

# When Artificial Intelligence meets Culture: The MESOC Serapeum

Dragan Čišić and Božidar Kovačić (University of Rijeka), Francesco Molinari (University of València)

## Highlights

- The MESOC Serapeum is a collection of Artificial Intelligence tools offered to researchers and policy analysts interested in exploring the societal impacts of culture.
- The first working version took advantage of a repository of 573 free and open-source academic papers dealing with the topic and especially describing relevant case studies.
- The repository is being expanded to include grey (policy) literature texts, with a special focus on European cities.
- The areas of societal impact are those identified by the EU Agenda for Culture as “crossover themes”: health and well-being, urban renovation and regeneration, social cohesion and public participation.
- The contribution of cultural activities is defined according to the 10 domains introduced by the UNESCO Framework of Cultural Statistics.
- The URL is: <http://mesoc-serapeum.eu/>.

## Background

Natural Language Processing (NLP) is the subfield of Artificial Intelligence (AI) that aims to process, analyse and ultimately understand natural language in written or spoken form.

Over the past few years, a major breakthrough in NLP has occurred with the introduction of the so-called transformers [Vaswani et al., 2017]. These are deep learning models – a class of machine learning algorithms inspired by the structure and function of the human brain – which have been proven to considerably outperform the more traditional NLP tools in a variety of key tasks, such as language translation, question answering, text summarization, semantic search and text classification.

Transformers use a novel mechanism, called attention, which enables them to capture long-range dependencies and contextual information at input level, making it possible to generate more coherent and meaningful outputs.

Several architectures developing the basic transformer model introduced in 2017 have been created and trained with textual data, which include, among others: BERT [Devlin et al., 2019], DistilBERT [Sanh et al., 2019], T5 [Raffel et al., 2019], GPT-2 [Radford et al., 2019], and the GPT-3 model [Brown et al., 2020] which is said to be superseded by GPT-4 during the first half of 2023.

## Rationale

The MESOC project consortium has created a repository of 573 free and open-source academic papers dealing with the societal impact of culture and especially describing relevant case studies. The repository is now being expanded to include grey (policy) literature texts, with a special focus on European cities.

However, it is never easy nor practical for a researcher or a policy analyst to handle a huge and increasing number of literature texts. Just to make a single example, the number of sentences in the documents currently stored in the MESOC repository exceeds 300,000.

In these conditions, it is quite difficult to decide what is relevant to read and consider as a guideline or at least as an inspiration for future actions. Not to mention the possibility of not finding sufficient time to analyse the texts in question and highlight their key points or to establish connections with related documents having similar contents.

## Solution

The MESOC Serapeum is a collection of Artificial Intelligence tools offered to researchers and policy analysts interested in exploring the societal impacts of culture.

The areas of societal impact are those identified by the EU Agenda for Culture as “crossover themes”: health and well-being, urban renovation and regeneration, social cohesion and public participation.

The contribution of cultural activities is defined according to the 10 domains introduced by the UNESCO Framework of Cultural Statistics.

The MESOC Serapeum is freely accessible from the following URL: <http://mesoc-serapeum.eu/>.

## Available functionalities

The main features of the MESOC Serapeum so far developed include the following:

- Semantic search
- Document review
- Clustering and topic analysis
- Transition variable search and view
- Societal impact analysis
- Thesaurus and taxonomy

## Semantic search

Semantic search is a technique that uses machine learning algorithms to understand the context, intent, and meaning behind a user's query, rather than relying solely on keyword matching. It could be defined as “search with meaning”, as it delivers more relevant and accurate results by analysing the relations between words and concepts and by understanding the user's intent based on search history and other contextual clues.

In the Serapeum, semantic search is implemented for all documents, abstracts and full texts.

## Document review

The AI system reviews existing documents to assess the societal impact of culture. This activity includes document clustering, text summarizing, keyword analysis, and other analytical tools and methods.

There is also a document mapping function, presenting the geographical area an article is about.

## Clustering and topic analysis

Clustering and topic analysis are two important techniques used in NLP for organizing and summarizing large collections of textual data.

Clustering is used to identify subgroups of similar documents, which can be helpful in information retrieval.

Topic analysis, on the other hand, is useful for identifying key themes and trends in a large corpus of texts.

As computers work with numbers, text has to be transformed into multidimensional numbers as vectors. We use here a 4096-dimensional space to transform words, sentences, and documents into vectors of features. Objects that are near each other should belong to the same cluster. Objects that are far from each other should belong to different clusters.

## Transition variable search and view

Transition variables show us the paths of transformation and the channels of materialization of impact, in a richer and more complex analysis than the cause-effect linearity. Transition processes are complex and non-linear, but they induce important changes across time. Transition variables enable a better contextualisation of the concrete change processes in specific places and times. They can be observed in shorter periods of time than the expected impacts.

The MESOC Serapeum extracts transition variables from texts as contextual elements, which can be measured, confirming that the cultural policy or practice under inspection is generating public value and/or affecting, at least to some extent, the target individuals or groups.

We first selected a transformer model, which we trained with few-shot learning methods on a small dataset using specific objectives and hyperparameters for transition variables extraction. Fine-tuning allowed the model to adapt to the specific nuances and patterns of the transition variable mining task. The results are presented as candidate transition variables.

## Societal impact analysis

Societal impact is the term used to describe the changes in e.g. the quality of life, health and wellbeing of people. The MESOC Serapeum generates impacts for every combination of the 3 “crossover themes” of the EU Agenda for Culture and the 10 cultural domains of the UNESCO Framework of Cultural Statistics.

To execute this task, DialogGPT (antecedent of ChatGPT) is used, a pre-trained language model developed by OpenAI that generates human-like text responses in a conversational context. DialogGPT is trained on a massive dataset of texts, allowing it to benefit from a diverse range of knowledge and understanding of different topics. It uses the latest advances in AI and deep learning to generate context-aware and coherent responses based on the inputted text.

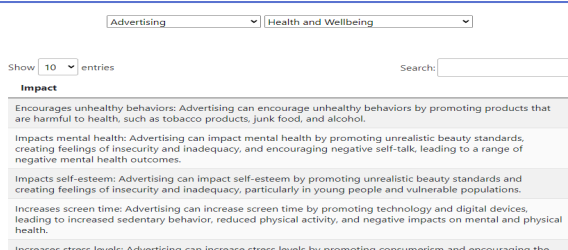


Figure 1 Impacts generated with DialogGPT

## Thesaurus and taxonomy

Through rich metadata and links, the MESOC Serapeum provides a powerful tool for further knowledge creation and research on the structural model of the Societal Dimension of Culture.

The MESOC thesaurus and taxonomy are accompanied by a knowledge graph, showing the impacts of arts and culture on wellbeing, health, urban renewal, and the economy.



Figure 2 Taxonomy graph with links defining impacts

## References

Brown, T. B., et al. (2020) Language Models are Few-Shot Learners. ArXiv, abs/2005.14165.

Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2019) BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. ArXiv, abs/1810.04805.

European Commission (2018) A new European Agenda for Culture. [https://ec.europa.eu/culture/sites/default/files/2020-08/swd-2018-167-new-european-agenda-for-culture\\_en.pdf](https://ec.europa.eu/culture/sites/default/files/2020-08/swd-2018-167-new-european-agenda-for-culture_en.pdf).

Radford, A., et. al. (2019) Language models are unsupervised multitask learners. OpenAI blog, 1(8), 9.

Raffel, C., et al. (2019) Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. ArXiv, abs/1910.10683.

Sanh, V., et al. (2019) DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. ArXiv, abs/1910.01108.

UNESCO Framework for Cultural Statistics (FCS) (2009). <http://uis.unesco.org/sites/default/files/documents/measuring-cultural-participation-2009-unesco-framework-for-cultural-statistics-handbook-2-2012-en.pdf>

Vaswani, A., et al. (2017) Attention is all you need. Advances in neural information processing systems, 30.

Gong, Y. et al. (2023) Cross-domain few-shot learning based on pseudo-Siamese neural network. Sci Rep 13, 1427. <https://doi.org/10.1038/s41598-023-28588-y>

## Acknowledgments and disclaimer

This research has received funding from the European Union’s (EU’s) Horizon 2020 Research and Innovation Programme under Grant Agreement No. 870935. However, the opinions expressed here are solely of the authors and do not engage any of the EU granting institutions.

## Technical overview

The MESOC Serapeum is based on transformers, and for different tasks different systems are used, namely BERT, ROBERTA, DISTILBERT, XLNET, T5, GPT2, GPT3 and Chat GPT. Some of the transformers have been pretrained and then finetuned for the tasks used (summarization, question and answers, keywords extraction), while for the categorization and extraction of transition variables, sentence transformers have been trained on the texts from the MESOC database. Semantic search is created using Siamese neural networks (Gong et al., 2023).