Statistics, Technology and Data Science for Economic and Social Development

Book of short papers of the ASA Rome Conference
18 to 20 September 2024

"Measuring and Interpreting World Changes with Statistics, Data Science and AI"

Book of Short Papers of the
Sapienza University of Rome Conference

Supplement to Volume 37/2

Guest Editors:
Leonardo S. Alaimo, Fabio Crescenzi,
Luigi Fabbris, Filomena Maggino, Maurizio Vichi

A.S.A CONTACTS
Principal Contact
Francesco Palumbo (Editor in Chief)
editor@sa-ijas.org

Support Contact
Domenico Vistocco (Editorial Manager)
ijas@sa-ijas.org

JOURNAL WEBPAGE
https://www.sa-ijas.org/ojs/index.php/sa-ijas

Statistica Applicata – Italian Journal of Applied Statistics is associated to the following Italian and international journals:

QTQM – Quality Technology & Quantitative Management (http/web.it.nctu.ew/~qtqm/)

SINERGIE – Italian Journal of Management

Statistica Applicata – Italian Journal of Applied Statistics (ISSN:1125-1964, E-ISSN:2038-5587) applies the Creative Commons Attribution (CC BY) license to everything we publish.

Pubblished: April 2025

Published by Firenze University Press
Powered by Firenze University Press

Firenze University Press
Università degli Studi di Firenze
via Cittadella, 7, 50144 Firenze, Italy
www.fupress.com

Statistica Applicata – Italian Journal of Applied Statistics is a four-monthly journal published by the Associazione per la Statistica Applicata (A.S.A.), Largo Gemelli 1 – 20123 Milano, Italy (phone + 39 02 72342904). Advertising:Cleup sc, via G. Belzoni, 118/3 – 35128 Padova, Italy (phone +39 049 8753496 – Fax +39 049 9865390), email: info@cleup.it.
Rules for manuscript submission: https://www.sa-ijas.org/ojs/index.php/sa-ijas/about/submissions
Subscription: yearly €103.30; single copy €40.00; A.S.A. associates €60.00; supporting institutions: €350.00. Advertisement lower than 70%. Postal subscription Group IV, Milan. Forum licence n. 782/89. CLEUP SC on behalf of ASA, 7 March 2023.

doi.org/10.26398/asaproc.0082

Artificial intelligence and future scenarios: Preliminary results from an application on family policies in north-eastern Italy

Mario Bolzan[a], Yuri Calleo[b], Simone Di Zio[c], Marco Marozzi[d], Manuela Scioni[a]
[a] Department of Statistical Sciences, University of Padova, Padova, Italy.
[b] Spatial Dynamics Lab, University College Dublin, Ireland.
[c] Department of Legal and Social Sciences, University "G. d'Annunzio" of Chieti-Pescara, Italy
[d] Department of Mathematics and Informatics, University of Ferrara, Italy

1. Introduction

In an era of great uncertainty, many public authorities are prioritizing the development of policies focused on family support. These initiatives aim to assess strategies that enhance family well-being and mitigate potential challenges. By investing in family-friendly policies, such as affordable childcare, parental leave, and financial support, national and regional entities aim to promote positive outcomes and reduce risks associated with economic and social pressures. These efforts are important in creating resilient family structures and fostering a positive environment for future generations. In this paper, we aim to contribute to research on *anticipatory governance*, a well-established approach that focuses on imagining, exploring, and preparing for future challenges and changes before they fully emerge (Fuerth, 2009). Unlike traditional reactive governance – which addresses problems after they arise – anticipatory government seeks to anticipate potential issues and implement proactive measures to address them. Different key aspects characterize this approach, including foresight and scenario planning (Calleo and Pilla, 2023), adaptive policies, innovation and technology, collaboration and engagement, and resilience building. We are currently in what is defined as the "age of data", where the proliferation of big data and advancements in artificial intelligence (AI) make data crucial for developing anticipatory solutions (Di Zio et al. 2023). These technologies offer a deeper, more detailed understanding of potential future developments with increased precision and speed. This integration of advanced data analytics into governance processes not only improves the accuracy of actions but also supports more strategic and adaptive policymaking, ultimately contributing to more resilient and forward-thinking governance. In addition, there is considerable interest in exploring the "scenarios" that institutions, perhaps unconsciously, through their promotion of policies, envision for the future, and compare them with the viewpoints of experts.

These considerations are particularly relevant in the realm of family policies, a constantly evolving and complex area exposed to rapid changes in social dynamics, economic conditions, and demographic trends (Bolzan, 2018). Anticipatory governance plays a crucial role in shaping effective family policies by enabling policymakers to proactively address emerging challenges and opportunities. As family dynamics and structures evolve, it is essential for policies to be designed with an eye toward future trends and plausible scenarios rather than simply reacting to current issues. For example, shifts in demographic patterns, such as declining birth rates or an ageing population, necessitate policies that can adapt to these changes before they create significant challenges. Economic factors also illustrate the importance of anticipatory governance. Economic downturns or shifts in the job market can have profound effects on family stability and well-being. By anticipating these economic trends, policymakers can develop measures to support families during periods of financial instability, such as by enhancing unemployment benefits or creating job training programs. Technological advancements present another area where anticipatory

governance is vital. As technology continues to evolve, it affects family life in various ways, from changing work environments to altering communication patterns and parenting practices. Policymakers who anticipate these changes can better prepare family support systems to remain relevant and effective. Social changes are also a key consideration. Evolving attitudes toward issues such as gender roles, same-sex marriage, and parental leave require policies that are not only current but also forward-thinking. In essence, anticipatory governance helps ensure that family policies reflect and support these changing social norms, promoting inclusivity and equity.

This paper examines the legal initiatives undertaken and implemented by the administrations of the four regions of the Northeast of Italy: Friuli Venezia Giulia, Trentino-Alto Adige, Veneto, and Emilia Romagna during the four years from 2018 to 2022, to extract from them the advised scenarios for the family's future and compare them with those obtained from the experts.

Nevertheless, analyzing multiple documents, whether manually or through text-mining techniques, may provide only a partial achievement of the goal due to the lack of a clear understanding of the issue and the difficulty of extraction. To overcome these challenges, the research objectives (RO) of this paper are the following:

- RO_1: Identify and extract regional legal initiatives: investigate and extract legal initiatives from selected regions in Northeastern Italy and analyze these initiatives using artificial intelligence models.
- RO_2: Analyse regional family issues: employ Generative Pre-Trained Transformers to identify and outline common issues and needs experienced by families within the region.
- RO_3: Develop policy recommendations: utilize Generative Pre-Trained Transformers to generate a list of potential policy recommendations tailored to each region, which can be presented to policymakers for further discussion and refinement.

The rest of the paper is organized as follows: Section 2 details the methodology adopted, Section 3 presents and discusses the results, and Section 4 concludes.

2. Materials and methods

The methodology adopted for this paper started from the extraction of legal initiatives from the official websites of the four regions analyzed: Friuli Venezia Giulia, Trentino-Alto Adige, Veneto, and Emilia-Romagna in the four years (2018-2022). The extraction is performed using either each site's internal search engine or the advanced features of the Google search engine. All documents related to regional legal initiatives containing the word "family" or similar terms were extracted. The documents thus obtained were analyzed one by one to eliminate any duplicate or irrelevant documents. The extraction allows us to obtain a data matrix containing all the relevant information to import into the model, including the region of reference, type of document, associated link, associated administrative number, date, title, and text. To facilitate semi-automated analysis of the matrix, we employed ChatGPT-4, an advanced generative AI model. Developed by OpenAI, ChatGPT-4 (www.chatgpt.com) is an AI language model engineered to comprehend and generate text that mimics human language. This model utilizes a transformer-based machine learning architecture, which enables it to effectively process and produce natural language outputs based on the provided input. The matrix is subsequently converted to Portable Document Format (PDF) to facilitate uploading. Four primary prompts were then utilized within the document for a deeper analysis. The selection of prompts is inherently subjective and has a substantial impact on the final outcomes. Nonetheless, we opted to use general prompts with straightforward formulations, defined as follows:

- **P1**: Act like a policymaker. Considering these official documents, provide a list of policies that support families in the region [*insert region here*], Italy.
- **P1.1**: Considering this list of family support policies of the region [*insert region here*],

Italy, eliminates those that are not relevant.

- **P2**: Act like a family member. Considering these official documents, what issues and needs emerge for families in the [*insert region here*] region in the text?
- **P2.1**: Considering this list of emerging needs of the region [*insert region here*], Italy, eliminates those that are not relevant.

From a statistical point of view, when we import the document into the model, asking a specific question, the initial step involves analyzing the content of the document employing natural language processing (NLP) models to understand the document's content, including 1) tokenization; 2) entity recognition; and 3) structural analysis. In our case, since we have a document D for each region, consisting of words $w_1, w_2, \dots, w_n$, the process involves breaking D into tokens $t_1, t_2, \dots, t_m$, where each token t_i, can be a word or a sub-word. Once this process is concluded, the entity recognition involves identifying entities E from the tokens adopting a named entity recognition (NER) (Wang et al. 2023), thus allowing for the extraction of relevant information crucial for answering the specific prompt considering the documents. In this case, if t_i corresponds to a location, E_i might be tagged as a location entity, using a conditional random field (CRF) model or transformers. Finally, the structure of the text is analysed adopting a probabilistic parsing, where given a sequence of tokens $(t_1, t_2, \dots, t_m)$, a parse tree P is generated to represent syntactic structures. This can be denoted as $P = argmax_p \Pr(P|t_1, t_2, \dots, t_m)$, where $\Pr(P|t_1, t_2, \dots, t_m)$ represents the probability of parse tree P given the token sequence.

Specifically for our questions, with regards to P1 and P2, the model looks for relevant sections in D. Let Q be the query (the prompt selected), and S be the section of D. The relevance score $R(Q,S)$ is computed, adopting a cosine similarity (Ethayarajh, 2019) in the embedding space defined as:

$$R(Q,2) = \frac{\boldsymbol{E}_Q \cdot \boldsymbol{E}_S}{||\boldsymbol{E}_Q||\,||\boldsymbol{E}_S||}, \tag{1}$$

where $\boldsymbol{E}_Q$ and $\boldsymbol{E}_S$ are the embedding of the query and the section, respectively. For each relevant section S, policies P_i are extracted based on key phrases and entities. If S contains phrases associated with policies, the model uses learned patterns to identify these, utilising a classification approach defined as:

$$\hat{P} = argmax_p \Pr(P_i|S). \tag{2}$$

On the other hand, with regards to P2, if the model identifies family issues by looking for specific patterns or entities related to problems within D, this is denoted as $I = \{i_j \mid \Pr(i_j \mid t_1, t_2, \dots, t_m > \theta\}$. Where i_j are issues and θ is a threshold for filtering. Finally, to summarise the issues and needs, we adopt an extractive summarisation technique where the score for each sentence s in D is calculated as: $S(s) = f(s, I)$, where f is a function that scores the relevance of the sentence s to the issues I. Since, within the extracted policies some of them were on other topics (e.g., related to climate change), we decided to use an exclusion criterion by providing the GPT with prompt 1.1 and 2.1. In this case, once we have a list of policies ($L = \{p_1, p_2, \dots, p_n\}$) generated from the initial analysis, the model automatically uses proper keywords and key phrases to identify the main terms or phrases that describe the focus of each policy and the relevance criteria and contextual embeddings to represent policies and relevance criteria in a high-dimensional space (Ha et al. 2023). At the end of the process, we obtain a list L of policies and issues emerged from the four regions in short time frames ready for evaluation.

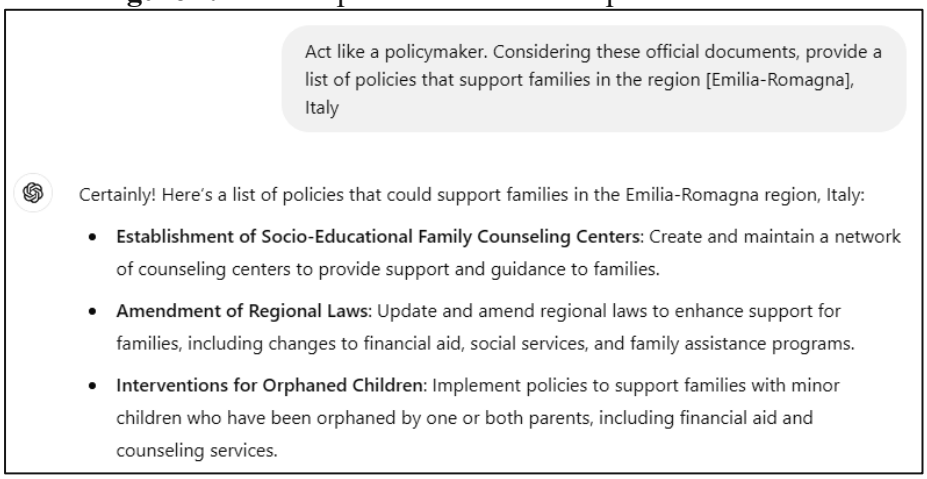
3. Results and discussion

The GPT-4 model was utilized to efficiently analyze administrative documents extracted from

the four regions. This approach enabled the extraction of a comprehensive list of intervention policies and provided a deeper understanding of the emerging needs highlighted within the documents. These results represent a dual advancement in the scientific literature: 1) First, they demonstrate the ability to rapidly extract information from a corpus of documents – within a matter of seconds – without requiring interpretative efforts from the analysts, as is often necessary with techniques like topic modelling. 2) Second, the approach enables the identification of actionable insights, in this case, formulating response policies for families in various regions based on the extracted information. Specifically, the model generated $L = 50$ preliminary policies for each region using the P1 prompt. Nevertheless, one limitation was the inclusion of policies unrelated to the family theme, as the model, after identifying a certain number of specific reference policies, continued to generate additional policies that were similar but not directly relevant. This issue arose from the model's tendency to extend the search beyond the essential reference policies. To address this issue, we employed the refined prompt 1.1, which instructed the model to eliminate non-relevant policies. As a result, the model produced 16 relevant policies for Emilia Romagna, 41 for Friuli Venezia Giulia, 47 for Trentino Alto Adige, and 47 for the Veneto region.

Considering the P2 prompt, the model efficiently identified the needs of families as expressed in the analyzed texts, extracting a total of 20 distinct needs across the four regions. In this instance, the results appeared consistent with the family theme and the prompt used. However, to ensure accuracy, we applied the refined prompt P2.1, allowing the model to further assess the relevance of the identified needs. The refinement process confirmed the relevance of all 20 needs extracted for each region, with no irrelevant needs identified. An example of the model's output is illustrated in Figure 1.

Figure 1. An example of the model's output.



The responses generated by the AI application are inherently influenced by the diverse formal languages employed by regional administrations, which reflect varying cultural perspectives and sensitivities regarding family structures and the social and economic challenges unique to each territory. Across all regions, economic assistance to families is prominently featured, particularly targeting those facing significant hardships such as large families, single parents, and individuals with disabilities. Additionally, there is a consistent emphasis on enhancing and expanding access to social and health services. However, notable differences between the four regions warrant further investigation. Below are specific elements observed in each region that were not explicitly present in the others. The Friuli Venezia Giulia region adopts a multifaceted approach to addressing several key social issues. Firstly, it places significant emphasis on combating school dropout rates and promoting academic success. This includes targeted programs designed to prevent bullying and address educational poverty, thereby supporting students who are at risk of falling behind. In addition to educational initiatives, the region is active in developing and managing social tourism activities. These programs aim to offer affordable travel opportunities for families, enhancing their

quality of life and providing valuable recreational experiences. Moreover, the Friuli Venezia Giulia region implements comprehensive strategies to tackle domestic violence. This includes providing support for victims and their families and ensuring that those affected by such violence receive the necessary help and resources to rebuild their lives.

In the Veneto region, several distinctive programs address specific family and social issues. The region places a strong emphasis on substance abuse prevention through targeted interventions aimed at reducing addiction within families and supporting affected individuals. Additionally, Veneto recognizes the need for legal assistance for families facing issues such as domestic violence, child custody disputes, or immigration concerns, and provides relevant support services. The region also prioritizes digital inclusion by ensuring that all families have access to technology and digital literacy training, which is essential for full participation in the modern digital environment. Furthermore, Veneto promotes inter-generational activities designed to strengthen relationships between different age groups, fostering greater cohesion within families and communities.

In the Trentino Alto Adige region, the approach to supporting families includes a focus on substance abuse and addiction services. The region offers a range of support and treatment options designed to assist families affected by these issues, aiming to mitigate the impact of substance abuse and provide pathways to recovery. Moreover, Trentino Alto Adige invests in cultural and recreational programs that enhance family life and foster community cohesion. These initiatives are intended to enrich the lives of families through diverse cultural experiences and recreational activities, strengthening social bonds and promoting a vibrant community environment.

In the Emilia Romagna region, several critical issues impact families in the region. One prominent challenge is digital inclusion; many families face barriers to accessing high-speed internet and digital devices, which limits their opportunities for online education, remote work, and participation in social activities. Environmental health also poses a significant concern, as poor environmental conditions, including pollution and inadequate waste management, adversely affect the well-being and quality of life of families in the region. Furthermore, the region struggles with underdeveloped transportation infrastructure. The public transportation network's limitations make it difficult for families to commute efficiently for work, school, and other essential activities.

These findings highlight areas where the AI approach is effective while also underscoring the need for continued experimentation with new methodologies and analytical tools to address these complex issues. It is important to acknowledge the role of regional socio-economic and cultural factors in shaping the outcomes of family policies. The divergence in policy effectiveness across regions such as Friuli Venezia Giulia, Trentino-Alto Adige, Veneto, and Emilia Romagna suggests that local contexts significantly influence how policies are perceived and implemented. For instance, the success of educational and social tourism initiatives in Friuli Venezia Giulia can be attributed to the region's strong emphasis on social cohesion and community-based interventions. Similarly, Veneto's focus on legal assistance and digital inclusion highlights the region's proactive stance in addressing modern societal challenges. However, these differences also underscore the limitations of a one-size-fits-all approach in policy development. The effectiveness of family policies is not solely dependent on the policies themselves but also on the underlying socio-economic fabric and cultural attitudes within each region. This nuanced understanding of regional dynamics should be a focal point for future policy formulation, ensuring that strategies are tailored to the specific needs and contexts of each area.

4. Conclusions

Anticipatory governance, as explored in this study, is a multifaceted approach that integrates statistical analysis, sociological insights, and artificial intelligence (AI) to proactively address future challenges in family policies. This paper investigated the application of anticipatory governance within the context of family policies in Northeastern Italy, considering advanced AI models, specifically GPT-4, to facilitate the extraction, analysis, and proposal of relevant policies across

four distinct regions. The research underscored the efficacy of AI-driven natural language processing techniques in rapidly analyzing extensive collections of administrative documents. This approach allowed for the swift identification of both current policies and emerging needs, demonstrating that tasks which traditionally require extensive human effort can now be executed in a fraction of the time and with negligible financial cost. The findings from this study are significant in several respects. First, the application of GPT-4 revealed its capability to generate comprehensive lists of preliminary policies, which were subsequently refined to a more targeted selection relevant to each region. This capability not only streamlined the policy analysis process but also enhanced the precision of policy identification, thereby facilitating more informed decision-making in public administration. Moreover, the study introduced a novel integration of AI tools within the realm of policy analysis, marking a departure from conventional methodologies that rely heavily on manual effort. This innovation has the potential to revolutionize the field by enabling more dynamic and responsive governance models that can anticipate and adapt to emerging societal needs more effectively.

For future works, an important direction involves further refinement of the AI models to improve their ability to contextualize and interpret the nuances of policy documents across different regions and policy areas. Additionally, expanding the application of these models to include real-time analysis of ongoing policy discussions could enhance the ability of policymakers to respond proactively to changing circumstances. From a computational perspective, future work could focus on enhancing the robustness and scalability of the AI algorithms used in policy analysis. This could involve the development of more sophisticated techniques for handling large datasets, improving the interpretability of AI-generated outputs, and ensuring that the models are able to generalize effectively across different contexts. Additionally, integrating these AI tools with other data sources, such as demographic and economic data, could provide a more comprehensive basis for policy formulation and evaluation. In conclusion, this study highlights the transformative potential of AI in the domain of policy analysis and anticipatory governance. By continuing to refine and expand these tools, we can progress towards a more agile and responsive model of governance, better equipped to address the complex challenges of the future.

References

Bolzan, M. (2018). *Domani in famiglia*. Franco Angeli.

Calleo, Y., Pilla, F. (2023). Delphi-based future scenarios: A bibliometric analysis of climate change case studies. *Futures*, **149**, 103143.

Di Zio, S., Calleo, Y., Bolzan, M. (2023). Delphi-based visual scenarios: An innovative use of generative adversarial networks. *Futures*, **154**, 103280.

Ethayarajh, K. (2019). How contextual are contextualized word representations? Comparing the geometry of BERT, ELMo, and GPT-2 embeddings. *arXiv preprint arXiv*:1909.00512.

Fuerth, L. S. (2009). Foresight and anticipatory governance. *Foresight*, **11**(4), pp. 14-32.

Ha, J. H., Lee, D. H., Choi, B. J. (2023, November). Text mining based GPT method for analyzing research trends. In *International Conference on Intelligent Human Computer Interaction*. Cham: Springer Nature Switzerland, pp. 26-31.

Wang, S., Sun, X., Li, X., Ouyang, R., Wu, F., Zhang, T., ... Wang, G. (2023). Gpt-ner: Named entity recognition via large language models. *arXiv preprint arXiv*:2304.10428.