3	5	1	2	4	5	20
Calabria	5	6	1	4	4	5	25
Campania	5	9	1	6	4	5	30
Lazio	4	5	1	2	4	5	21
Lombardia	10	5	1	2	4	7	29
Molise	4	6	1	6	4	5	26
Toscana	5	12	1	3	4	8	33
Veneto	6	3	1	4	4	6	24
							208
93			37		78		208
42		51	8	29	32	46	

208
documenti

2M
tokens

AI vs Umano – Disegno dello studio



AI vs Umano – Prompt

Sei un dipendente pubblico che deve riscrivere dei documenti istituzionali italiani per renderli **semplici e comprensibili per i cittadini**. Ti verrà fornito un documento pubblico e il tuo compito sarà quello di riscriverlo applicando regole di semplificazione **senza però modificare il significato** del documento originale. Ad esempio potresti rendere le **frasi** più **brevi**, **eliminare le perifrasi**, **esplicitare** sempre il **soggetto**, utilizzare **parole** più **semplici**, trasformare i verbi **passivi** in verbi di forma **attiva**, spostare le **frasi parentetiche** alla **fine** del periodo.

AI vs Umano – Risultati

	<i>Original</i>	<i>Human1</i>	<i>Human2</i>	<i>GPT-3.5-Turbo</i>	<i>GPT-4</i>	<i>LLaMA 3</i>	<i>Phi 3</i>
<i>Tokens</i>	33,295	34,135	29,755	30,032	31,722	36,035	36,056
<i>Sentences</i>	1,314	1,506	1,744	1,515	1,840	1,944	1,900
<i>Tokens per Sentences</i>	25.33	22.66	17.06	19.53	17.24	18.53	18.97
<i>Sentences per Documents</i>	164.25	188.25	218.00	189.37	230.00	243.00	237.50
<i>Gulpease Index</i>	44.31	49.72	50.64	48.49	51.34	50.26	50.16
<i>Flesch Vacca Index</i>	19.97	34.23	33.63	30.33	36.75	34.09	33.75
<i>NVdB (%)</i>	73.28	80.44	76.89	78.28	81.07	80.18	80.16
<i>Passive (%)</i>	20.87	15.78	17.71	13.99	12.00	15.81	15.72
<i>READ-IT BASE (%)</i>	75.91	68.62	51.00	66.61	55.00	58.37	57.69
<i>READ-IT LEXICAL (%)</i>	93.64	85.37	89.71	91.96	90.29	77.13	75.74
<i>READ-IT SYNTACTIC (%)</i>	63.72	53.14	40.09	38.42	29.92	40.97	41.24
<i>READ-IT GLOBAL (%)</i>	86.48	69.24	61.34	68.69	54.60	59.26	58.37
<i>Semantic Similarity (%)</i>	-	96.52	97.26	96.06	95.80	94.96	94.96
<i>Edit distance (%)</i>	-	35.84	29.20	49.21	52.14	55.48	55.44

AI vs Umano – Risultati

Le semplificazioni automatiche (GPT-4) possono essere paragonate a quelle manuali.

Le semplificazioni automatiche (GPT-4) presentano metriche di complessità leggermente migliori di quelle manuali.

Le semplificazioni automatiche (GPT-4) presentano una similarità semantica inferiore a quelle manuali

AI vs Umano – Limitazioni

I documenti sono stati semplificati troppo

Lo stile è stato profondamente cambiato

Possibile perdita di valenza giuridica

Valutazione giuridica dei testi semplificati con AI

Fiorentino, Russodivito, Ganfi, Oliveto (in stampa) «Validazione e confronto tra semplificazione automatica e semplificazione manuale di testi in italiano istituzionale ai fini dell'efficacia comunicativa» in AI-Linguistica.
Linguistic Studies on AI-Generated Texts and Discourses

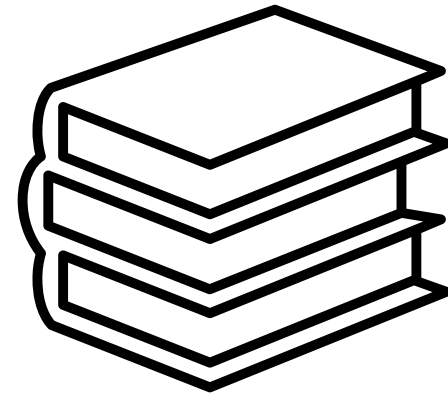


Validazione giuridica



3

**esperti
giuristi**



60

**estratti
semplificati**

Validazione giuridica – Risultati

Sono presenti tutte le informazioni essenziali?

96 %

Sono stati eliminati dei tecnicismi giuridici?

13 %

Sono stati introdotti errori?

12 %

Validazione giuridica – Limitazioni

Lo stile dei documenti semplificati non è coerente con quello lo stile
burocratico amministrativo

Il processo di semplificazione in alcuni casi non conserva il valore
giuridico

Prompt chaining vs zero-shot prompt

*Fiorentino, Ganfi, Russodivito, Cioffi, Simonelli (in stampa) «Simple Italian and AI: strenghts and weaknesses»
in Lebende Sprachen*

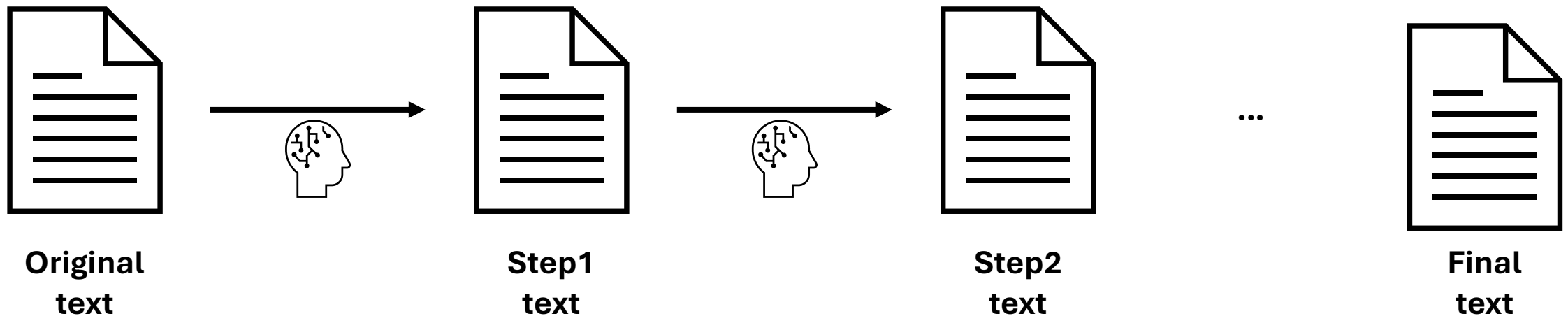


Prompt chaining

La semplificazione di un documento amministrativo è un **operazione complessa**.

Per mantenere la valenza giuridica è necessario avere controllo sul processo di semplificazione. Per fare ciò possiamo sfruttare il principio “**divide et impera**”.

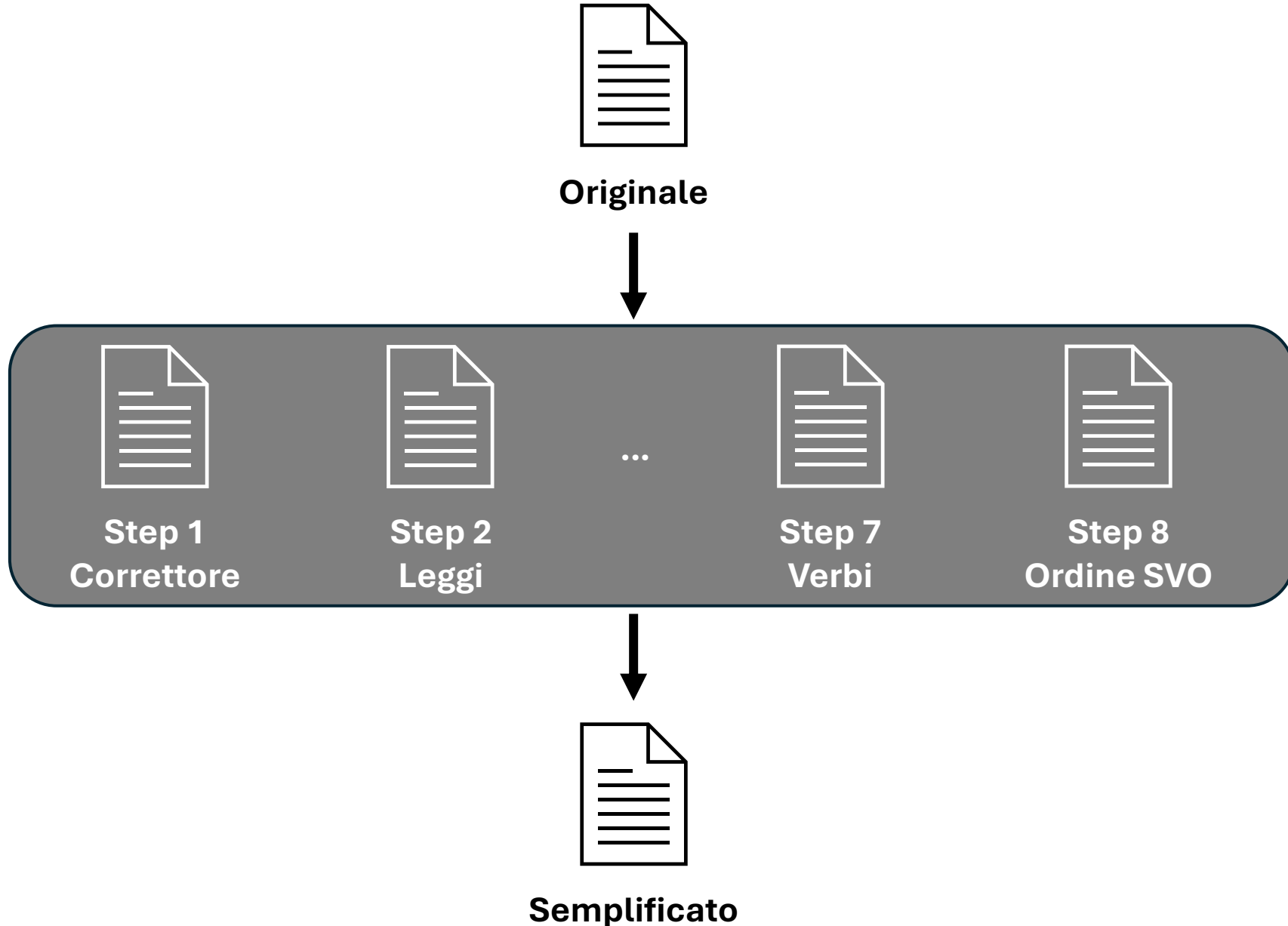
Tramite il **Prompt Chaining** è possibile avere una semplificazione **progressiva** e più **accurata**.



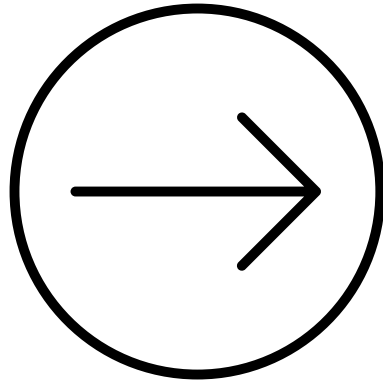
Prompt chaining – Regole di semplificazione

- 1) Correttore/Proofreading
- 2) Normalizzazione delle leggi
- 3) Sostituzione dei connettivi formali/burocratici
(altresì, nella fattispecie, ...)
- 4) Sostituzione delle espressioni formali/burocratiche
(è fatto divieto, ...)
- 5) Divisione delle frasi lunghe
- 6) Sostituzione delle nominalizzazioni deverbali
- 7) Trasformazione verbi passivi
- 8) Ordine SVO, soggetto esplicito, subordinate ai margini della principale

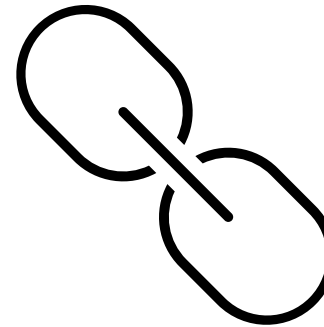
Prompt chaining – Regole di semplificazione



Prompt chaining – Confronto



ZERO-SHOT
prompting

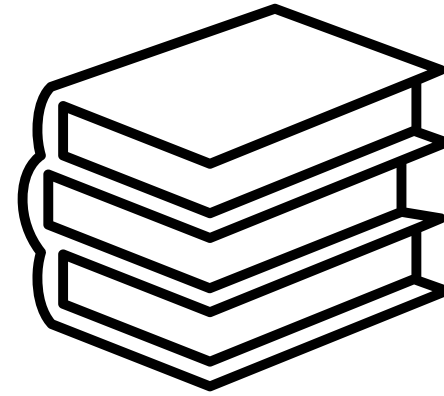


CHAIN
prompting

Prompt chaining – Confronto



2
giuristi



52 coppie
originale-semplificato

Prompt chaining – Risultato

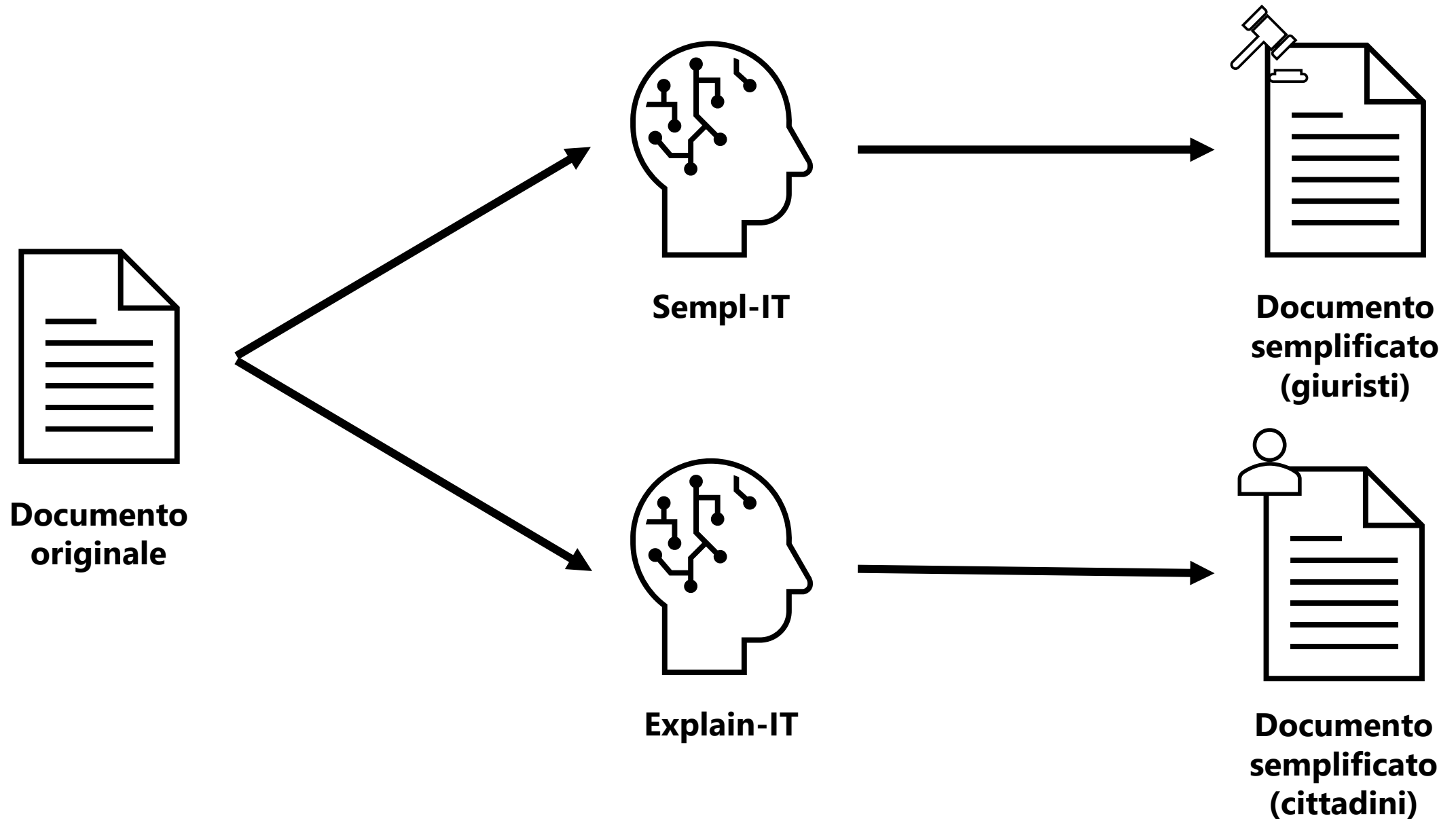
~76 %

Le coppie **ORIGINALE VS PROMPT CHAIN** sono
“**Perfettamente equivalenti**”

~51 %

Le coppie **ORIGINALE VS PROMPT ZERO-SHOT** sono
“**Abbastanza equivalenti**”.

Prompt chaining – Prospettive

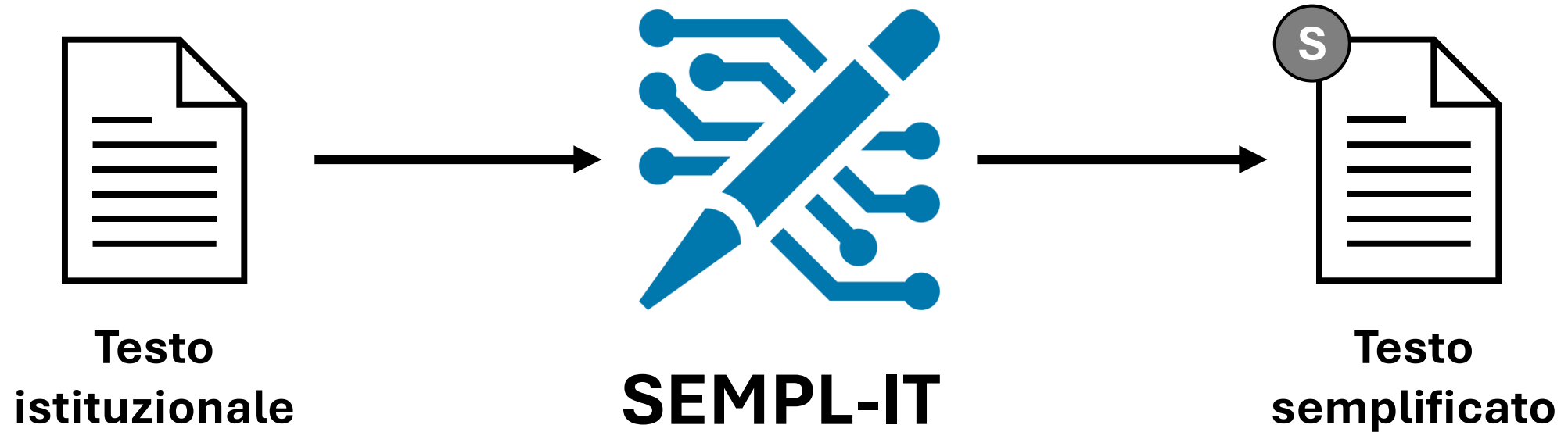


Creazione di un LLM per la semplificazione



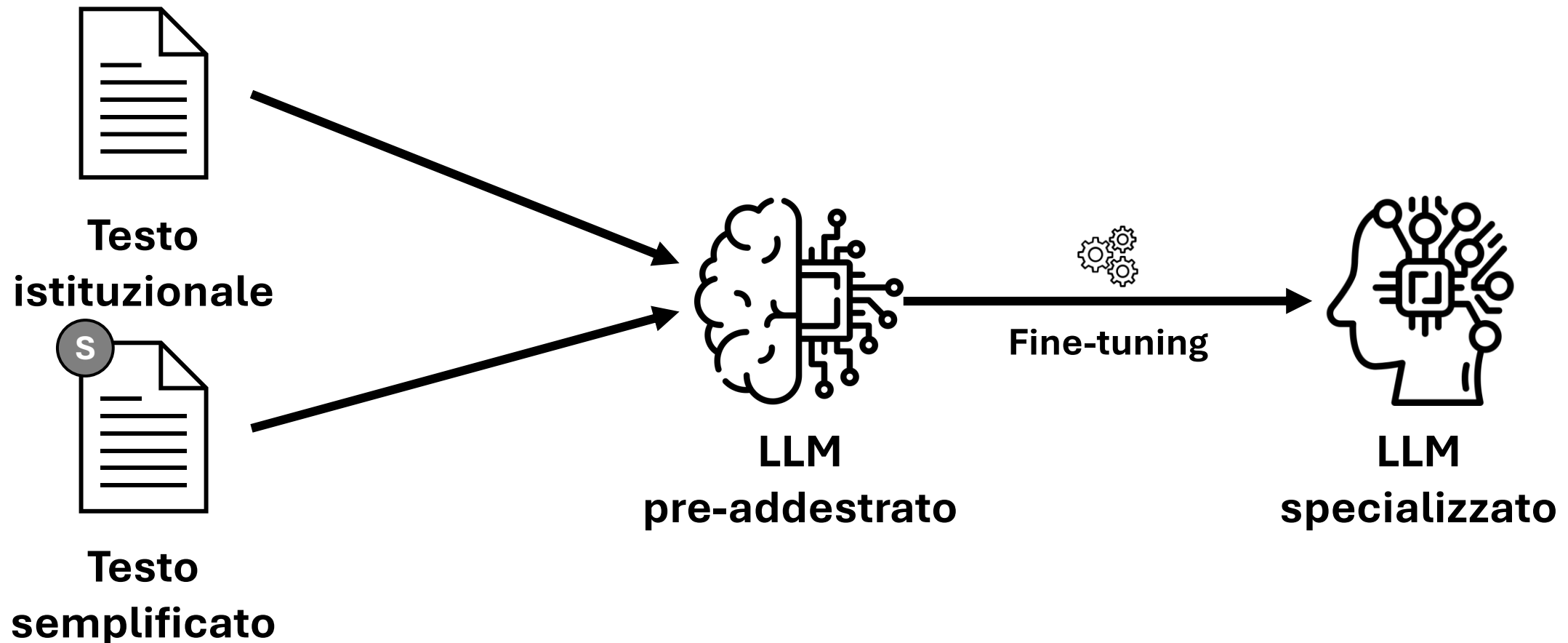
SEMPLE-IT

SEMPLE-IT è un **Automatic Text Simplifier** basato su intelligenza artificiale specializzato nella semplificazione di documenti istituzionali.

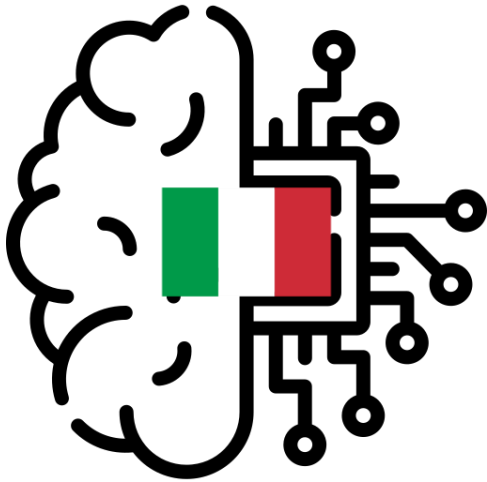


SEMPLE-IT – Fine-tuning

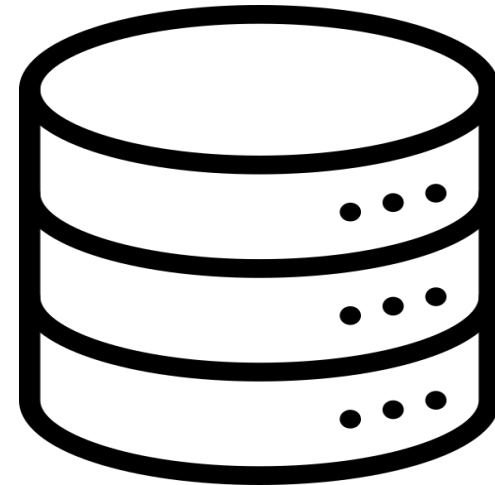
La **specializzazione** (fine-tuning) di un LLM consiste nel fornire numerose coppie di testi originali e semplificati, affinché apprenda il modo in cui semplificare il testo.



Fine-tuning – Requisiti



**Modello
pre-addestrato**

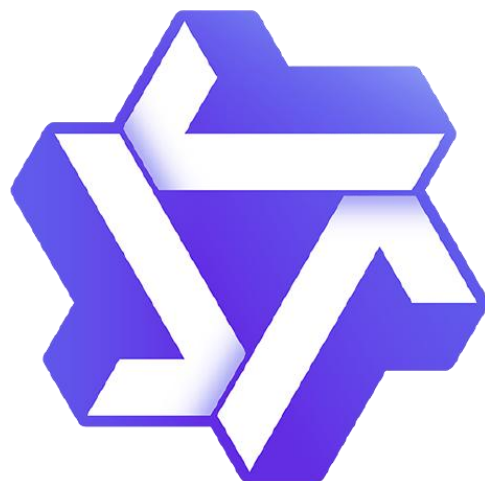


**Corpus
parallelo**

Modello pre-addestrato

Per l'addestramento di SEMPL-IT, è stato scelto il modello **Qwen2.5** nella variante da **7B parametri**.

- Ottime performance linguistiche sull'italiano.
- Architettura moderna ed efficiente



arXiv:2412.15115v2 [cs.CL] 3 Jan 2025

 **Qwen**

2025-01-06

Qwen2.5 Technical Report

Qwen Team

 <https://huggingface.co/Qwen>
 <https://modelscope.cn/organization/qwen>
 <https://github.com/QwenLM/Qwen2.5>

Abstract

In this report, we introduce Qwen2.5, a comprehensive series of large language models (LLMs) designed to meet diverse needs. Compared to previous iterations, Qwen 2.5 has been significantly improved during both the pre-training and post-training stages. In terms of pre-training, we have scaled the high-quality pre-training datasets from the previous 7 trillion tokens to 18 trillion tokens. This provides a strong foundation for common sense, expert knowledge, and reasoning capabilities. In terms of post-training, we implement intricate supervised finetuning with over 1 million samples, as well as multistage reinforcement learning, including offline learning DPO and online learning GRPO. Post-training techniques significantly enhance human preference, and notably improve long text generation, structural data analysis, and instruction following.

To handle diverse and varied use cases effectively, we present Qwen2.5 LLM series in rich configurations. The open-weight offerings include base models and instruction-tuned models in sizes of 0.5B, 1.5B, 3B, 7B, 14B, 32B, and 72B parameters. Quantized versions of the instruction-tuned models are also provided. Over 100 models can be accessed from Hugging Face Hub, ModelScope, and Kaggle. In addition, for hosted solutions, the proprietary models currently include two mixture-of-experts (MoE) variants: Qwen2.5-Turbo and Qwen2.5-Plus, both available from Alibaba Cloud Model Studio.

Qwen2.5 has demonstrated top-tier performance on a wide range of benchmarks evaluating language understanding, reasoning, mathematics, coding, human preference alignment, etc. Specifically, the open-weight flagship Qwen2.5-72B-Instruct outperforms a number of open and proprietary models and demonstrates competitive performance to the state-of-the-art open-weight model, Llama-3-405B-Instruct, which is around 5 times larger. Qwen2.5-Turbo and Qwen2.5-Plus offer superior cost-effectiveness while performing competitively against GPT-4o-mini and GPT-4o respectively. Additionally, as the foundation, Qwen2.5 models have been instrumental in training specialized models such as Qwen2.5-Math (Yang et al., 2024b), Qwen2.5-Coder (Hui et al., 2024), QwQ (Qwen Team, 2024d), and multimodal models.

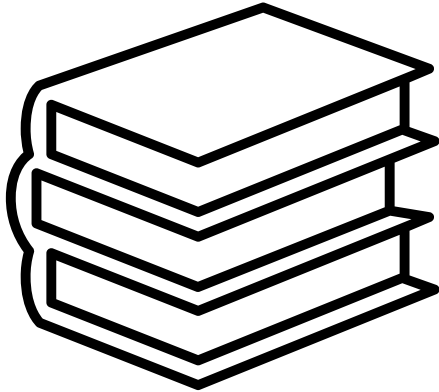


Figure 1: In the iterative development of the Qwen series, data scaling has played a crucial role. Qwen 2.5, which leverages 18 trillion tokens for pre-training, has demonstrated the most advanced capabilities within the Qwen series, especially in terms of domain expertise, underscoring the importance of scale together with mixture in enhancing the model's capabilities.

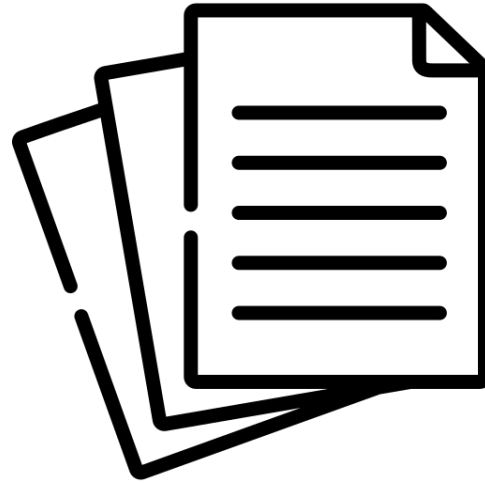
1

Corpus

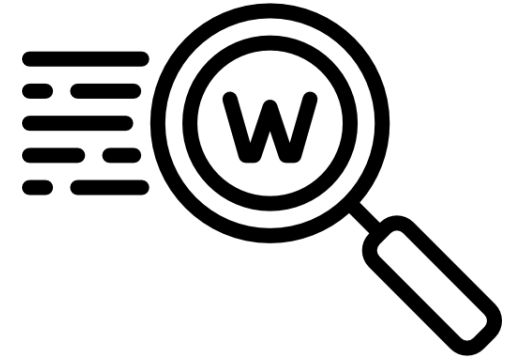
Per l'addestramento di SEMPL-IT, è stato utilizzato un sottoinsieme del corpus Italst.



24
documenti



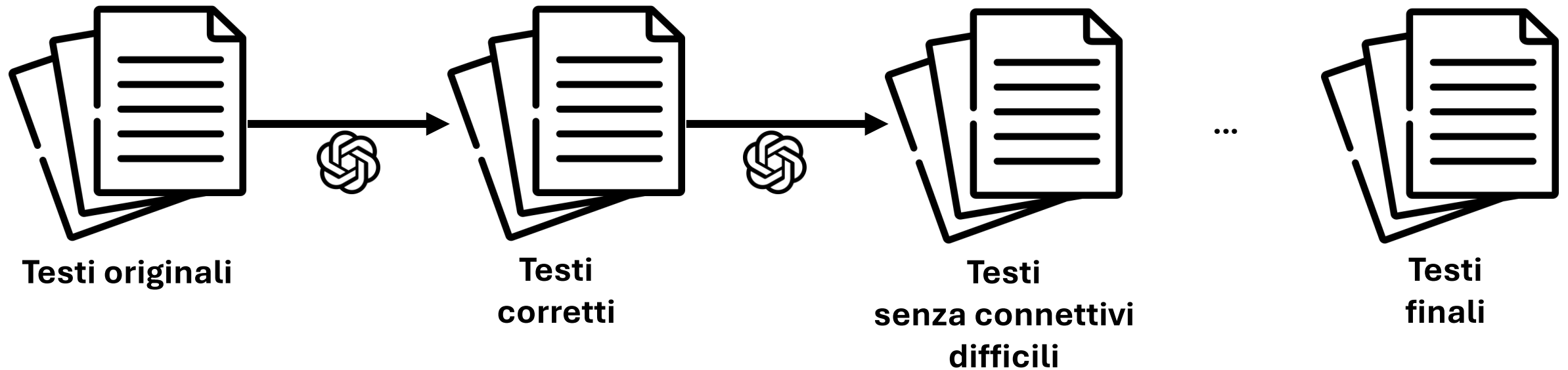
1020
capitoli



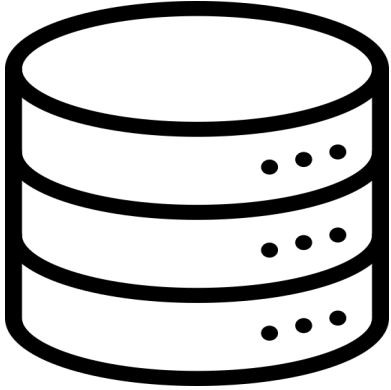
180 000
tokens

Corpus – Semplificazione

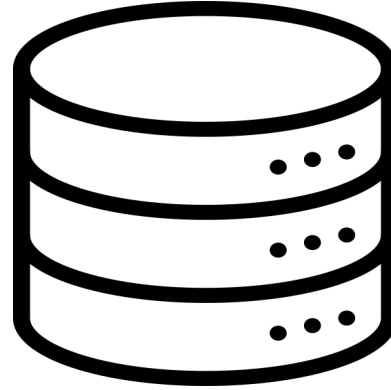
Per la creazione del corpus parallelo abbiamo semplificato il corpus di partenza applicando ogni regola di semplificazione tramite **GPT-4o** e il **prompt chaining**.



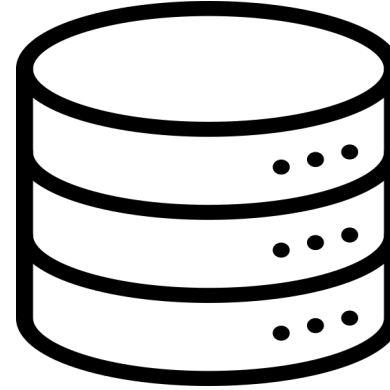
Corpora paralleli



Proofreading



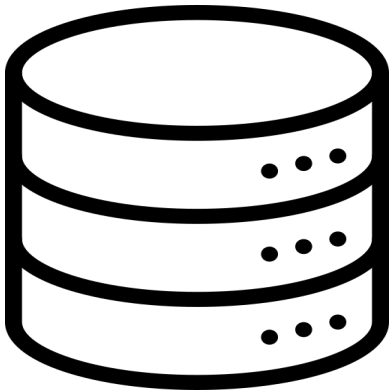
**Normalizzazione
riferimenti**



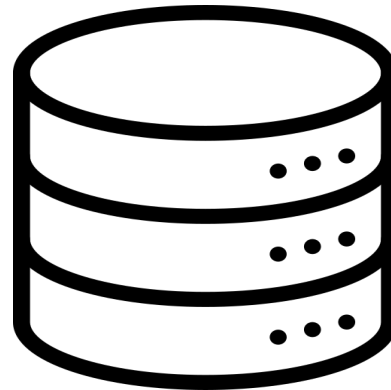
**Connettivi
difficili**



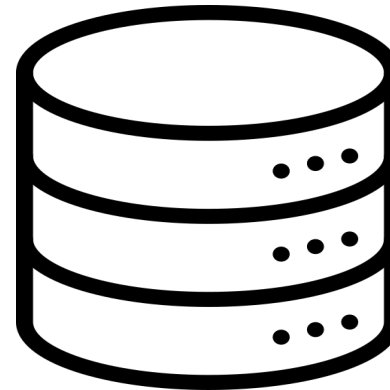
**Espressioni
burocratiche**



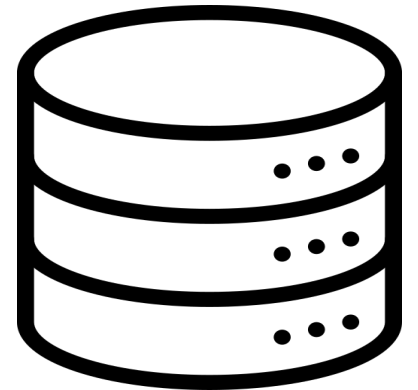
Divisione frasi



Nominalizzazioni



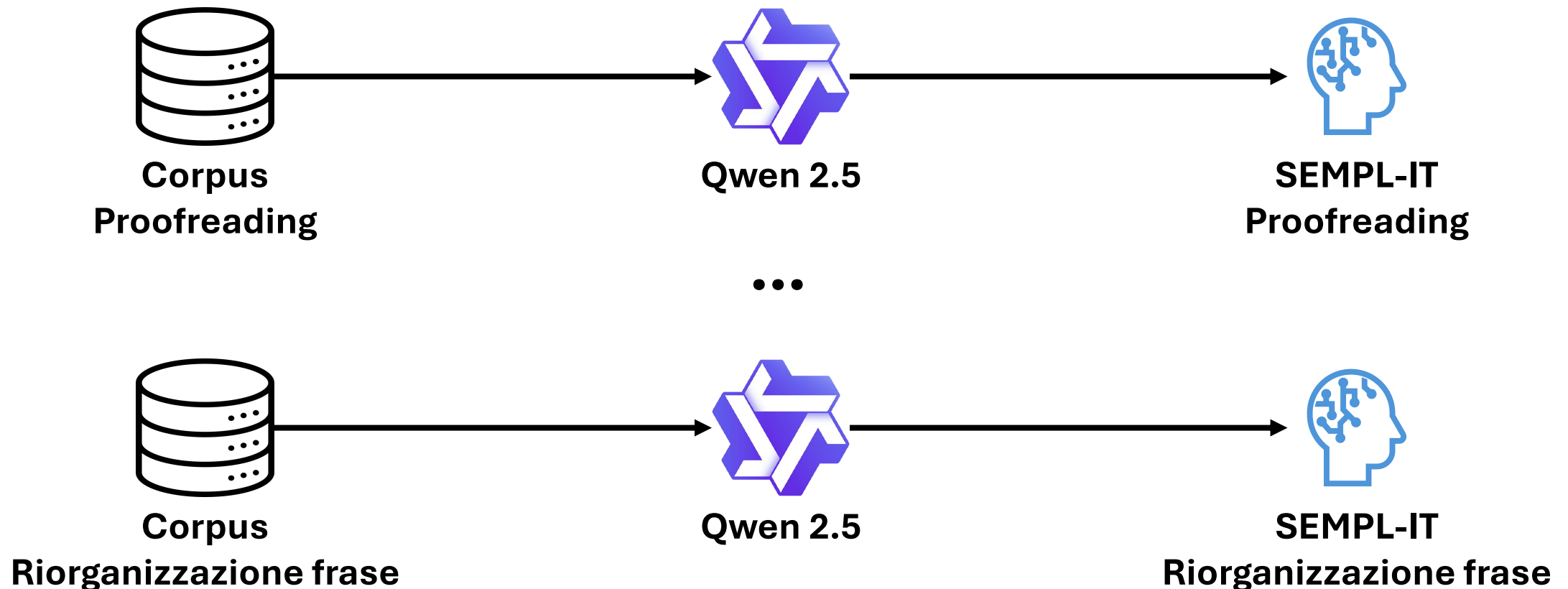
Verbi passivi



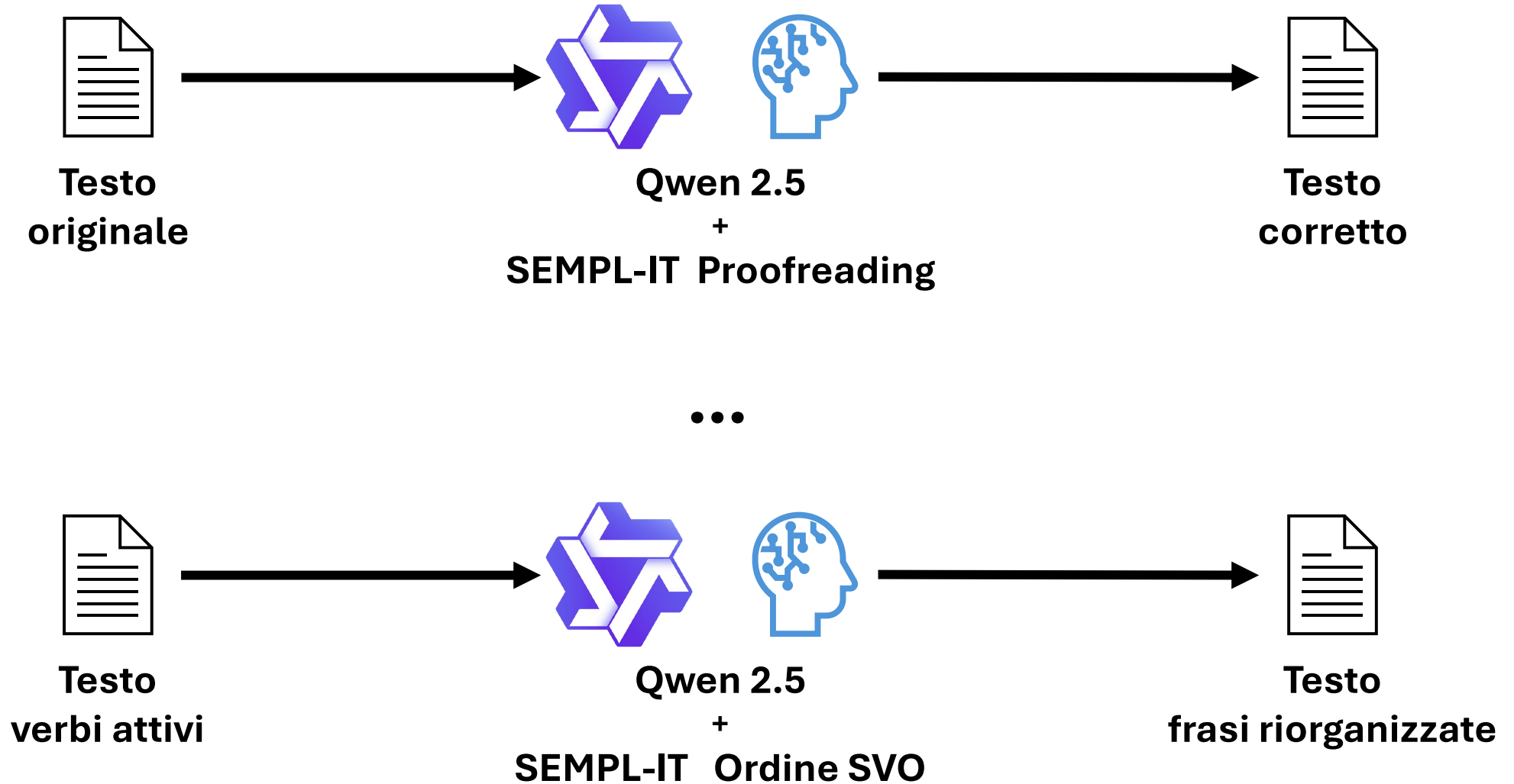
**Riorganizzazione
frase**

SEMPLE-IT – Fine-tuning

Sono stati addestrati 8 modelli specializzati nell'applicazione di una regola di semplificazione tramite una procedura di finetuning particolare chiamata PEFT (Parameter Efficient Finetuning).



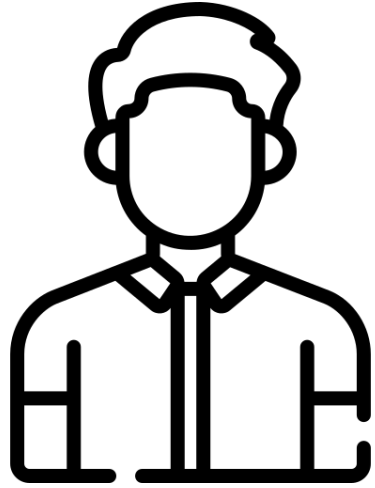
SEMPLE-IT – Utilizzo



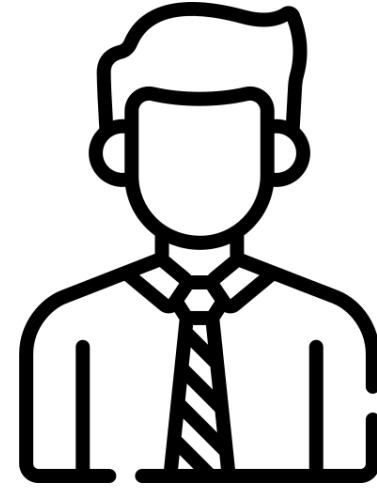
SEMP-IT – Web App



SEMPLE-IT – Web App

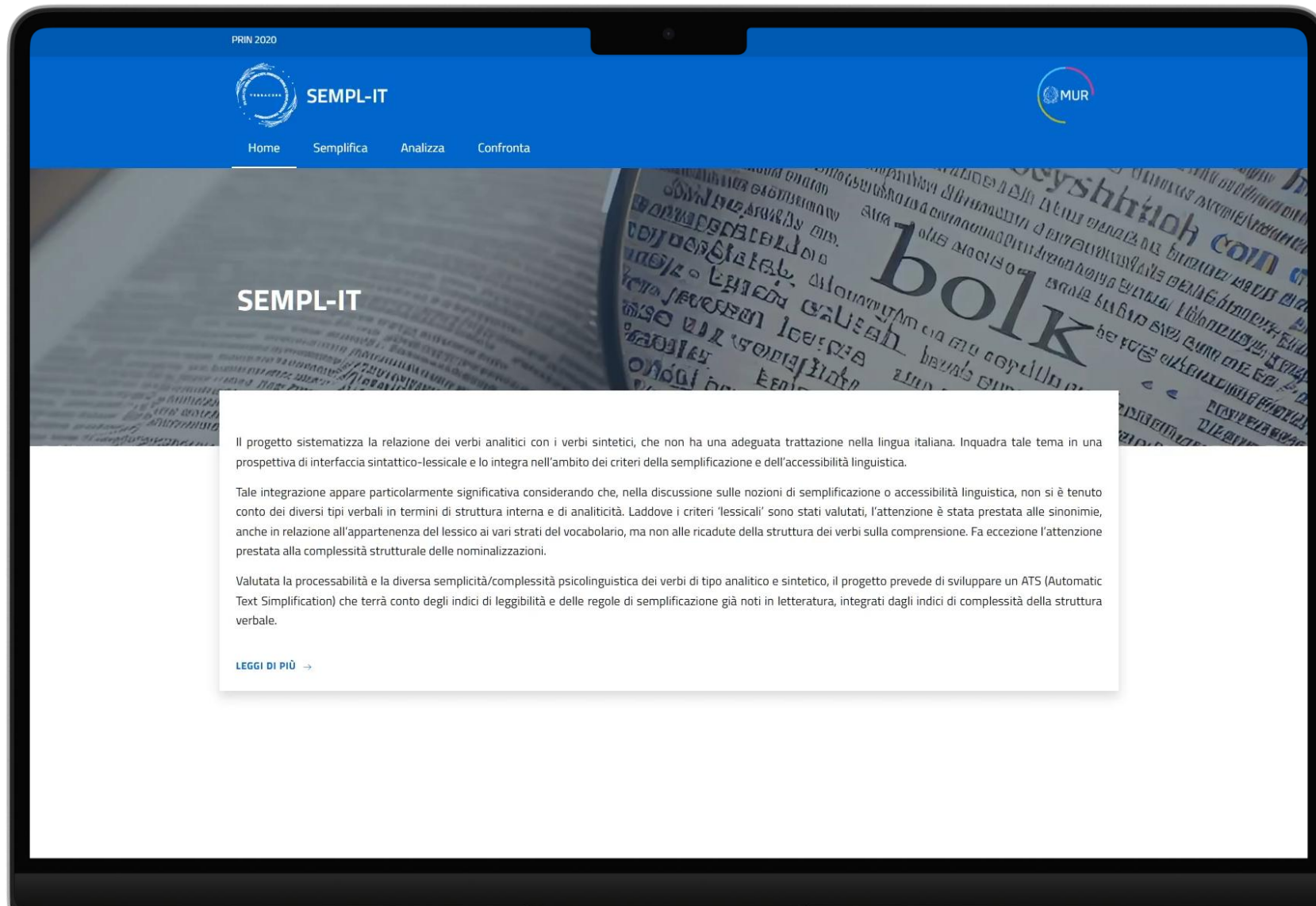


Cittadini

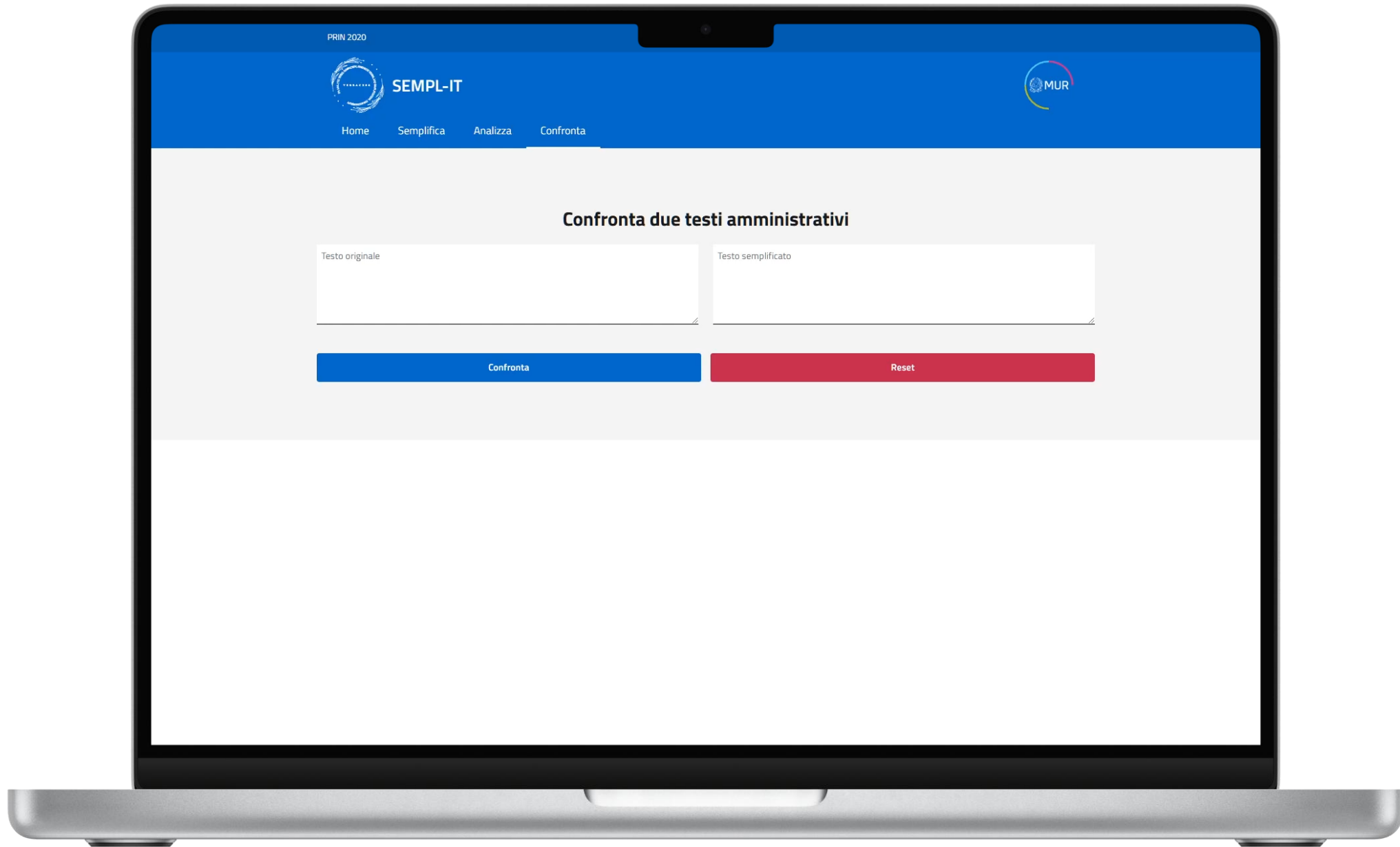


Esperti

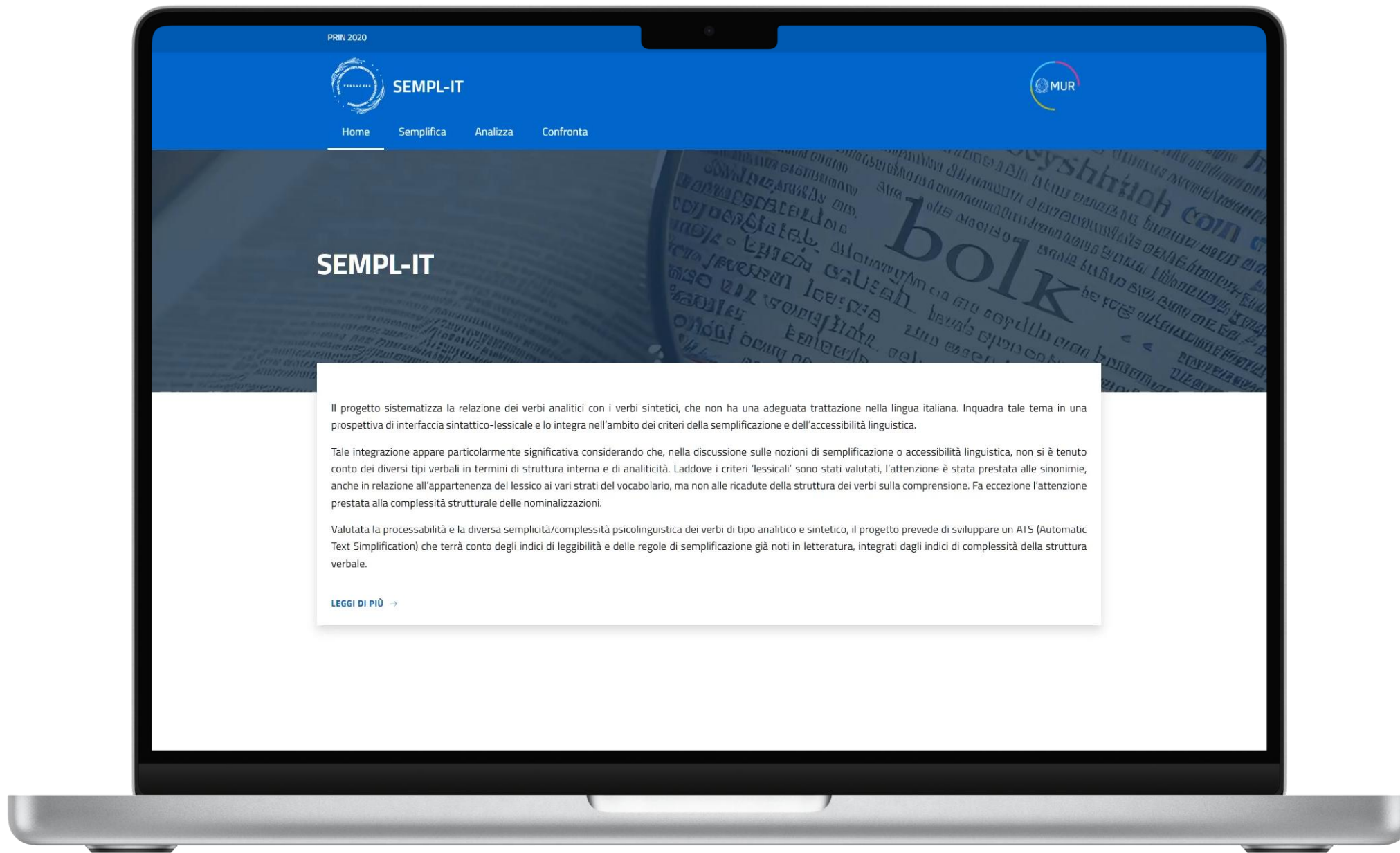
SEMPLE-IT – Analisi



SEMPLE-IT – Confronta



SEMPLE-IT – Semplifica



PRIN 2020



SEMPLE-IT



Home

Semplifica

Analizza

Confronta

SEMPLE-IT

Il progetto sistematizza la relazione dei verbi analitici con i verbi sintetici, che non ha una adeguata trattazione nella lingua italiana. Inquadra tale tema in una prospettiva di interfaccia sintattico-lessicale e lo integra nell'ambito dei criteri della semplificazione e dell'accessibilità linguistica.

Tale integrazione appare particolarmente significativa considerando che, nella discussione sulle nozioni di semplificazione o accessibilità linguistica, non si è tenuto conto dei diversi tipi verbali in termini di struttura interna e di analiticità. Laddove i criteri 'lessicali' sono stati valutati, l'attenzione è stata prestata alle sinonimie, anche in relazione all'appartenenza del lessico ai vari strati del vocabolario, ma non alle ricadute della struttura dei verbi sulla comprensione. Fa eccezione l'attenzione prestata alla complessità strutturale delle nominalizzazioni.

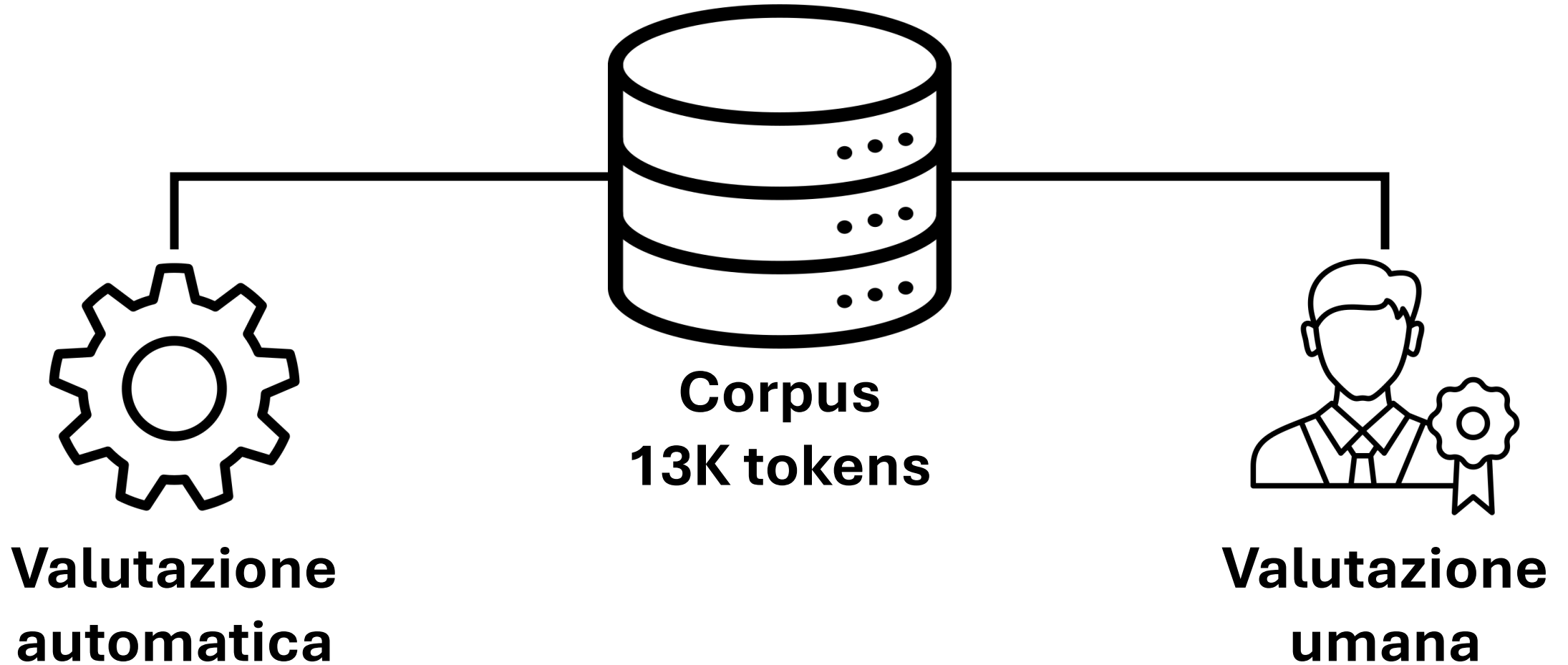
Valutata la processabilità e la diversa semplicità/complessità psicolinguistica dei verbi di tipo analitico e sintetico, il progetto prevede di sviluppare un ATS (Automatic Text Simplification) che terrà conto degli indici di leggibilità e delle regole di semplificazione già noti in letteratura, integrati dagli indici di complessità della struttura verbale.

[LEGGI DI PIÙ](#) →

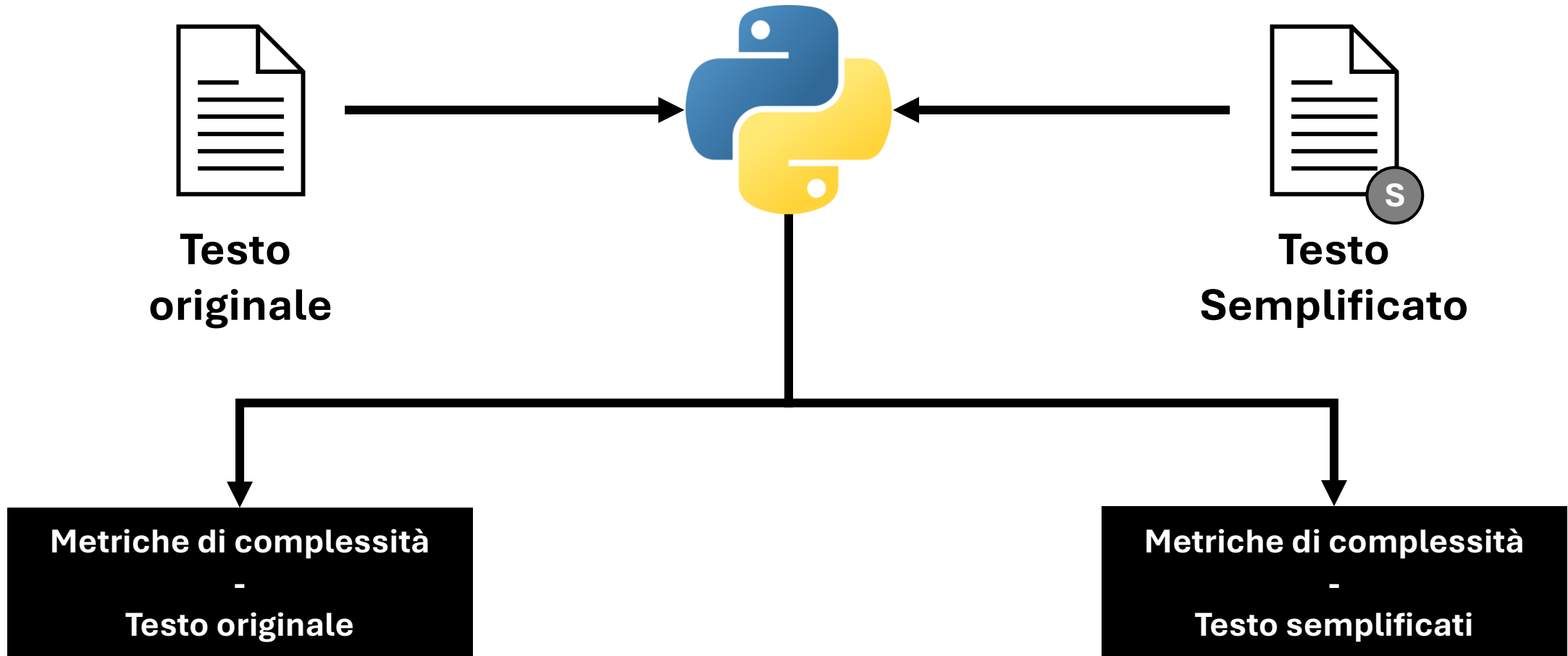
Validazione SEMPL-IT



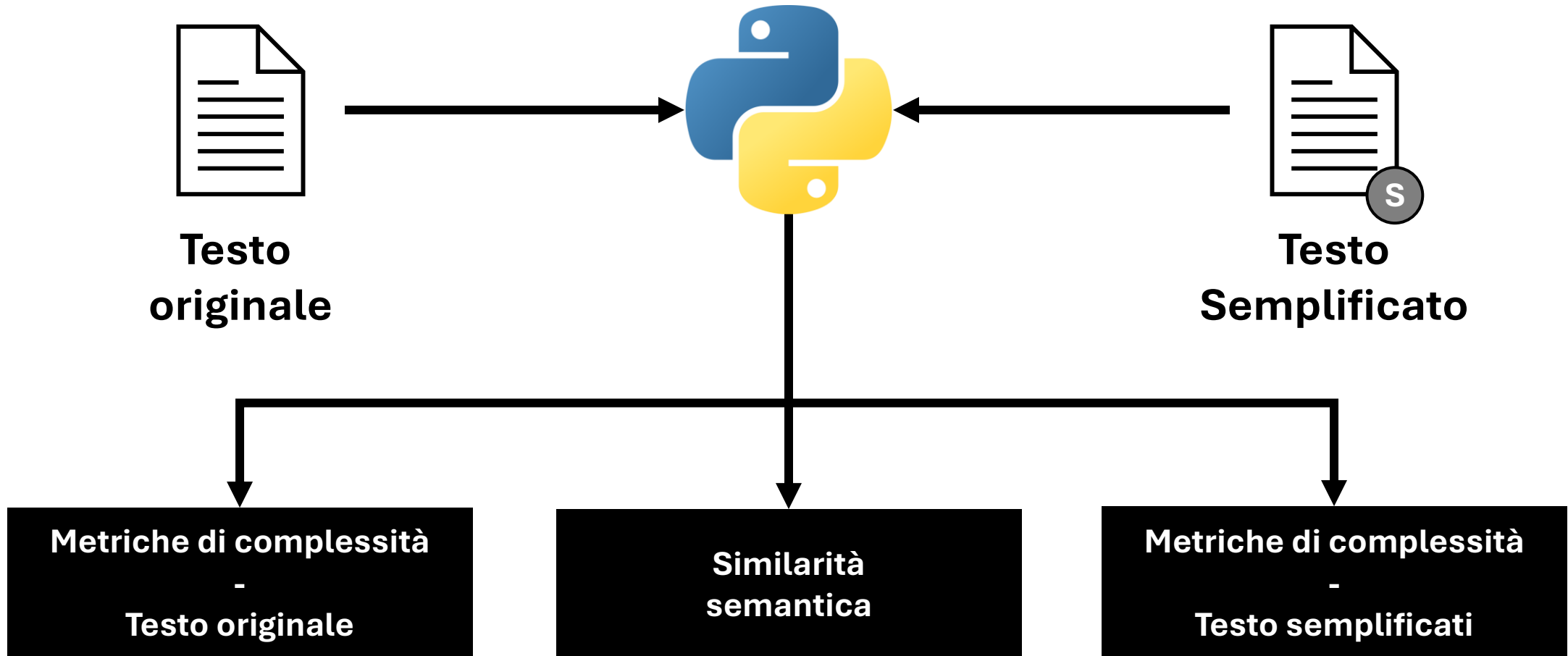
Validazione



Validazione automatica



Validazione automatica



Validazione automatica – Risultati

	Token	Frase	% Verbi Attivi	% Vdb	Connettivi Difficili	Indice Gulpease	Flesch Vacca	Similarità Semantica
Originale	13720	490	84.17	72.32	166	41.37	16.56	-
Proofreading	13754	503	83.54	72.28	167	41.73	17.11	99.75
Leggi	13729	504	83.54	72.30	167	41.72	17.20	99.68
Connettivi difficili	13671	504	83.50	72.84	129	41.78	17.55	99.53
Espressioni burocratiche	13208	505	84.27	73.21	109	42.21	18.43	99.14
Divisione frasi	13531	706	80.61	74.03	97	46.86	28.00	98.90
Nominaliz.	13471	707	82.26	75.63	94	47.19	28.21	98.75
Verbi	13433	707	87.86	75.60	94	47.28	28.52	98.51
Riorganizzaz. Fase	13422	707	87.79	75.72	94	47.37	28.93	98.47

Validazione automatica – Risultati

Il numero di frasi e la percentuale di verbi attivi
aumentano significativamente

Il numero di connettivi difficili
diminuisce significativamente

Gli indici Gulpease e Flesch-Vacca
migliorano significativamente

La similarità semantica è molto **elevata**

Validazione manuale – Risultati

È stata applicata la regola di semplificazione?

[SI, NO]

Sono state eseguite delle operazioni aggiuntive?

[SI, NO]

Sono stati introdotti degli errori?

[SI, NO]

Validazione manuale – Risultati

	È stata applicata la regola di semplificazione?	Sono state eseguite delle operazioni aggiuntive?	Sono stati introdotti degli errori?
Proofreading	94,74 %	42,11 %	14,04 %
Leggi	100 %	7,02 %	10,53 %
Connettivi	66,67 %	29,82%	1,75%
Espressioni	91,23 %	24,56 %	10,53 %
Divisione Frase	100 %	0 %	7,02 %
Nominaliz.	78,95 %	22,81 %	31,58 %
Verbi	82,46 %	43,86 %	33,33 %
Riorganizzaz. Fase	57,89 %	7,02 %	5,26 %

Commento linguistico – Limiti

Modifica il numero e non elimina la nominalizzazione

Nominalizzazione

Originale

Inoltre, hanno situazioni familiari e personali complesse e necessitano di un intervento multidisciplinare. **Questo intervento** dovrebbe coinvolgere più servizi del sistema sanitario e sociale territoriale.

LLM

Inoltre, hanno situazioni familiari e personali complesse e necessitano di interventi multidisciplinari. **Questi interventi** dovrebbero coinvolgere più servizi del sistema sanitario e sociale territoriale.

Eliminazione dei verbi copulativi o passivi e stravolgimento della struttura concettuale dell'originale:

Verbi

Originale

L'obiettivo **è favorire** stili di vita sani

LLM

L'obiettivo **favorirà** stili di vita sani

Commento linguistico – Limiti

Stravolge la struttura concettuale dell'originale, dopo aver trasformato in attivo un passivo o aver trasformato una nominalizzazione in un verbo pieno

Verbi	
Originale	la comunicazione deve essere sottoscritta da un legale rappresentante
LLM	la comunicazione sottoscrive un legale rappresentante

Verbi	
Originale	La garanzia deve essere prestata con le modalità
LLM	L'ente deve prestare la garanzia

Nominalizzazioni	
Originale	Ci sono prove della produzione e diffusione di buone pratiche.
LLM	Ci sono prove che producono e diffondono buone pratiche.

Commento linguistico – Limiti del Prompt chaining

Propagazione dell'errore: un errore nella sostituzione dei verbi ha condotto a delle strutture incoerenti sul piano semantico.

La riorganizzazione della frase consolida l'errore di interpretazione della fase di prima.

Verbi

Originale

Disposizioni anticipate di **trattamento: indirizzare e fruire,**
armonizzare i sistemi di tenuta e deposito delle DAT

LLM

Indirizziamo e fruiamo delle disposizioni anticipate di trattamento.

Riorganizzazione della frase

Originale

Indirizziamo e fruiamo delle disposizioni anticipate di trattamento.

LLM

Indirizziamo le disposizioni anticipate di trattamento e fruiamo
delle disposizioni anticipate di trattamento.

Commento linguistico – Semplificazioni efficaci

Viene sostituita l'espressione burocratica con una più disponibile

Espressioni	
Originale	la presa in carico multidisciplinare e multiprofessionale
LLM	la gestione multidisciplinare e multiprofessionale

Espressioni	
Originale	Vengono fornite ad ogni paziente informazioni riguardanti la preparazione ai trattamenti
LLM	I pazienti ricevono informazioni sulle preparazioni ai trattamenti

Espressioni	
Originale	Ad essi competono le seguenti funzioni
LLM	Le loro funzioni sono

Commento linguistico – Semplificazioni efficaci

Sostituzione delle costruzioni con nominalizzazioni o passivi con strutture più accessibili

Nominalizzazioni	
Originale	Vi è evidenza dell'implementazione di tutte le buone pratiche regionali e delle raccomandazioni ministeriali di pertinenza.
LLM	Vi è evidenza che tutte le buone pratiche regionali e le raccomandazioni ministeriali di pertinenza siano state implementate.

Verbi	
Originale	La vendita dei biglietti e degli abbonamenti per gli autobus urbani, nonché dei titoli di parcheggio a raso, avviene presso tutti gli esercizi commerciali convenzionati.
LLM	I commercianti convenzionati vendono i biglietti e gli abbonamenti per gli autobus urbani, nonché i titoli di parcheggio a raso.

Commento linguistico – Semplificazioni efficaci

Capacità di recuperare ed esplicitare una anafora zero nelle catene anaforiche

Divisore di frasi	
Originale	Inoltre, ha scelto gli indicatori chiave per valutare la performance clinico-assistenziale/riabilitativa, basandosi sulla "scienza" o sull'"evidenza" che supporta ogni indicatore.
LLM	Inoltre, ha scelto gli indicatori chiave per valutare la performance clinico-assistenziale/riabilitativa. Questa scelta è stata basata sulla "scienza" o sull'"evidenza" che supporta ogni indicatore.

Divisore di frasi	
Originale	La demenza presenta una complessità trasversale all'ambito clinico, assistenziale e riabilitativo, che ha ripercussioni cliniche, operative ed organizzative
LLM	La demenza presenta una complessità trasversale all'ambito clinico, assistenziale e riabilitativo. Questa complessità ha ripercussioni cliniche, operative ed organizzative.

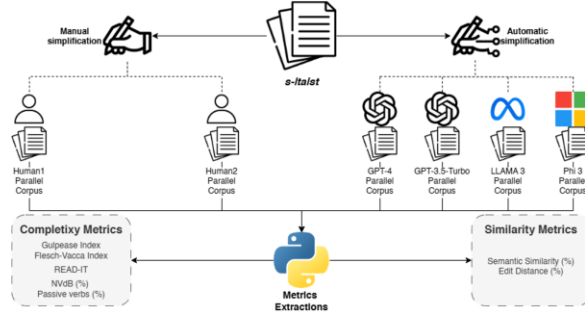
Commento linguistico – Efficacia del Prompt chaining

Riorganizzazione della frase → ristabilisce l'ordine meno marcato, dopo che un testo ha subito varie trasformazioni

Riorganizzazione di frasi	
Originale	Questi interventi i competenti Uffici comunali o i singoli cittadini segnalano.
LLM	I competenti Uffici comunali o i singoli cittadini segnalano questi interventi.

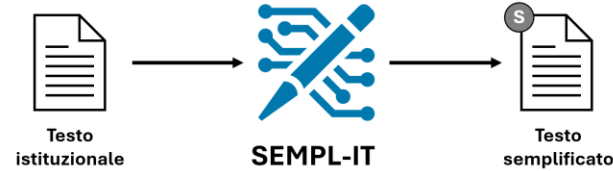
Riassunto

AI vs Umano – Disegno dello studio



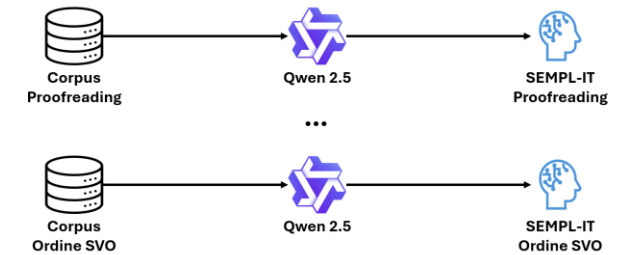
SEMPLE-IT

SEMPLE-IT è un **Automatic Text Simplifier** basato su intelligenza artificiale specializzato nella semplificazione di documenti istituzionali.



SEMPLE-IT – Fine-tuning

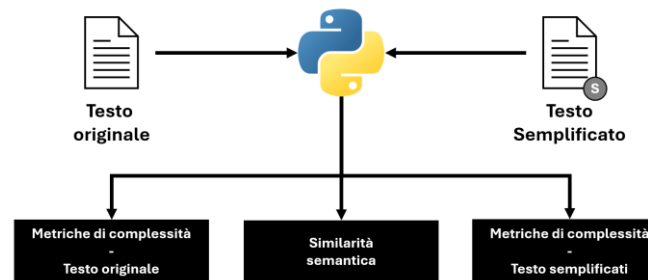
Sono stati addestrati 8 modelli specializzati nell'applicazione di una regola di semplificazione tramite una procedura di finetuning particolare chiamata PEFT (Parameter Efficient Finetuning).



SEMPLE-IT – Web App

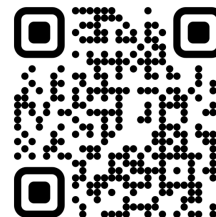


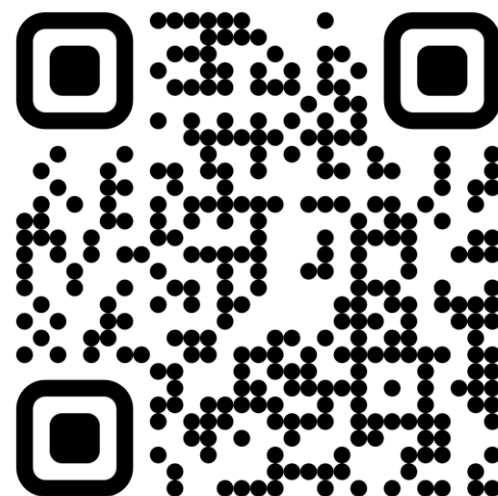
Validazione automatica



Validazione manuale – Risultati

	È stata applicata la regola di semplificazione?	Sono state eseguite delle operazioni aggiuntive?	Sono stati introdotti degli errori?
Proofreading	94,74 %	42,11 %	14,04 %
Leggi	100 %	7,02 %	10,53 %
Connettivi	66,67 %	29,82 %	1,75 %
Espressioni	91,23 %	24,56 %	10,53 %
Divisione Frasi	100 %	0 %	7,02 %
Nominaliz.	78,95 %	22,81 %	31,58 %
Verbi	82,46 %	43,86 %	33,33 %
Riorganizzaz. Fase	57,89 %	7,02 %	5,26 %





Grazie per l'attenzione!

giuliana.fiorentino@unimol.it

vittorio.ganfi@unimol.it

marco.russodivito@unimol.it