From Words to Action: A National Initiative to Overcome Data Scarcity for the Slovene LLM

Špela Arhar Holdt
University of Ljubljana
arharhs@ff.uni-lj.si

Špela Antloga
University of Ljubljana; University of Maribor
spela.antloga@fri.uni-lj.si

Tina Munda
University of Ljubljana
tina.munda@gmail.com

Eva Pori
University of Ljubljana
eva.pori@ff.uni-lj.si

Simon Krek
University of Ljubljana; IJS
simon.krek@ijs.si

Abstract

Large Language Models (LLMs) have demonstrated significant potential in natural language processing, but they depend on vast, diverse datasets, creating challenges for languages with limited resources. The paper presents a national initiative that addresses these challenges for Slovene. We outline strategies for large-scale text collection, including the creation of an online platform to engage the broader public in contributing texts and a communication campaign promoting openly accessible and transparently developed LLMs.

1 Introduction

Extremely large language models, such as GPT-4, have demonstrated remarkable advancements across various natural language processing tasks, sparking widespread interest in their applications. However, their reliance on vast and diverse datasets makes them inherently biased toward well-resourced languages. For languages like Slovene, with a smaller speaker base and limited data availability, this disparity poses a significant challenge, hindering the development of robust language-specific LLMs.

Recent studies have highlighted similar challenges faced by other low-resource languages, underscoring the need for comprehensive multilingual evaluation and language-specific model development. (Lai et al., 2023) provide an in-depth assessment of ChatGPT's performance across multiple languages, revealing its uneven effectiveness in low-resource contexts. Their evaluation, covering seven tasks and 37 languages, exposes significant performance gaps in both low- and extremely low-resource languages. Likewise, (Alam et al., 2024) explore the broader landscape of LLMs for low-resource languages, addressing their multilingual, multimodal, and dialectal complexities. Their findings emphasize the necessity of language-specific initiatives and reveal the persistent limitations of LLMs for medium- to low-resource languages, largely due to the lack of representative datasets.

While these studies highlight the performance-related limitations of current LLMs for low-resource languages, an equally pressing concern is their accessibility. The proprietary nature and high computational demands of existing models restrict access for many research organizations and smaller companies, underscoring the need for open-access alternatives. Addressing this issue requires not only the development of computationally efficient models, but also the creation of large, diverse, and high-quality datasets tailored to specific languages.

For smaller language communities, such as Slovene, the search for texts to be included in LLMs must extend well beyond readily available online sources, as these alone are insufficient to support the development. In this paper, we introduce a national initiative aimed at overcoming these obstacles for Slovene, detailing our strategies for large-scale text collection and community engagement for the development of openly available Slovene LLMs.

2 Project framework and previous work

LLMs have introduced a major shift in the field of Natural Language Processing (NLP), offering more efficient fine-tuning for common NLP tasks while simplifying their implementation (Brown et al., 2020). The datasets serve as the foundational infrastructure analogous to a root system that sustains and nurtures the development of LLMs (Liu et al., 2024). Therefore, preparing language data for LLMs is a crucial step that directly impacts their performance across various tasks.

The effectiveness of NLP tasks relies heavily on the scale of the language model's training, which is directly influenced by access to large, diverse datasets spanning various domains (Kaplan et al., 2020). Moreover, diversity in training data plays a crucial role in enhancing the generalization capabilities of large models, enabling downstream tasks, as outlined in (Ali et al., 2019), to effectively leverage knowledge even with limited training data (Brown et al., 2020).

In the ongoing *PoVeJMo—Adaptive Natural Language Processing with Large Language Models* project,[1] we aim to develop several computationally efficient and open-access large language models trained on Slovene language data. These models will also be adapted and evaluated for selected industry use cases, such as enhancing Slovene speech recognition and synthesis for industrial systems, preparing museum materials and interactive systems, as well as for medical applications and infrastructure code generation.

We have first gathered 9.2 billion tokens from freely available sources, including existing language corpora and other openly accessible Slovene-language data, providing a solid foundation for the project. The initial dataset includes different types of text, such as news articles up to and including September 2023 (Kosem et al., 2023), academic works (Žagar et al., 2022), web crawls (mC4 (Raffel et al., 2020), MaCoCu (Bañón et al., 2023), CC100 (Wenzek et al., 2019)), and various Slovene reference and specialized corpora included in the Metafida database (Erjavec, 2023)). The GaMS-1B-Chat language model, with one billion parameters, has been trained on this language material (Vreš et al., 2024).

While this initial model provides valuable insights into the effects of training on Slovene data, it is roughly a thousand times smaller than the largest models (e.g., used for the latest version of ChatGPT), makes errors frequently, and highlights the need for further text collection to improve performance.[2]

3 National text-collection campaign

We have estimated that the development of a sufficiently large model requires approximately 40 billion additional words. The estimation of the required amount of training data is heuristic and approximate, derived from two approaches. The first approach is based on the findings of the LLama model study (Touvron et al., 2023). The study illustrates the decline in the loss function as the number of training tokens increases, where, on average, two tokens correspond to one word. For most models, including LLaMa 7B, which is the closest equivalent to GaMS, the most significant decrease in the loss function occurs up to approximately 100 billion tokens (that is, around 50 billion words). We expect to collect around 10 billion words from freely available online resources (currently 9.2 billion; see Section 2), while an additional 40 billion will need to be gathered using alternative approaches.[3] The second approach relies on LLM scaling laws, which suggest that a model of size *x* requires at least *5x* to *10x* words for effective training, though in practice the requirement is often even greater. Given that the larger GaMS model is expected to have approximately 10 billion parameters, it would require a minimum of 50 billion words to achieve optimal performance.

This ambitious undertaking necessitated the development of a comprehensive communication and operational strategy. The key components of this strategy include: (1) identifying and engaging potential contributors of textual materials; (2) developing efficient mechanisms for text submission; (3) implementing secure and scalable storage systems; (4) establishing effective and reliable processes for tracking and documenting all text acquisition activities; and (5) defining the metadata framework to ensure systematic organization and accessibility of collected texts. Beyond these operational aspects, strategy (6) addresses legal and ethical considerations associated with data collection and usage, while also prioritizing (7) promotional and dissemination activities to build public awareness and support. A key objective of these efforts is to emphasize the importance of developing a large-scale language model for Slovene, highlighting its far-reaching implications for tech-

[1] The project is funded between 2023 and 2026 by the Slovenian Research and Innovation Agency ARIS and the EU Recovery and Resilience Facility. More information on `https://www.cjvt.si/povejmo/en/project/`.

[2] GaMS-1B-Chat can be tested at `https://povejmo.si/klepet/` (at the moment, the interface is only available in Slovene).

[3] The observed trend indicates that the loss function continues to decrease even to 1,500 billion tokens, and in the case of LLaMa 3, where 3,000 billion tokens were used, the loss function still exhibited a downward trajectory, although more pronounced for larger models.

nological innovation and cultural preservation.

Our text collection campaign operates through two main strategies. On the one hand, we engage with large-scale text providers such as national libraries, publishing houses, media organizations, government ministries, and other significant contributors, strongly advocating for the value and potential impact of creating a Slovene LLM and encouraging them to contribute their textual resources. On the other hand, we reach out to individuals, inviting them to donate their own texts to both actively support and directly contribute to the co-creation of the Slovene language model.

Each of these two groups presents unique requirements and demands distinct approaches to communication and engagement. Beyond the technical challenges of acquiring and processing the material, the primary obstacles lie in legal constraints. Current Slovene legislation permits the use of copyrighted material for data mining. However, this does not apply to material available on the web unless it is provided under an appropriate license. However, this does not include the material available on the web. This legal framework imposes significant limitations on the ability to gather and utilize existing Slovene texts for model training.

Addressing this issue involves two potential approaches. The first option is to advocate for legislative reform, urging lawmakers to amend the copyright laws to accommodate the specific needs of language technology development. Such a legal adjustment could facilitate broader access to textual resources while ensuring that intellectual property rights are respected in a manner compatible with technological advancements. However, this approach is inherently time-intensive and comes with no guarantee of success. It places the outcome largely outside of our control, as it depends on the willingness of policymakers to adopt the proposed changes and the eventual implementation of new legal frameworks.

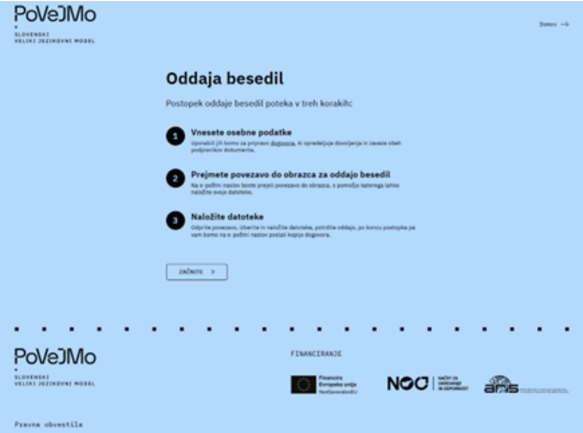
Given these uncertainties, we have determined that a more pragmatic and immediate approach is to actively seek permission from copyright holders to use their texts for the purpose of building a Slovene LLM. This strategy, while more labour-intensive, allows us to make direct progress without waiting for external factors to align. This approach requires substantial effort on our part, as it involves identifying relevant stakeholders, initiating discussions, addressing potential concerns, and negotiating agreements. A significant challenge in this process arises from the hesitation of some stakeholders, particularly media outlets and publishers, who are concerned about the uncertain societal implications of artificial intelligence and its potential impact on their work.

3.1 Engaging with large-scale providers

Our experiences with addressing large-scale text providers so far suggest that stakeholders often prefer to maintain the status quo, choosing to wait and observe who among them will take the first step. This creates a paradoxical situation: while applications such as ChatGPT, which rely on data from other languages, are already widely used also in the Slovene language, there is hesitation about building a Slovene-specific model or, more precisely, about allowing access to copyrighted text to build the model. The primary concern seems to be uncertainty about what will happen with the texts, specifically how and why the data will be used, whether there is potential for misuse, and what safeguards are in place to ensure that the texts are handled responsibly and ethically. A general conclusion could be drawn that stakeholders are willing to use an English-based model, which requires no contribution of their own texts, but they might be hesitant about contributing when it comes to developing a Slovene model.

In response, our communication strategy emphasizes the importance of preserving Slovene as a digital language. We argue that a Slovene language model is essential to ensure that the language remains comparable and competitive with other similar languages in the digital age. This initiative is framed as a collective effort, where every contribution helps to achieve a shared benefit—enhancing and improving resources for everyone. Additionally, we highlight the advantage of building this model independently, rather than solely relying on foreign corporations. By taking control of the process, we can ensure transparency in the types of texts included and maintain the ability to use the model for our specific needs.

In the meantime, the European Union has also recognized the importance of this issue and is actively working toward creating a publicly available large language model for all European languages. This initiative aims to provide equal opportunities for both commercial and non-commercial use



PoVeJMo
SLOVENSKI
VELIKI JEZIKOVNI MODEL

Oddaja besedil

Postopek oddaje besedil poteka v treh korakih:

1. **Vnesete osebne podatke**
   Uporabili jih bomo za pripravo dogovora, ki opredeljuje dovoljenja in zaveze obeh podpisnikov dokumenta.
2. **Prejmete povezavo do obrazca za oddajo besedil**
   Na e-poštni naslov boste prejeli povezavo do obrazca, s pomočjo katerega lahko naložite svoje datoteke.
3. **Naložite datoteke**
   Odprite povezavo, izberite in naložite datoteke, potrdite oddajo, po koncu postopka pa vam bomo na e-poštni naslov poslali kopijo dogovora.

PoVeJMo
SLOVENSKI
VELIKI JEZIKOVNI MODEL

FINANCIRANJE

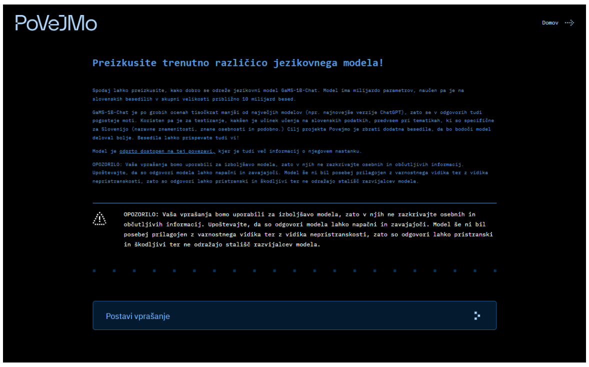
Figure 1: The section of the web portal where the interested participants can provide their texts (https://zbiranje.povejmo.si/).

across European languages. In this context, the amount of Slovene text collected directly impacts the positioning of the Slovene language within this broader European framework. Thus, one of our messages to the public is that building a Slovene language model is a matter of national interest. It ensures that Slovene can be effectively integrated into products and services developed by companies, benefiting both businesses and the wider community.

3.2 Engaging with individuals

To address individuals or groups that do not fall into the category of large-scale text providers and to encourage their contribution of textual resources for the development of the Slovene language model, we designed a targeted promotional strategy. As part of this effort, we participated in radio and television programs focused on language or technology-related topics, where we explained the concept of LLMs, the processes involved in their development, their potential applications, and the importance of individual engagement. In addition, we organized or participated in roundtable discussions and expert panels that facilitated debates on the advantages and concerns surrounding the creation of such a model.

To facilitate the widest possible text collection, we developed a web portal (https://povejmo.si/) where participants can submit their texts (Figure 1) while accessing essential information about the phases of developing a large generative language model, details of the text collection campaign, and answers to frequently asked questions. To ensure trust and encourage participation, we highlighted both the purpose and values of the text-collection project. The project's goals include supporting developmental independence by creating a Slovene-specific language model, ensuring a controlled and secure process for data handling, promoting open accessibility of the model, improving the machine understanding and generation of Slovene, overcoming language barriers, and capturing Slovenia's national specifics. The project is guided by three key values: openness, ensuring transparency, clear methodologies, and secure data handling; ethics, with a commitment to privacy, anonymity, and proper consent; and inclusiveness, fostering the participation of diverse groups to reflect the richness of the Slovene language.



PoVeJMo

Preizkusite trenutno različico jezikovnega modela!

Spodaj lahko preizkusite, kako dobro se odreže jezikovni model GaMS-1B-Chat. Model ima milijardo parametrov, naučen pa je na slovenskih besedilih v skupni velikosti približno 10 milijard besed.

GaMS-1B-Chat je po grobih ocenah tisočkrat manjši od največjih modelov (npr. najnovejše verzije ChatGPT), zato se v odgovorih tudi pogosteje moti. Koristen pa je za testiranje, kakšen je učinek učenja na slovenskih podatkih, predvsem pri tematikah, ki so specifične za Slovenijo (naravne znamenitosti, znane osebnosti in podobno.) Cilj projekta Povejmo je zbrati dodatna besedila, da bo bodoči model deloval bolje. Besedila lahko prispevate tudi vi!

Model je odprto dostopen na tej povezavi, kjer je tudi več informacij o njegovem nastanku.

OPOZORILO: Vaša vprašanja bomo uporabili za izboljšavo modela, zato v njih ne razkrivajte osebnih in občutljivih informacij. Upoštevajte, da so odgovori modela lahko napačni in zavajajoči. Model še ni bil posebej prilagojen z varnostnega vidika ter z vidika nepristranskosti, zato so odgovori lahko pristranski in škodljivi ter ne odražajo stališč razvijalcev modela.

⚠ OPOZORILO: Vaša vprašanja bomo uporabili za izboljšavo modela, zato v njih ne razkrivajte osebnih in občutljivih informacij. Upoštevajte, da so odgovori modela lahko napačni in zavajajoči. Model še ni bil posebej prilagojen z varnostnega vidika ter z vidika nepristranskosti, zato so odgovori lahko pristranski in škodljivi ter ne odražajo stališč razvijalcev modela.

Figure 2: The section of the web portal where the interested participants can test the existing model (https://povejmo.si/klepet/).

In Appendix A, we include the translated *Agreement for the use of copyrighted works in connection with text collection in the PoVeJMo project* to demonstrate the implemented legal solution.

The web portal also features an interactive section where users can try GaMS-1B-Chat, developed on the current version of the language model (Figure 2). As mentioned in Section 2, the current version still underperforms, however, this limitation also serves as a call for action: if we aim to develop a more accurate and robust model, a significantly larger dataset of Slovene texts is essential.

4 Data Storage and processing protocol

We have implemented a comprehensive protocol for securely managing the storage, encryption, transmission, and processing of data throughout all stages of its lifecycle. Data is received through network connections or on portable storage devices, which may contain unencrypted or optionally encrypted data using a temporary key set by the data owner. Once received, data undergoes decryption (if necessary) on a secure system, is re-encrypted with a new encryption key, and then securely uploaded to the primary storage system. Backup copies of the encrypted data are created and stored on a secondary system to ensure disaster resilience. During processing, unencrypted data is used exclusively for the duration of the analysis or transformation. Owners define specific rules for further sharing of their data, such as limiting its use to the project or permitting broader access to other researchers within Slovenia, the EU, or beyond.

Data is transferred between systems using encrypted channels (e.g., SSL/TLS, SSH). Data minimization principles are applied, providing only the necessary subsets of data for specific use cases, packaged as encrypted ”data bundles”. All data transfers are tracked to ensure accountability, with detailed records of which data was shared and with whom. Processing on high-performance computing systems (e.g., Vega supercomputer) or specialized research infrastructures (e.g., FRIDA at the Faculty of Computer and Information Science) adheres to strict security protocols. Data is stored in encrypted form and decrypted only temporarily during processing. After processing, unencrypted data is securely deleted. Future improvements may include leveraging Confidential Computing technologies for enhanced security.

Access management prioritizes security and simplicity, given the initial small group of users. Encryption keys are manually tracked and managed by the project lead, with backups maintained securely. Data bundles for specific use cases are encrypted with unique keys assigned to each user, employing AES symmetric encryption for efficiency and security. As the number of users or datasets grows, more advanced access management systems will be introduced.

5 Conclusion and Future Work

This paper has outlined the national initiative to overcome data scarcity and support the development of a Slovene large language model. By implementing diverse text-gathering scenarios, including a user participation portal and an extensive communication strategy, the project has actively addressed the unique challenges faced by languages with a small speaker base.

At the time of this paper’s submission, the foundational infrastructure for collecting Slovene-language texts has been fully established, including a user-friendly portal for contributions and robust protocols for data storage and processing. A nationwide promotional campaign was launched, aiming to mobilize broad participation from both institutions and individuals. By the time of the workshop, we anticipate being able to report on the first campaign’s outcomes, including the volume and diversity of collected texts and their suitability for training the Slovene large language model. These results will provide valuable insights into the scalability and replicability of such initiatives for other less-resourced languages.

Acknowledgments

The research program *Language Resources and Technologies for Slovene* (P6-0411) and the projects Adaptive Natural Language Processing with Large Language Models and Large Language Models for Digital Humanities (GC-0002) are funded by the Slovenian Research and Innovation Agency.

References

Firoj Alam, Shammur Absar Chowdhury, Sabri Boughorbel, and Maram Hasanain. 2024. Llms for low resource languages in multilingual, multimodal and dialectal settings. In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics: Tutorial Abstracts*, pages 27–33.

Abbas Raza Ali, Marcin Budka, and Bogdan Gabrys. 2019. Towards meta-learning of deep architectures for efficient domain adaptation. In *PRICAI 2019: Trends in Artificial Intelligence: 16th Pacific Rim International Conference on Artificial Intelligence, Cuvu, Yanuca Island, Fiji, August 26–30, 2019, Proceedings, Part II 16*, pages 66–79. Springer.

Marta Bañón, Malina Chichirau, Miquel Esplà-Gomis, Mikel L. Forcada, Aarón Galiano-Jiménez, Cristian García-Romero, Taja Kuzman, Nikola Ljubešić, Rik van Noord, Leopoldo Pla Sempere, Gema Ramírez-Sánchez, Peter Rupnik, Vít Suchomel, Antonio Toral, and Jaume Zaragoza-Bernabeu. 2023. http://hdl.handle.net/11356/1795 Slovene web corpus MaCoCu-sl 2.0. Slovenian language resource repository CLARIN.SI.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind

Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901.

Tomaž Erjavec. 2023. http://hdl.handle.net/11356/1775 Corpus of combined slovenian corpora metaFida 1.0. Slovenian language resource repository CLARIN.SI.

Jared Kaplan, Sam McCandlish, Tom Henighan, Tom B Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu, and Dario Amodei. 2020. Scaling laws for neural language models. *arXiv preprint arXiv:2001.08361*.

Iztok Kosem, Jaka Čibej, Kaja Dobrovoljc, Tomaž Erjavec, Nikola Ljubešić, Primož Ponikvar, Mihael Šinkec, and Simon Krek. 2023. http://hdl.handle.net/11356/1879 Monitor corpus of slovene trendi 2023-09. Slovenian language resource repository CLARIN.SI.

Viet Dac Lai, Nghia Trung Ngo, Amir Pouran Ben Veyseh, Hieu Man, Franck Dernoncourt, Trung Bui, and Thien Huu Nguyen. 2023. Chatgpt beyond english: Towards a comprehensive evaluation of large language models in multilingual learning. *arXiv preprint arXiv:2304.05613*.

Andy Liu, Mona Diab, and Daniel Fried. 2024. Evaluating large language model biases in persona-steered generation. *arXiv preprint arXiv:2405.20253*.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of machine learning research*, 21(140):1–67.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. 2023. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*.

Domen Vreš, Martin Božič, Aljaž Potočnik, Tomaž Martinčič, and Marko Robnik Šikonja. 2024. Generative model for less-resourced language with 1 billion parameters. pages 485–511.

Guillaume Wenzek, Marie-Anne Lachaux, Alexis Conneau, Vishrav Chaudhary, Francisco Guzmán, Armand Joulin, and Edouard Grave. 2019. Ccnet: Extracting high quality monolingual datasets from web crawl data. *arXiv preprint arXiv:1911.00359*.

Aleš Žagar, Matic Kavaš, Marko Robnik-Šikonja, Tomaž Erjavec, Darja Fišer, Nikola Ljubešić, Marko Ferme, Mladen Borovič, Borko Bošković, Milan Ojsteršek, and Goran Hrovat. 2022. http://hdl.handle.net/11356/1449 Abstracts from the KAS corpus KAS-abs 2.0. Slovenian language resource repository CLARIN.SI.

Appendix A

Copy of the Agreement for the use of copyrighted works in connection with text collection in the PoVeJMo project

Article 1 (Introductory Provisions)

1.1. The Contractor is a scientific research organization engaged in building open-access large language models for the Slovenian language, among other activities, within the framework of the program Adaptive Natural Language Processing with Large Language Models (PoVeJMo) (hereinafter: PoVeJMo program), which is implemented under the Public Call for Co-financing of Long-term Large Research and Innovation Collaborative Programs at TRL 3-6 within the Recovery and Resilience Plan (RRP).

1.2. With the aim of developing open-access large language models for the Slovenian language, the Contractor collects various types of texts. [NAME SURNAME] manages this process on behalf of and for the Contractor's account.

Article 2 (Subject of the License)

2.1. By this license, the Copyright Holder transfers to the Contractor all economic copyrights, related rights, and other rights of the author as defined by the Slovenian Copyright and Related Rights Act (hereinafter: ZASP), necessary for the development of open-access large language models for the Slovenian language, particularly the right of reproduction, the right to make available to the public, and the right of adaptation.

2.2. The transfer of rights to the Contractor is non-exclusive, royalty-free, indefinite in duration, and unlimited in territorial scope, allowing the Contractor to build open-access large language models for the Slovenian language, including the execution of the long-term PoVeJMo program.

2.3. The Copyright Holder agrees that the Contractor may freely transfer the granted rights to third parties for the purpose of developing open-access large language models for the Slovenian language.

2.4. The subject of this agreement pertains to the rights of the following copyrighted works:

- Copyrighted Work 1
- Copyrighted Work 2
- ...

Article 3 (Obligations of the Copyright Holder)

3.1. The obligations of the Copyright Holder include:

- Enabling the Contractor access to the copyrighted works in digital form via the online portal (https://povejmo.si/) or by another method agreed upon in an annex to this agreement.
- Collaborating with the Contractor to ensure the successful development of open-access large language models for the Slovenian language, particularly in the implementation of the PoVeJMo program.

3.2. The Copyright Holder guarantees that they hold all rights to the copyrighted works specified in Article 2.4, thereby enabling the Contractor to obtain all necessary permissions for their use in developing open-access large language models for the Slovenian language.

Article 4 (Obligations of the Contractor)

4.1. The obligations of the Contractor include:

- Using the copyrighted works specified in Article 2.4 of this agreement and any copies thereof created during the development of open-access large language models for the Slovenian language solely for that purpose and storing them in a secure environment that ensures an appropriate level of security, proportionate and limited to what is necessary for safe storage and prevention of unauthorized use.
- The works from Article 2.4 are exclusively used for the development of open-access large language models for the Slovenian language.

Article 5 (Data Protection)

5.1. The Contractor and the Copyright Holder agree to protect and process any personal data in accordance with the provisions of the Slovenian Personal Data Protection Act (Official Gazette of the Republic of Slovenia, No. 163/22, hereinafter: ZVOP-2) and Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, repealing Directive 95/46/EC (OJ L 119, 4.5.2016, hereinafter: General Data Protection Regulation).

Article 6 (Final Provisions)

6.1. The Contractor and the Copyright Holder agree that this agreement shall be governed by Slovenian law, and any matters not regulated by this agreement shall be governed by the provisions of ZASP and the Slovene Obligations Code (hereinafter: OZ).

6.2. This license is drawn up in two (2) identical copies, one (1) for each party. Any modifications or amendments to this agreement are possible only with mutual consent and in writing.

6.3. The Contractor and the Copyright Holder commit to resolving any disputes amicably. If a dispute cannot be resolved, the competent court at the Contractor's registered office shall have jurisdiction over dispute resolution.