BAR Brazilian Administration Review

ANPAD

Research Article

Assessment Method for Generative AI Technology in Foresight and Policy Design in Public Management: Expanding AI Trustability for Anticipatory Governance

Mateus Panizzon[1, 2], Raquel Janissek-Muniz[2], Natália Marroni Borges[2], Amanda Cainelli[2]

[1] Universidade Caxias do Sul, Business School, Caxias do Sul, RS, Brazil
[2] Universidade Federal do Rio Grande do Sul, Escola de Administração, Porto Alegre, RS, Brazil

How to cite: Panizzon, M., Janissek-Muniz, R., Borges, N. M., & Cainelli, A. (2025). Assessment method for generative AI technology in foresight and policy design in public management: Expanding AI trustability for anticipatory governance. *BAR-Brazilian Administration Review, 22*(3), e240196

DOI: https://doi.org/10.1590/1807-7692bar2025240196

Keywords:
anticipatory governance; artificial intelligence; foresight; policy design; public management

JEL Code:
O38, H83, O31, C63, D83

ABSTRACT

Objective: to develop and evaluate a generative AI system prototype and assessment method that supports anticipatory governance by integrating foresight and policy design, enabling stakeholders to anticipate and proactively address emerging challenges in public policy. **Methods:** the study uses a design science research approach, combining institutional and explainable AI frameworks. It designs and assesses a generative AI prototype through three case scenarios focusing on environmental, electoral, and labor regulations, and expands results to an assessment protocol. **Results:** the analysis demonstrates the strengths and limitations of generative AI in AG systems. The study produces a systemic framework and an assessment protocol for evaluating AI's role in augmenting AG capabilities, focusing on enhancing trust and reliability. **Conclusions:** the article's main contribution is the proposed assessment protocol that contributes to both theory and practice by providing a replicable method for enhancing trustability in AI-driven AG. The findings support researchers and policymakers in reflecting on and utilizing responsible AI to navigate complex geopolitical, environmental, and societal challenges.

Received:
November 19, 2024.
This paper was with the authors for two revisions.

Accepted:
March 28, 2025.

Publication date:
August 18, 2025.

Corresponding author:
Mateus Panizzon
Universidade Caxias do Sul, Business School
Rua Francisco Getúlio Vargas, n. 1130, Petrópolis,
CEP 95070-560, Caxias do Sul, RS, Brazil

Universidade Federal do Rio Grande do Sul,
Escola de Administração
Rua Washington Luiz, n. 855, Centro Histórico,
CEP 90010-460, Porto Alegre, RS, Brazil

Funding:
The authors thank to Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Brazil) to the financial support. Process number: 409898/2023-6

Conflict of Interests:
The authors stated that there was no conflict of interest.

Editors-in-Chief:
Ricardo Limongi
(Universidade Federal de Goiás, Brazil)

Associate Editor:
Sang-Bing Tsai
(WUYI University, China)

Reviewers:
Two anonymous reviewers.

Peer Review Report:
The disclosure of the Peer Review Report was not authorized by its reviewers.

Editorial assistants:
Eduarda Anastacio and Simone Rafael (ANPAD, Maringá, Brazil).

Data Availability: BAR – Brazilian Administration Review encourages data sharing but, in compliance with ethical principles, it does not demand the disclosure of any means of identifying research subjects.

Plagiarism Check: BAR maintains the practice of submitting all documents received to the plagiarism check, using specific tools, e.g.: iThenticate.

Peer review: is responsible for acknowledging an article's potential contribution to the frontiers of scholarly knowledge on business or public administration. The authors are the ultimate responsible for the consistency of the theoretical references, the accurate report of empirical data, the personal perspectives, and the use of copyrighted material. This content was evaluated using the double-blind peer review process. The disclosure of the reviewers' information on the first page is made only after concluding the evaluation process, and with the voluntary consent of the respective reviewers.

INTRODUCTION

With the advancement of AI technologies and their impact on several levels of society, the need to monitor emerging AI developments to understand scenarios and provide more responsive policy regulations has become critical for public policy and public management, for both opportunities for seizing and risk mitigation. In this sense, the function of anticipatory governance (AG), combining foresight, roadmapping and public policy design, integrating stakeholders' vision, becomes critical to support governments at several levels to enable sustainable transformations and address global challenges (Guston, 2014; Maffei et al., 2020; Muiderman et al., 2022; Panizzon & Janissek-Muniz, 2025). However, there is a gap in the literature about the role of generative AI on anticipatory governance, in terms of automation and augmentation of tasks and capabilities, and how AI agents can be designed taking responsible AI into consideration. Therefore, this article situates the discussion in the role of generative AI for AG, and the role of xAI for AI.

This research was developed within the observatory of AI technologies for public management, which aims to create initiatives to leverage anticipatory governance capabilities for public management in Brazil, integrating further understanding of AI technologies' impact from the external and internal points of view. From an external point of view, we consider the external monitoring of emerging AI research and technologies to investigate their impact level and implications for public policy. From an internal point of view, we mean understanding the adoption of AI technologies already in place in public management (Van Noordt & Misuraca, 2022). This integrated view of the observatory is an essential perspective for measuring AI impacts from the system's external and internal points of view. In this article, we focus on the external perspective, more specifically, on a generative AI (GEN_AI) assistant based on a large language model (LLM) designed to explore the impacts of given AI technologies on public policy design and redesign at both levels of emerging science and technology and innovations by highly funded startups. The introduction of AI technologies in the foresight process has been increasingly studied in recent years (Farrow, 2020; Bevolo & Amati, 2020; Geurts et al., 2022). However, adopting AI requires analyzing how much this agent can support humans or stakeholders in dimensions such as learning, automation, augmentation, and innovation.

By learning, we mean how the AI agent can support policymakers to gather more knowledge about a given subject, accelerating learning (Srinivasan, 2022). By automation, we mean how the AI agent can substitute specific manual human tasks in foresight that have less value to its process, like data collection and cleaning (Tyson & Zysman, 2022). By augmentation, we mean how the AI agent can increase human capabilities for different stakeholders to deal with a complex analysis, including the impacts of AI technologies, global challenges, and redefinition of policy regulations (Hassani et al., 2020). By innovation, we mean how new tasks and roles can emerge due to this technology, due to technology entanglement and job reconfiguration (Verma & Singh, 2022), reshaping functions for foresight analysts to new levels. To understand how this AI agent can serve as a more autonomous tool or become a support for the collective human intelligence process, we adopted an approach to analyze the AI agent in three scenarios. Therefore, the research question that drives this study is: *How is the perception of foresight and AI experts over a generative AI assistant for anticipatory governance, and can an assessment protocol emerge from this evaluation?*. A better understanding of this perception helps advance the prototype's conception for further testing with non-expert users with a structured method.

Therefore, this study aims to analyze the trustability perception of foresight, AI, and policy design experts over a generative AI assistant for anticipatory governance in three case scenarios. Based on the perception of AI trustability, we were able to abstract an assessment method for generative AI to anticipatory governance for further replication. In this context, we mean how the AI agent can automate by substituting specific manual human tasks in foresight that have less value in its process, like data collection and cleaning (Tyson & Zysman, 2022). By augmentation, we mean how the AI agent can increase human capabilities for different stakeholders to deal with a complex analysis, which are the impacts of AI technologies, global challenges, and redefinition of policy regulations (Hassani et al., 2020). And by innovation, we mean how new tasks and roles can emerge due to this technology, due to technology entanglement and job reconfiguration (Verma & Singh, 2022), reshaping functions for foresight analysts to new levels. To understand how this AI agent can serve as a more autonomous tool or become a support for the collective human intelligence process, we adopted an approach to analyze the AI agent in three scenarios. For the sake of this study, we analyzed, through a GEN_AI assistant, three cases: electoral laws, environmental laws, and labor laws, and how AI technologies impact these regulations. We selected these scenarios given the wide-open themes of global challenges that involve multiple stakeholders, such as democracy and AI (Coeckelbergh, 2023), climate change

and AI (Coeckelbergh & Sætra, 2023), and impacts of AI on professions (Boobier, 2018; Georgieff & Hyee, 2022; Tolan et al., 2020), that are critical for anticipatory governance (Puaschunder, 2019). These themes are highly related to impacts on geopolitical tensions, climate challenges, and social inequalities.

The vital issue of analyzing a GEN_AI assistant's performance for AG is that GEN_AI does have advantages; however, it currently faces technical limitations such as model collapse or model autophagy disorder (Dohmatob et al., 2024; Shumailov et al., 2023), AI hallucinations (Athaluri et al., 2023; Emsley, 2023; Salvagno et al., 2023), and AI bias (Mikalef et al., 2022; Suresh & Guttag, 2021), requiring more AI and foresight literacy to be better adopted by users (Jokinen et al., 2023), especially for collective processes. Therefore, this research seeks to understand the potential of the self-adoption of this AI in different levels of assistance (user autonomy without experts, user with experts, support to experts). This approach is particularly important from an institutional lens, especially considering both foresight and AI adoption in public management.

THEORETICAL REFERENCE

Anticipatory governance and foresight

According to Guston (2014), anticipatory governance is "a broad-based capacity extended through society that can act on a variety of inputs to manage emerging knowledge-based technologies while such management is still possible" (p. 1) — to address this challenge, AG motivates activities designed to build capacities in foresight, engagement, and policy design, integrating knowledge from scientists, engineers, policymakers, and other publics. Since then, studies such as Maffei et al. (2020), Ohta (2020), and Heo and Seo (2021) expanded AG from technologies to other issues, such as climate change and societal challenges, seeking to understand AG in Austria, Japan, UK, Netherlands, Finland, and Korea for policy design with foresight support. However, it is essential to notice that, in 2009, Fuerth (2009) already wrote about the intersections between foresight and AG, mentioning the 'forward engagement' as a "particular approach to anticipatory governance, drawing upon complexity theory for assessment of issues requiring government policy, including network theory for proposed reforms to legacy systems of governance to enable them to manage complexity under conditions of accelerating change; and cybernetic theory to propose feedback systems to allow ongoing measurement of the performance of policies against expectations" (p. 1). In 2012, the author published a new study (Fuerth, 2012) describing concepts developed between 2001 and 2011 and refined during a series of workshops held at the National Defense University. In essence, his concept for AG encompasses anticipatory governance, a systems-based approach that enables governance to cope with accelerating, complex forms of change. Anticipatory governance is a 'system of systems' comprising a disciplined foresight-policy linkage, networked management and budgeting to mission, and feedback systems to monitor and adjust. Both concepts address common elements such as the need for foresight and anticipation or proactivity for policy design; a systems-based approach, comprehending the integration of several components from foresight to policymaking; adaptability and flexibility to be able to respond to new technologies' impacts; engagement and interaction, built upon collective intelligence; and responsiveness to feedback among the process.

Although these elements usually sound like a rule of thumb, for Störmer et al. (2020), there is a presentism and short-term bias in policymaking which impacts policy design, focusing on current problems, that affects agenda setting with limited participation of actors, policymaking without the support of foresight and long-term view, which involves budgeting, implementation, and evaluation with a linear and not interactive process. Foresight, on the other hand, is more concerned with future and emerging issues, gathering weak signals, in which collective interpretation requires joint efforts from academia (in which new basic research drives disruptive technologies), from industry (in which new research can become impactful), and from government/society (in which regulations can be anticipated). Both models, short-term and long-term, have their functions. However, when it comes to understanding technological impact, short-term does not provide sufficient performance. This has led to regulations attempting to catch up with innovations, becoming reactive rather than anticipative. This creates hidden costs, dysfunctions among several levels of society, and uncertainty due to the lack of a legal framework.

One of the key examples of a good application of AG can be observed in Kolliarakis and Hermann (2020), on the *Towards European Anticipatory Governance for Artificial Intelligence*, where a series of workshops created critical interactions regarding the emergence of new AI technologies and future scenarios (from the perspective of researchers), the capabilities to implement it on an industry level (from the perspective of big techs), and the need to anticipate regulations (from the standpoint of policymakers). In 2024, the EU Artificial Intelligence Act was launched (Laux et al., 2024) and became the first regulation for AI at this level. Due to AI cognitive capabilities, this legal framework reflects important drivers for industry and society to address.

Therefore, it is notable that AG can reshape perceptions and learning by integrating science and technology producers with regulatory responsibilities and industry funding. This integration enables enriched discussion and anticipation capabilities to address new AI developments and impacts, making the approach more proactive and less responsive.

However, the pace of AI technology research, creation, and deployment is currently noticeably fast. Considering that LLM generative AI can even speed up software coding and database manipulation, shorter cycles of AI systems implementation can be expected for the next few years. For instance, Hugging Face, a global AI community, has more than 624,402 AI models available (2024). The portal *There is an AI for That* features 12,463 AIs for 15,287 tasks and 4,804 jobs (2024). The ethical AI database, however, has only 298 validated startups (2024). Overall, there is a high investment flow in generative AI technologies (2024), and expectation with the development of LAM (large agent models) and AutoAI (Cao, 2022; Radanliev & Roure, 2023). It means that traditional foresight methods, which typically require more time for framing, scanning, scenario development, and impact analysis, need to reevaluate process cycles, balancing learning while capturing the timing of key emerging technologies. The same applies to policy design, which relies on technology interpretation of current regulations. Therefore, the same generative AI can be better understood to support both integrated foresight and policy design processes in the context of AG.

AG, foresight, and policy design with GEN_AI

On public management, consultation on databases such as Science Direct and evidence shows how 'artificial intelligence' and 'public management,' despite less than ten articles from 1994 to 2017, started to grow from 18 in 2018 to 116 in 2024, evidencing the integration of these two fields. In the foresight field, the discussion of AI impacts on foresight is not new (Bevolo & Amati, 2020; Farrow, 2020; Geurts et al., 2022) and addresses the implications of AI technologies on phases such as scoping, scanning, scenario building, impact assessment, and options strategizing. That can include the adoption of ontology generation, NLP, text mining, data processing, simulation, and several other types of AI technologies (Geurts et al., 2022). However, the inherent complexity, uncertainty, and unstructured data related to the foresight process lead to more hybrid approaches, integrating human and machine capabilities. This article is more concerned with considering the specific GEN_AI technologies.

Generative AI, by common understanding, can generate new content from a specific prompt and a trained database, ranging from text, image, audio, video, 3D models, synthetic data, and styles, which can help the conversion process. Each one of these objects operates with distinguished algorithms and development pathways. For instance, Goodfellow et al. (2020) introduced generative adversarial networks (GANs) to improve the generation of realistic images, using neural networks — generator and discriminator — for image-related tasks. Vaswani et al. (2017) introduced the generative pre-trained transformer, which provided the basis for large language models (LLMs), enabling them to deal with large text corpora. These models focus on predicting or generating the next token in a sequence, inferring the most likely response to a prompt input. This is grounded in statistical models that understand the probabilities of text distribution in training data and adopt a self-attention mechanism for better context over text corpora. Despite being two fields of research, foresight can be both a consumer of image and text creation to better deal with visions of the future and scenario narratives. Therefore, advancements in generative AI, such as GANs or LLMs, directly impact foresight and policy design activities. However, since most aspects of foresight and policy design deal with text data, we focus on understanding LLMs for this study.

Large language models (LLMs) represent a critical advancement in natural language processing (NLP), characterized by their ability to generate a coherent and contextually relevant text corpus across various domains. It is important to note that the neural network architecture (Shen et al., 2023) and the transformer architecture, introduced by Vaswani et al. (2017) in the seminal paper "*Attention is All You Need*," are central issues of LLMs. This architecture eschews traditional recurrent layers in favor of attention mechanisms, enabling LLM models to weigh the importance of different words in a sentence, regardless of their positional distance. This is essential because words can have different meanings due to their context; for example, 'interest' can be motivational or financial. That was a common problem when dealing with traditional text mining techniques. However, LLMs must be pre-trained on diverse and extensive corpora to operate correctly. This enables high model performance and fine-tuning to develop a deep understanding of language patterns and nuances, providing valuable responses to prompts. Therefore, it can be observed that the generative capabilities of LLMs are not merely syntactic but extend to generating text that is contextually appropriate and semantically rich. However, there are limitations that users need to be aware of.

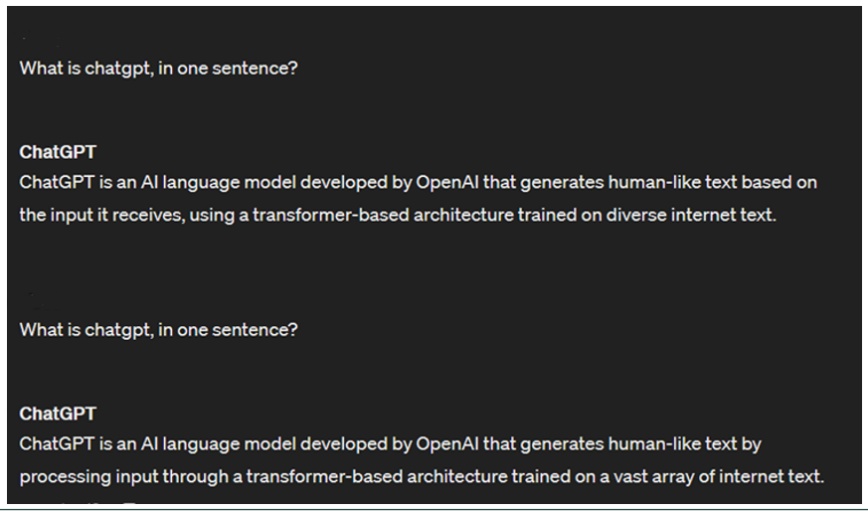
Despite LLMs accelerating text creation based on large databases — which can be necessary for scanning activities, scenario narratives, and impact analysis — they currently face technical limitations such as model collapse (Dohmatob et al., 2024; Shumailov et al., 2023), AI hallucinations (Athaluri et al., 2023; Emsley, 2023; Salvagno et al., 2023), and AI bias (Mikalef et al., 2022; Suresh & Guttag, 2021), requiring more AI and foresight literacy to be better developed by experts and non-expert users (Jokinen et al., 2023). All phenomena are new and are being investigated more extensively by research groups. Model collapse occurs when an LLM ingests synthetic data over time, or data produced by AI (such as blogs written by AI and not originally by humans), which often lack diversity. This affects the distribution curve, making the LLM less accurate over time and reducing its ability to provide valuable answers.

AI hallucinations are considered inherent to LLMs and occur when the outputs generate specific fragments of text that are not 100% accurate. This can be a partial or total distortion of reality. AI bias, on the other hand, refers to biases in the final response, such as algorithmic discrimination, that originate from data bias, representation bias, measurement bias, development bias, aggregation bias, evaluation bias, or even learning bias (Suresh & Guttag, 2021).

The central point in these limitations is that, according to Bender et al. (2021), humans tend to attribute meaning where there is none, misleading themselves or others while taking synthetic text as meaningful and credible, even when it is incorrect. All this requires users who adopt LLMs to have some foresight and AI literacy (Jokinen et al., 2023) to better identify possible mistakes from the LLM solutions they use in this context.

At the same time, many LLM developers use humans in the loop to train, tune, and validate LLMs at several stages of the AI development lifecycle to address these limitations. Therefore, users do not usually see this human knowledge in the backend of an LLM GEN_AI.

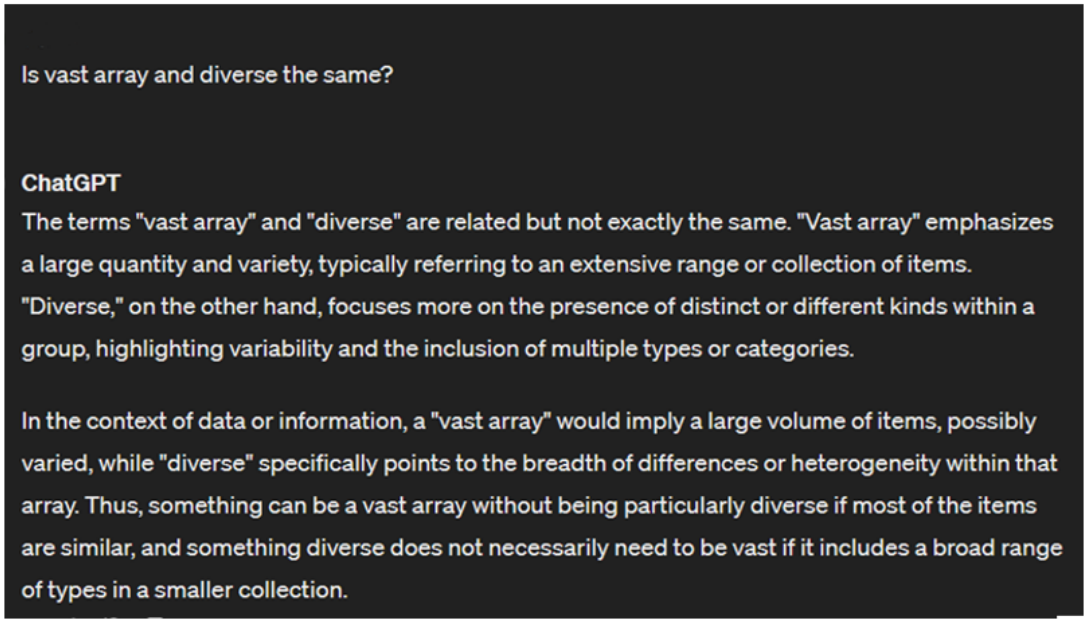
However, more importantly, it is critical to deal with the generative nature of AI. A given LLM model will always express itself differently when answering the same question, even if it is asked by the same user. This can have positive and negative implications, depending on the context of the knowledge service. One example of understanding this phenomenon is asking a GPT the same question and observing how the answer may vary (Figure 1). Notice the end of the answer. Especially when dealing with more complex responses, changes in wording can lead to different interpretations and meanings, influencing perception creation and action.



Source: OpenAI. (2024). ChatGPT 4 . https://openai.com/pt-BR/

Figure 1. Example of different responses over the same question with GEN_AI (Test 1).

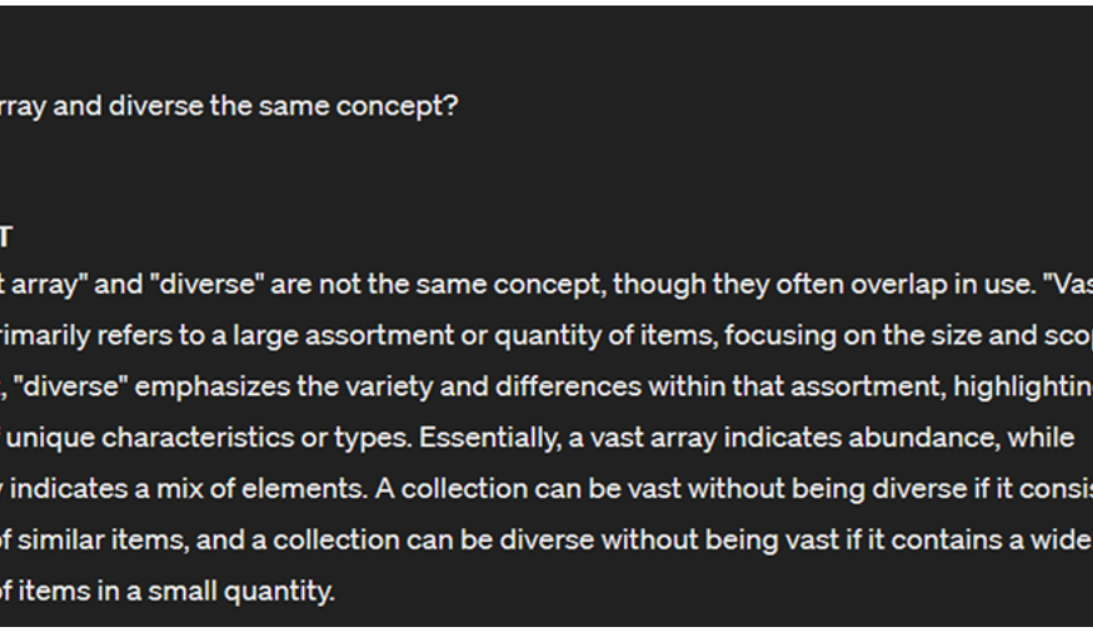
What is interesting about this test is that when asked whether 'vast array' and 'diverse' have the same meaning, the LLM indicates that they are different concepts (Figure 2). This implies, by essence and logic, that the two previous answers have significant differences.



Source: OpenAI. (2024). ChatGPT 4 . https://openai.com/pt-BR/

Figure 2. Example of response with GEN_AI (Test 2).

Then, we reinforced the same question by adding 'same concept' (Figure 3). Therefore, due to the response provided by logic, there is an inherent contradiction in the answer to "What is ChatGPT, in one sentence?"



Source: OpenAI. (2024). ChatGPT 4 . https://openai.com/pt-BR/

Figure 3. Example of response with GEN_AI (Test 3).

This could be a slight difference in the context of the example. However, one must consider this phenomenon in a complex discussion involving foresight and policy design, with several concepts embedded and interpreted by non-expert users adopting this technology for automation. They expect that, from a specific prompt, responses about scenarios, impacts, and policy implications will be quickly provided and with clarity of concepts and definitions. How will this shape perception and meaning?

Although several LLM solutions are in place in 2024 (OpenAI ChatGPT, Google Gemini, Microsoft Copilot, Anthropic Claude, open-source LangChain), they all face these issues at different levels. These phenomena

become more complex as they ingest more synthetic data built upon AI and published by humans. Therefore, assessing generative AI solutions with a scientific approach and expert analysis will be more than necessary. Grounded in this problem, we present the technology and methodology adopted to assess LLM AI technology for AG in the next section.

Explainable AI in an institutional and anticipatory governance context

Traditionally, advanced AI models such as deep learning systems (e.g., neural networks) are considered 'black boxes' due to their complex decision-making processes, making it difficult to explain how they arrive at specific conclusions (Islam et al., 2021; Saeed & Omlin, 2023). Explainable AI (xAI) refers to a set of methods and techniques designed to make AI models — including complex generative AI systems, and more specifically, models like GPT (generative pretrained transformer) — more transparent and interpretable for human users.

In anticipatory governance, where future-oriented decisions can support AI, explainable AI is crucial to ensure stakeholders can understand and trust the AI's outputs. Generative AI, which can produce novel data or predictions based on learned patterns, amplifies this complexity. When using GPT models for anticipatory governance, the model generates complex scenarios and potential futures based on vast training data. While highly sophisticated, these predictions often lack transparency regarding how they were derived.

The need for xAI is amplified in this context, as decisions derived from GPT-generated insights must be justifiable, particularly when influencing public policy, crisis management, or long-term governance planning (Islam et al., 2021; Saeed & Omlin, 2023). Therefore, generative AI creates an urgent need for interpretability in scenarios where anticipatory governance is used for policy design, crisis management, and long-term planning. xAI seeks to address these issues by offering clear, accessible explanations for how generative AI models function and generate outputs, ensuring these systems align with governance objectives and values (Longo et al., 2024; Nagahisarchoghaei et al., 2023).

When using GPT models for anticipatory governance, explainability is necessary to ensure that decision-makers understand both the data sources and the reasoning process of the model. The principles of xAI in this context can be broken down into three fundamental orientations: source-oriented, representation-oriented, and logic-oriented explanations (Table 1) (Longo et al., 2024; Nagahisarchoghaei et al., 2023; Saeed & Omlin, 2023).

Table 1. Approaches for XAI.

Approach	Description	Example
Source-oriented explanations:	These explanations focus on providing insights into the origins of the data or features that influenced the AI model's predictions. In anticipatory governance, where decisions must be based on reliable data, source-oriented explanations help policymakers understand which datasets, historical data, or external knowledge sources contributed to the AI's decision-making process (Nagahisarchoghaei et al., 2023). This is critical in ensuring transparency and reliability, particularly when decisions impact large populations. These explanations focus on identifying the sources of data that GPT relied on to generate predictions. Since GPT models are trained on vast datasets, source-oriented explanations are critical in ensuring that the model's outputs are based on reliable, relevant, and up-to-date data. For example, if GPT suggests a policy intervention for climate change, it is important to trace which scientific reports, historical data, or expert opinions contributed to the recommendation.	Example: When GPT forecasts future energy demands, a source-oriented explanation might reveal that it used data from global energy reports and past governmental energy policies (Nagahisarchoghaei et al., 2023).
Representation-oriented explanations:	This principle emphasizes how the AI model internally represents and processes data to reach its conclusions (Longo et al., 2024). It is crucial in understanding how specific inputs (e.g., policy variables, environmental data) are processed through the model's layers or transformations. In anticipatory governance, representation-oriented explanations help decision-makers grasp the structure of the model and how various inputs are weighted or transformed to produce outputs. This type of explanation addresses how GPT internally represents inputs (such as policy variables or socioeconomic data) and transforms them to generate predictions. In anticipatory governance, where accurate representation of data (like population demographics or economic indicators) is critical, understanding how GPT processes this information is crucial for trust and accuracy in decision-making.	Example: A representation-oriented explanation might show how GDP data is weighted within the model when GPT predicts economic outcomes over the next decade (Longo et al., 2024).
Logic-oriented explanations:	Logic-oriented explanations focus on the reasoning process of the AI system — how the model uses its internal logic (such as rules, if-then statements, or decision trees) to arrive at a conclusion (Saeed & Omlin, 2023). This is essential for understanding not just what decision the AI made, but why it made that decision. In governance contexts, this type of explanation is critical for justifying policies or scenarios that AI models suggest, especially when human oversight is required. Since GPT relies on probabilistic language models, it's vital to understand why the model chose a particular sequence of ideas or words to predict certain scenarios. In governance contexts, logic-oriented explanations ensure that the reasoning process of GPT aligns with human expectations and ethical standards.	Example: GPT recommends stricter regulations on carbon emissions; a logic-oriented explanation would detail the reasoning behind the recommendation, such as past correlations between emissions levels and environmental impact (Saeed & Omlin, 2023).

Note. Proposed by the authors.

Logic-oriented explanations: Logic-oriented explanations focus on the reasoning process of the AI system — how the model uses its internal logic (such as rules, if-then statements, or decision trees) to arrive at a conclusion (Saeed & Omlin, 2023). This is essential for understanding not just what decision the AI made,

Therefore, when applying xAI in generative AI systems for anticipatory governance, especially taking into account the institutional context, more than a system explanation, it is critical to understand that institutional explanations are required in systems that deal with public needs such as health (Theunissen & Browning, 2022), security, housing, and other governmental affairs. That's why, especially in public contexts, several critical factors must be considered for AI adoption:

(a) Trust and accountability: trust in AI outputs is essential, especially when addressing future scenarios that influence long-term policies. xAI enhances trust by providing interpretable insights into how generative AI models arrive at predictions, ensuring that AI-driven policy decisions are justifiable and transparent (Longo et al., 2024; Nagahisarchoghaei et al., 2023).

(b) Ethical and regulatory compliance: xAI ensures compliance with regulations like the European General Data Protection Regulation (GDPR) by providing clear explanations and justifications for AI-generated decisions. This ensures that governance decisions are ethically sound and legally defensible (Islam et al., 2021; Saeed & Omlin, 2023).

(c) Bias and fairness: in anticipatory governance, xAI helps identify and mitigate biases in generative AI systems. This is crucial to ensure that AI models do not produce predictions that exacerbate inequalities, thereby promoting fairness in governance outcomes (Saeed & Omlin, 2023; Yang et al., 2023).

(d) Human-AI interaction and trustworthiness: xAI plays a vital role in ensuring that AI-generated insights are presented in ways that match the needs of different stakeholders, from high-level officials to technical experts. Contextualized explanations foster trust and improve decision-making in governance (Longo et al., 2024; Saeed & Omlin, 2023).

(e) Adaptation and personalization: xAI ensures that explanations evolve as generative AI models adapt to new data or changing conditions. This ensures that policymakers can continue to rely on AI outputs without losing interpretability or trust (Islam et al., 2021; Longo et al., 2024).

There are some specific techniques, such as retrieval-augmented generation (RAG), is an advanced approach that enhances explainable AI by combining retrieval-based and generation-based models. This hybrid approach improves the transparency and interpretability of generative AI systems in anticipatory governance, incorporating experts and human-in-the-loop influence over the final AI output. Especially in the field of AG, this is critical for:

(a) Evidence-based explanations: RAG allows the AI system to retrieve relevant information from external data sources, justifying its decisions. In anticipatory governance, where AI forecasts future scenarios or makes policy recommendations, RAG can pull from large datasets or relevant policy documents to back its predictions. This enhances the trustworthiness of AI systems by grounding predictions in real-world data (Islam et al., 2021; Nagahisarchoghaei et al., 2023).

(b) Mitigating hallucination: generative AI models are prone to 'hallucinations,' where they generate incorrect or fabricated outputs. RAG mitigates this issue by ensuring the generative model bases its predictions on factually accurate and retrievable information from external sources. This is crucial in policymaking, where decisions based on AI predictions can have far-reaching consequences (Islam et al., 2021).

(c) Contextualizing explanations: RAG provides contextualized explanations by dynamically retrieving domain-specific information. In governance, where stakeholders require specific justifications for decisions, RAG ensures that AI outputs are linked to real-world data, facilitating better understanding among policymakers (Longo et al., 2024).

(d) Enhancing human-AI collaboration: RAG fosters better collaboration by providing easy-to-verify references and justifications for its predictions. Policymakers can review the retrieved data and cross-check AI-generated predictions, thus increasing the reliability and accountability of AI-driven governance systems (Islam et al., 2021).

With the awareness of the xAI context, and more specifically, an institutional xAI, we proceeded to artifact design and assessment, considering elements to be observed, such as model collapse (Dohmatob et al., 2024; Shumailov et al., 2023), which means signs of degenerative response in the model that can compromise the output credibility; AI hallucinations (Athaluri et al., 2023; Emsley, 2023; Salvagno et al., 2023), where the AI generates incorrect output that cannot be supported by cross-checking; AI bias (Mikalef et al., 2022), which involves identifying traces of algorithmic discrimination or other specific bias; generative interference, where the generative nature of the AI provides differences in the response structure, compromising the tool's goal and the credibility and usefulness of the information — considering a qualitative expert analysis.

TECHNOLOGY AND METHODOLOGY

This is a design science research in terms of epistemology and method. That means experts designed and assessed an artifact from a qualitative perspective, including software development. This GEN_AI agent was designed to help researchers and policymakers gain insight into the regulatory impacts of emerging tech. It's

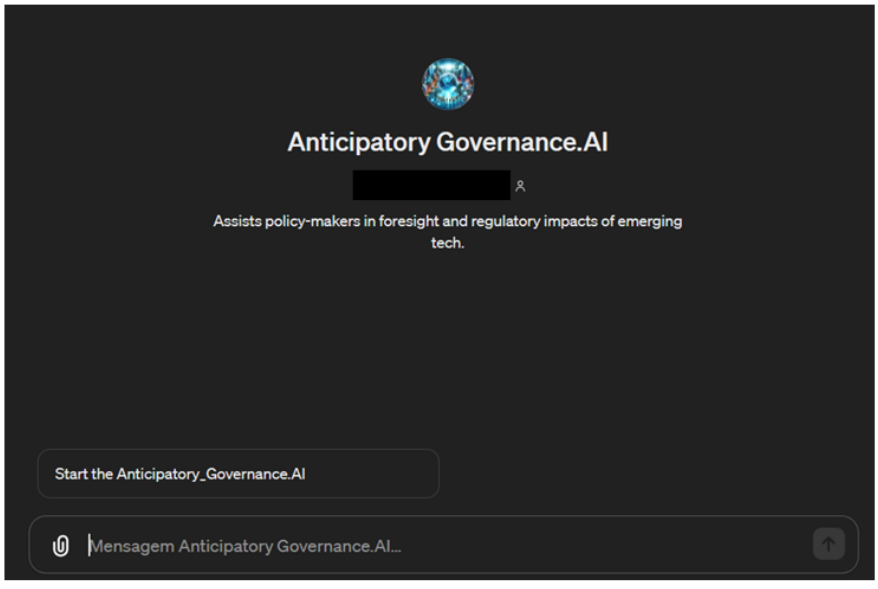
based on OpenAI ChatGPT 4 (2024). As a GPT, it was programmed with specific behavioral instructions and 15 steps that take about 10 minutes to process end-to-end, creating a chain-of-thought reasoning. That approach enables contextual cumulative knowledge from the LLM, increasing the probability of finding information from the problem identification to emerging tech analysis, scenario building, roadmapping, possible impacts, and risk analysis of a given public policy in a given country. Since it runs over GPT-4, the capabilities of web browsing to access real-time data and code interpreters were activated, enabling data analysis and other features.

The GEN_AI agent was designed to be contextual over a specific country of analysis, contextual over a specific policy domain (health, security, labor, environment...), and to adopt a human-in-the-loop perspective (Grønsund & Aanestad, 2020). In this sense, these 15 steps have intermediary gates, where humans can input the need for more specification and review information, avoiding that the entire process runs without human interference, which could significantly interfere with the final result. Therefore, each step can be revised, further details requested, or new data added.

For the cases, we selected three scenarios: one for electoral policy, another for environmental policy, and another for labor policy. In these cases, we will describe in the following sections the outputs of the GPT, as well as an analysis conducted, including a more in-depth study of fact-checking. For the report, we won't mention the names of private companies and startups cited by the GPT. Four evaluations by four PhD experts in foresight, roadmapping, AI, and public policy were involved for artifact evaluation, based on five criteria in the context of xAI described in the evaluation session.

ANALYSES OF THE CASES

We ran the three scenarios separately, from end to end, using the anticipatory GEN_AI (Figure 4). For the article presentation, we provided a comparison for each case scenario at each main stage of the GPT in the AG flow. The AI begins inquiring about the country under analysis and then the policy domain. In this case, we informed 'Brazil' and 'electoral, environmental, and labor.' The following sets of comparisons refer to each step of the AI.



Source: Content generated by ChatGPT 4 (Anticipatory Governance.AI).

Figure 4. AG LLM welcome screen.

Stakeholders involved in the system

Stakeholder mapping and engagement are critical for AG in both foresight for collective intelligence and policy design for policy coordination and legitimacy (Heo & Seo, 2021; Kunseler et al., 2015; Myllyoja et al., 2022). After understanding the context of the analysis, the AI began analyzing the stakeholders in the given system, categorizing them into government, private sector, universities and institutions, civil society, and funding bodies. In this case, the initial response, in three scenarios, was more general, necessitating more detailed specifications. For this article, we suppressed the names of the companies. Upon observing the selected stakeholders in government, private sectors, universities and research institutions, civil society, and funding bodies, the responses provided were considered valid and in line with Brazilian reality (Table 2).

Table 2. LLM comparison of stakeholders involved in the system in the three scenarios.

Electoral Case	Environmental Case	Labor Case
Government TSE, Regional Courts, Ministry of Justice, National Congress	**Government** Ministry of Environment, IBAMA, ICMBio	**Government** Ministry of Labor and Employment (MTE), Ministry of Economy, Public Prosecutor's Office for Labor (MPT)
Private Sectors Electronic Voting System Providers, CyberSecurity Firms, Data Analytics Companies	**Private Sectors** Oil & Energy, Mining, Pulp & Paper Industry	**Private Sector** Major Employers and Industry Associations such as CNI, Tech Companies, and Startups
Universities and Research Institutions IBOPE	**Universities and Research Institutions** Federal Universities & INPA	**Universities and Research Institutes** —
Civil Society Transparency Brazil, OAB	**Civil Society** ISA, Amazon Watch, Greenpeace	**Civil Society** CUT, ETHOS
Funding Bodies International Foundations, National Government Budgets	**Funding Bodies** Amazon Fund, BNDES, World Bank, United Nations	**International Actors and Funding** International Labour Organization (ILO), World Bank and Inter-American Development Bank

Note. Content generated by ChatGPT 4 (Anticipatory Governance.AI).

Main classes of AI technologies (research-level)

The core idea of anticipatory governance, especially considering the foresight perspective, is to identify emerging technologies, in this case, AI technology. For that approach, we draw upon the standpoint of scientific foresight (Van Woensel, 2020), which addresses that emerging scientific research can lead to further technology development. These represent the AI tech classes observed by the GPT assistant, with specific details and explanations suppressed to focus on the classes. All AI technologies and classes made sense for the problem addressed (Table 3). However, in the Labor Case, AI did not identify technologies but articles, resulting in an AI hallucination. For example, the first article referenced, *"Automation and job loss: The Brazilian case"*, doesn't exist with this exact title. This evidence is a limitation that requires a sense of awareness among non-experts on the topic.

Table 3. LLM comparison of main classes of AI technologies in the three scenarios.

Electoral Case	Environmental Case	Labor Case
AI and Voter Fraud Detection	AI for Real-Time Monitoring of Deforestation	Automation and Job Replacement Risks: A Sectoral Analysis in Brazil
Cybersecurity and AI	Predictive Modeling in Climate Change Studies	The Impact of AI on Remote Work: Enhancing Productivity and Flexibility
Machine Learning for Voter Behavior Analysis	AI and Biodiversity Conservation	AI in Human Resources: Transforming Talent Acquisition and Management
Blockchain for Voting Integrity	AI in Water Resource Management	Economic Implications of AI Integration into the Brazilian Labor Market
Natural Language Processing for Social Media Monitoring	Automated Pollution Tracking Systems	Ethical Considerations of AI in the Workplace: A Brazilian Perspective
AI in Electoral Logistics	AI for Renewable Energy Management	The Role of AI in Enhancing Worker Skills and Education
AI for Accessible Voting	AI-Driven Waste Management Solutions	AI and the Gig Economy: Implications for Labor Law
Deep Learning for Image and Video Analysis	AI in Ecosystem Services Valuation	Predictive Analytics in Labor Markets: Forecasting Industry Trends
AI-driven Policy Simulation	AI for Microclimate Prediction	AI, Robotics, and the Future of Unions: A New Era for Collective Bargaining
AI Deepfake in Elections	Genetic Algorithms for Biodiversity Conservation	Social Impact of AI on Work: A Study on Worker Displacement and Welfare
	Neural Networks for Soil Health Monitoring	
	AI in Plastic Degradation Research	

Note. Content generated by ChatGPT 4 (Anticipatory Governance.AI).

AI innovations — Funded startups

While the previous analysis focuses on science and technology, it concentrates on startups receiving significant funding, enabling them to scale their solutions further and diffuse innovation (Oehmichen et al., 2023; Rojas & Tuomi, 2022). In this analysis, the generative aspect of GPT became more evident: while the Electoral Case provided a specific description of startups, the Environmental Case was organized by name, funding, focus, and government contracts. The Labor Case provided a simple description of all startups, confirming their existence through cross-checking. GPT presents new solutions in this phase that can impact the domains (Table 4).

Table 4. LLM comparison of main classes of AI innovations in the three scenarios.

Electoral Case	Environmental Case	Labor Case
A startup that uses blockchain and biometrics to enable secure and accessible mobile voting. Significant funding and trials in various U.S. states highlight its potential applicability for remote voting solutions in Brazil.	Startup A Funding: $150 million Focus: AI-driven deforestation monitoring systems Government Contracts: Partnered with the Brazilian government to monitor Amazon deforestation.	Specializes in AI-driven analytics for optimizing workplace productivity and employee well-being.
Known for providing electronic voting systems and services, Startup B also integrates AI to enhance the security and efficiency of its solutions. Its global presence and extensive experience make it a key player in the market.	Startup B Funding: $120 million Focus: AI for sustainable urban planning Government Contracts: Collaboration with several Brazilian cities to implement smart, sustainable urban developments.	AI platform that streamlines recruitment processes, enhancing the speed and accuracy of hiring.
Specializes in online voting and electoral management solutions, using encryption and AI to secure elections. Its technologies could help Brazil enhance transparency and reduce electoral fraud.	Startup C Funding: $90 million Focus: AI for water quality and usage analytics Government Contracts: Deployed technology in municipal water systems across Southeast Brazil to optimize water usage and quality monitoring.	Offers AI-powered skills assessment and training tools to help workers adapt to changing job demands.
Focuses on detecting deepfakes and synthetic media, which are crucial for combatting misinformation during elections. Their AI-driven tools can help ensure the authenticity of election-related communications.	Startup D Funding: $85 million Focus: Optimization of solar energy grids using AI Government Contracts: Works with the Ministry of Energy to enhance solar energy distribution.	Develops AI solutions for automating repetitive tasks in manufacturing and services, reducing the need for manual labor.
An innovative startup that offers a blockchain-based voting platform, ensuring transparency and security. Their technology could be used for smaller-scale elections or referendums in Brazil.	Startup E Funding: $80 million Focus: AI-powered wildlife monitoring and poaching prevention systems Government Contracts: Partnered with ICMBio to deploy in national parks.	Focuses on employee engagement and retention using AI to predict turnover and improve job satisfaction.
Develops AI-driven identity verification systems that could be integrated into electronic voting processes to ensure voter identity integrity.	Startup F Funding: $75 million Focus: AI for climate prediction and modeling Government Contracts: Provides climate data analytics to the Brazilian Ministry of Environment.	Provides AI-driven safety monitoring solutions to reduce workplace accidents and ensure compliance with labor regulations.
Offers a blockchain voting solution designed to be tamper-proof and transparent, potentially suitable for governmental elections if tested and certified by Brazilian electoral authorities.	Startup G Funding: $70 million Focus: AI for risk assessment of biodiversity data Government Contracts: Collaboration with environmental agencies to monitor biodiversity in the Atlantic Forest.	A platform that uses AI to match freelancers with gig opportunities, supporting the flexibility of the gig economy.
Develops AI tools for analyzing public opinion and voter sentiment from social media data, which could aid political parties and electoral bodies in understanding voter needs and trends.	Startup H Funding: $65 million Focus: AI solutions for waste management and recycling optimization Government Contracts: Implemented in major Brazilian cities to enhance waste sorting and recycling processes.	Uses machine learning to offer predictive insights into workforce planning and development.
Uses AI and machine learning to improve the accuracy and speed of vote counting. Their technology could enhance the auditability of electronic voting results in Brazil.		An AI tool that helps companies ensure compliance with labor laws, minimizing legal risks.
Specializes in AI tools for protecting voter registration databases and ensuring the integrity of voter rolls, a critical aspect of the electoral process that often faces challenges in Brazil.		Combines AI with educational technology to provide career-oriented training for specific industry needs.

Note. Content generated by ChatGPT 4 (Anticipatory Governance.AI).

Regulations affected by AI

After analyzing emerging and current technologies, the AI sought to identify regulations in the domain that may be affected, considering the discussion of how AI can reshape the body of rules (Andrade & Kontschieder, 2021; Pflanzer et al., 2023). In all three cases, the primary laws at the federal level were mentioned in the analysis and are coherent with the analysis. A more detailed description was suppressed for the table. The reference numbers of the laws are also correct (Table 5). This analysis aims to create a reference for further development.

Table 5. LLM comparison of Brazilian regulations affected by AI in the three scenarios.

Electoral Case	Environmental Case	Labor Case
Brazilian General Data Protection Law (LGPD)	Forest Code (Law no. 12,651, 2012)	*Consolidação das Leis do Trabalho (CLT)*
Electoral Code (*Código Eleitoral*, Law no. 4,737, 1965)	National Policy on Climate Change (Law no. 12,187, 2009)	*Lei Geral de Proteção de Dados (LGPD)*
Law on Political Parties (*Lei dos Partidos Políticos*, Law no. 9,096, 1995)	National Solid Waste Policy (Law no. 12,305, 2010)	*Marco Civil da Internet*
Law of Clean Record (*Lei da Ficha Limpa*, Complementary Law no. 135, 2010)	Water Law (Law no. 9,433, 1997)	Law no. 13,467 (2017) (Labor Reform Law)
Resolutions by the Superior Electoral Court (TSE)	Environmental Crimes Law (Law no. 9,605, 1998)	Law on Telework (Law no. 13,467, 2017)
Cybersecurity Decree (Decree no. 10,222, 2020)		

Note. Content generated by ChatGPT 4 (Anticipatory Governance.AI).

Foresight scenarios

With the context of country, domain, stakeholders, emerging technologies, ongoing innovations, and regulations, the GPT creates three scenarios: baseline, alternative, and disruptive perspectives (Wright et al., 2020). The differences in the overall structure (title, description, technologies involved, probability, impact, or potential changes) can be noticed in the three cases, reflecting variations in how information is presented due to the generative nature of AI (Table 6). The plausibility and probability of the scenarios could be considered coherent despite GPT not providing a timeline for future scenarios as a reference. However, since the purpose of scenarios is to build new visions or perspectives, the central point is what policymakers can learn from them. One issue to be addressed is that running several exercises may result in different scenarios. This can lead to information overload for the decision-maker, potentially undermining the credibility of the scenarios due to the abundance of information. This is a critical point of discussion in foresight: the cognitive acceptance of scenarios when participants are involved and are learning with the process and when scenarios are simply and quickly provided by a tool.

Table 6. LLM comparison of foresight scenarios.

Electoral Case	Environmental Case	Labor Case
Baseline Scenario: Gradual Integration Description: In this scenario, AI technologies are gradually integrated into the Brazilian electoral system, focusing primarily on enhancing existing cybersecurity measures and improving the efficiency of voter registration and verification processes. Technologies Involved: Basic AI applications in cybersecurity, enhanced biometric verification systems. Impact: Moderate improvements in electoral security and process efficiency, with slow adoption rates due to bureaucratic inertia and concerns over voter privacy. Probability: High (60%)	**Baseline Scenario: Gradual Integration** Description: AI technologies are incrementally integrated into existing environmental management systems. Technologies such as AI-driven deforestation monitoring and smart water management are adopted at a moderate pace, aligning closely with current technological uptake trends. Probability: High (70%) Impact: Improves efficiency and accuracy of environmental monitoring and enforcement but does so within the constraints of slowly evolving regulatory frameworks. Limited by budget constraints and gradual policy adaptations.	**1. Baseline Scenario: Gradual Integration** Description: AI technologies are progressively adopted within existing legal and business frameworks. Automation is primarily focused on enhancing productivity in sectors where it complements human labor. Remote working tools become more widespread, and AI-driven HR solutions are commonly used for talent management. Potential Changes: Incremental updates to labor laws to address remote work norms and data privacy concerns. Probability: High (70%) — This scenario assumes a continuation of current trends without drastic shifts in technology adoption or regulatory changes.
Alternative Scenario: Comprehensive Digital Transformation Description: Brazil embraces a more comprehensive digital transformation of its electoral process. This includes widespread adoption of blockchain for vote tracking, AI-driven analytics for real-time monitoring of electoral integrity, and extensive public engagement through digital platforms. Technologies Involved: Blockchain, advanced AI in data analytics and fraud detection, interactive digital platforms for voter education. Impact: Significant enhancement in transparency and trust in elections, increased voter participation, but raises issues regarding digital divide and misinformation. Probability: Medium (30%)	**Alternative Scenario: Proactive Policy Leadership** Description: The Brazilian government takes a proactive role in integrating AI into environmental policies. It enacts swift legislative updates, invests in AI research, and establishes partnerships with technology companies and academia. Probability: Medium (40%) Impact: Accelerates the deployment of AI solutions across a broader range of environmental issues, including biodiversity conservation, climate adaptation, and sustainable urban planning. This scenario optimizes outcomes but requires significant coordination and investment.	**2. Alternative Scenario: Rapid Transformation** Description: A surge in AI technology adoption driven by significant investments from both domestic and international tech firms. Extensive automation in manufacturing and service sectors leads to substantial job displacement, but also the emergence of new job categories in tech and AI maintenance. Potential Changes: Rapid legal adaptations required, including new frameworks for AI ethics, automation taxes, and re-skilling programs. Probability: Medium (40%) — While feasible, this scenario depends on aggressive investment and public acceptance of rapid tech changes.

(continue)

Table 6. LLM comparison of foresight scenarios. (continuation)

Electoral Case	Environmental Case	Labor Case
Disruptive Scenario: Fully Automated and Decentralized Elections Description: A radical shift where AI and blockchain lead to fully automated and decentralized voting processes. This scenario sees the implementation of sophisticated AI systems capable of managing entire elections with minimal human oversight. Technologies Involved: Advanced AI systems, comprehensive blockchain networks, autonomous decision-making algorithms. Impact: Drastic reduction in human error and fraud, unprecedented levels of transparency and efficiency. However, it may face strong opposition due to concerns over machine errors, potential biases in AI algorithms, and loss of traditional voting rituals. Probability: Low (10%)	Disruptive Scenario: Rapid Technological Breakthroughs Description: A series of rapid technological advancements occur, outpacing current legislative and infrastructural readiness. AI technologies such as genetic algorithms for species conservation and neural networks for microclimate prediction become highly advanced and widely available. Probability: Low (20%) Impact: Drastically alters the environmental management landscape. While offering significant benefits like real-time ecosystem monitoring and management, it poses challenges in terms of regulation, privacy, and socio-economic disparities due to uneven access to technologies.	3. Disruptive Scenario: AI-Driven Economy Description: AI fundamentally changes the nature of work. Most routine jobs are automated, and the gig economy shifts to project-based AI management roles. Significant disruptions in traditional employment structures lead to a redefinition of work, income, and social security. Potential Changes: Comprehensive overhaul of labor laws, introduction of universal basic income, and new regulations governing AI entities. Probability: Low (20%) — This scenario represents a radical departure from current trajectories and would require breakthroughs in AI capabilities and societal shifts.

Note. Content generated by ChatGPT 4 (Anticipatory Governance.AI).

Roadmapping

To provide context for the user, GPT is oriented to create a roadmap based on foresight and technology impacts (Hussain et al., 2017), incorporating time technologies, milestones, and expected impacts. It should be noted that the year structure changed, and for the Labor Case, a completely new structure was created when comparing scenarios (Tables 7, 8, and 9). Roadmapping generation is an intermediate layer for understanding the impact level at a given time, integrating foresight with policy design.

Table 7. Comparison of AI technology impact roadmap in the Electoral Case.

Year	Technology	Milestones	Expected Impact
2024	Basic AI and Cybersecurity	Implement AI-driven security enhancements in existing electoral systems.	Improve resistance against cyber-attacks; enhance data protection.
2025	Biometric Verification	Widespread adoption for voter identification during elections.	Reduce voter fraud; streamline voter registration processes.
2026	Blockchain for Vote Tracking	Pilot projects in local elections to test blockchain applications.	Increase transparency; build public trust in voting results.
2027	Advanced-Data Analytics	Deploy AI for analyzing voting patterns and fraud detection.	Enhance real-time monitoring and response during elections.
2028	Interactive Digital Platforms	Launch platforms for voter education and engagement.	Increase voter participation; improve electoral literacy.
2030	Full Blockchain Implementation	Nationwide rollout of blockchain-based voting systems.	Ensure tamper-proof elections; decrease administrative costs.
2032	Autonomous AI Systems	Trial fully automated AI systems for managing elections.	Minimize human error; optimize efficiency of electoral processes.
2035	Fully Automated Elections	Evaluate feasibility of nationwide automated elections.	Revolutionize election administration; possibly face societal pushback.

Note. Content generated by ChatGPT 4 (Anticipatory Governance.AI).

Table 8. Comparison of AI technology impact roadmap in the Environmental Case.

Timeframe	Technology	Milestones and Actions	Expected Impact
Short-term (1-2 years)	AI for Deforestation Monitoring	Implementation in critical regions like the Amazon.	Enhanced real-time monitoring and response capabilities.
	AI in Waste Management	Pilot projects in major cities.	Improved recycling rates and waste reduction.
Mid-term (3-5 years)	AI for Water Resource Management	Nationwide deployment in agriculture and urban areas.	Optimized water usage, and reduced waste and pollution.
	AI for Biodiversity Conservation	Roll out in protected areas and hotspots.	More accurate species tracking and habitat assessment.
Long-term (6-10 years)	AI for Climate Prediction and Modeling	Integration with national climate strategies.	Better preparedness and adaptation strategies.
	AI in Renewable Energy Management	Adoption by major energy providers.	Increased efficiency in renewable energy utilization.

Note. Content generated by ChatGPT 4 (Anticipatory Governance.AI).

Table 9. Comparison of AI technology impact roadmap in the Labor Case.

Year	Milestones	Scenario 1: Gradual Integration	Scenario 2: Rapid Transformation	Scenario 3: AI-Driven Economy
2024	Initial AI integration in HR and remote work	Begin implementation	Begin implementation	Begin implementation
2025	Review and first updates to labor laws	Minor updates	Significant legal adaptations	Major legal overhauls
2026	Expansion of AI in service sectors	Continued adoption	Rapid job displacement	Wide-scale automation
2027	A national debate on automation impacts	Evaluation of impacts	Public and policy response	Economic restructuring
2028	Implementation of AI ethics and regulation	Strengthening data privacy	Automation taxes, AI ethics laws	Regulation of AI entities
2029	Re-skilling and educational reforms	Targeted programs	Extensive re-skilling initiatives	Universal re-skilling
2030	Review of social security systems	Adaptations to benefits	New welfare models	Introduction of UBI
2031 and beyond	Long-term adaptation and innovation	Continuous improvement	Ongoing adjustments to new roles	New economic models

Note. Content generated by ChatGPT 4 (Anticipatory Governance.AI).

Policy implications and suggested law updates

All of this initial analysis can be considered from the perspective of anticipatory governance, referred to as foresight and AI impact analysis on policy (Mishra, 2023). This exercise's outcome is generating new information and new perspectives to better understand policy implications and regulation updates (Table 10). Therefore, this is the central point of the AI tool. Here, we present the first output provided by the tool. It is essential to mention that, in this part, the user is encouraged to go deeper and request more details, as this is the actionable phase where policymakers need to better understand limitations and opportunities for law updates. The generative nature of the tool provides a difference in how information is presented; however, the propositions are coherent in a first analysis.

Table 10. Comparison of policy implications and suggested law updates.

Electoral Case	Environmental Case	Labor Case
Suggested Law Updates for Brazil's Electoral System:	**Policy Implications and Suggested Legal Updates**	
Amendments to the Brazilian General Data Protection Law (LGPD): Detail: Introduce specific clauses that address the use and protection of voter data in AI-driven systems, including how data is collected, stored, and processed by AI technologies during elections. Purpose: To protect voter privacy and ensure compliance with data ethics as AI becomes more integrated into election systems.	**1. Enhancing the Forest Code (Law no. 12,651, 2012)** Update Needed: Incorporate provisions for AI-driven monitoring tools to ensure compliance and enhance enforcement capabilities. - Implication: Allows for real-time, accurate deforestation detection and automated reporting, leading to quicker governmental response and potential deterrence of illegal activities.	**1. Modernization of Labor Laws** Specific Law: *Consolidação das Leis do Trabalho (CLT)* Update Needed: Incorporate specific provisions for AI-driven automation, remote work, and data privacy in employment settings. Update regulations on telework to reflect new work modalities and enhance protection for gig economy workers.
Updates to the Electoral Code: Detail: Revise the code to include provisions for the use of blockchain technology and AI in voting processes, ensuring these technologies meet transparency and security standards. Purpose: To legally enable the adoption of blockchain and AI, setting standards for technology providers and electoral authorities.	**2. Amending the National Policy on Climate Change (Law no. 12,187, 2009)** Update Needed: Include AI as a central tool in climate modeling and risk assessment, specifying funding and development of AI-driven climate adaptation strategies. - Implication: Enhances the country's ability to predict climate events accurately and implement more effective adaptation measures, potentially reducing the impact of climate-related disasters.	**2. Data Protection in Employment** Specific Law: *Lei Geral de Proteção de Dados (LGPD)* Update Needed: Strengthen regulations regarding the use of personal and biometric data by employers, particularly with AI systems for monitoring and performance evaluations. Ensure that AI compliance aligns with global standards on data ethics and privacy.
Amendments to Cybersecurity Regulation: Detail: Develop stricter Cybersecurity guidelines and requirements for electronic voting systems, especially those using AI for data processing and fraud detection. Purpose: To bolster the security framework around electronic voting, preventing cyber threats and ensuring the integrity of election results.	**3. Revising the National Solid Waste Policy (Law no. 12,305, 2010)** Update Needed: Mandate the use of AI in waste management processes, including waste sorting and recycling optimization. - Implication: Improves efficiency and effectiveness of waste management systems, supports recycling industries, and reduces environmental pollution.	**3. Regulation of AI and Automation** New Legislation Needed: Draft and enact new laws specifically targeting the ethical use of AI and robotics in the workplace, including provisions for worker displacement and automation taxes to fund re-skilling programs.
Introduction of a New AI and Elections Act: Detail: Establish a comprehensive framework governing the development, testing, and deployment of AI systems in the electoral context, including ethical guidelines and testing protocols. Purpose: To provide a clear legal basis for using AI in elections, covering expenses from system development to post-election analysis, ensuring that these technologies are used responsibly and transparently.	**4. Updating the Water Law (Law no 9,433, 1997)** Update Needed: Integrate AI technologies into the management of water resources, specifically for predictive analytics on water usage and quality. - Implication: Provides a more sustainable approach to water management, ensuring better allocation and preservation of water resources across different sectors.	**4. Social Security Reform** Specific Law: Existing social security regulations Update Needed: Reformulate social security laws to accommodate non-traditional employment structures influenced by AI, such as project-based work and freelance engagements. Explore the feasibility of schemes like universal basic income (UBI) under the disruptive scenario.

(continue)

Table 10. Comparison of policy implications and suggested law updates. (continuation)

Electoral Case	Environmental Case	Labor Case
Revision of Funding and Campaign Finance Laws: Detail: Update regulations to address the use of AI in political campaigning, including transparency in AI-driven political advertising and the use of data analytics for campaign strategies. Purpose: To maintain fairness in electoral campaigning and prevent the misuse of AI under undue influence or voter manipulation.	5. Modernizing the Environmental Crimes Law (Law no. 9,605, 1998) Update Needed: Define and standardize the use of AI in detecting and documenting environmental crimes. - Implication: Strengthens enforcement mechanisms by using AI to ensure comprehensive coverage of environmental violations, leading to higher conviction rates.	5. Educational and Skills Development Update Needed: Evaluate laws and policies supporting life-long education and technology-focused curricula in educational programs to prepare the workforce for a technology-driven economy. Implement nationwide re-skilling initiatives aligned with anticipated industry demands.
		6. Regulatory Framework for Gig Economy New Legislation Needed: Develop comprehensive regulations that ensure fair labor practices within the gig economy impacted by AI-driven platforms. This includes clear definitions of employment status, rights, and benefits for gig workers.

Note. Content generated by ChatGPT 4 (Anticipatory Governance.AI).

Policy AI readiness

Based on the previous context, GPT calculates a policy AI readiness score to establish a more objective perspective. Despite creating a coherent metric, AI created different criteria for each case, considering risk level. The initial instruction for GPT was to keep the requirements open, aiming to create a score ranging from 0% to 100% based on the previous analysis. The Electoral Case received a readiness score of 53%, the Environmental Case 65%, and the Labor Case 59% (Table 11).

Table 11. Comparison of policy AI readiness.

Electoral Case	Environmental Case	Labor Case
Calculation of Readiness Score: Technology Adoption: 10/20 Reasoning: Some use of technology, like electronic voting machines, but limited AI integration.	Let's calculate the readiness and risk scores based on hypothetical yet plausible evaluations for Brazil: • Legislative Framework: 12/20 - Some laws are adaptable, but many require significant updates.	Let's now create a score to measure the readiness of policy alignment with emerging AI technologies in Brazil. This scoring will assess how prepared Brazil's current legislative framework is to handle the impacts od AI on the labor market. We'll consider existing laws, proposed updates, technological penetration, and the overall regulatory environment.
Legislative Framework: 8/20 Reasoning: Existing laws cover basics but lack specific provisions for newer technologies like AI and blockchain.	• Technological Adoption: 15/20 - Moderate to high adoption in some areas, particularly in monitoring deforestation and climate modeling.	Criteria for Readiness Scoring: - Existence of Relevant Laws: Availability of laws that directly address issues brought about by AI.
Institutional Capability: 12/20 Reasoning: Adequate institutional knowledge but more expertise, specifically in AI technologies.	• Stakeholder Engagement: 14/20 - Good engagement among government and NGOs, though private sector participation varies.	- Flexibility of Current Laws: Existing laws can adapt to new technologies without significant amendments.
Security and Integrity: 9/20 Reasoning: Moderate security measures are in place, but enhanced threats from AI need more robust defenses.	• Funding and Resources: 13/20 - Reasonable funding, especially with international support, but could be higher.	- Government Proactiveness: Level of government engagement in understanding and integrating AI advancements.
Public Trust and Engagement: 14/20 Reasoning: There is generally high trust in the electoral system, though concerns about technology use remain.	• Infrastructure: 11/20 - Infrastructure is developing, with urban areas better equipped than rural ones.	- Public and Corporate Awareness: Awareness and preparedness of companies and the public towards AI impacts.
Overall Readiness Score: (10+8+12+9+14) / 100} = 53%	Overall Readiness Score: Readiness Score= 12+15+14+13+11100x 100% = 65%	- Infrastructure and Support Systems: Availability of technological infrastructure and support systems for AI integration and education.
		Scoring: Each criterion will be rated on a scale from 0 to 20, where 0 indicates no readiness, and 20 indicates full readiness. The total score will be expressed as a percentage, with 100% representing complete readiness. Existence of Relevant Laws: 12/20 (Some laws like LGPD and labor reforms are in place, but more specific AI regulations are needed.)

Note. Content generated by ChatGPT 4 (Anticipatory Governance.AI).

Strategic questions to policymakers

To provide a more reflective than prescriptive tool, GPT concludes by presenting a set of actionable questions that policymakers should focus on to encourage deep reflection and actionable changes (Table 12).

Table 12. Comparison of questions to policymakers.

Electoral Case	Environmental Case	Labor Case
How can AI be safely and effectively integrated into the electoral system to enhance security and efficiency without compromising voter privacy?	How can the current legislative framework be quickly and effectively updated to integrate AI technologies in environmental management?	How can current labor laws (CLT) be updated to effectively address the implications of AI and automation in various sectors?
What legislative updates are necessary to accommodate and regulate emerging technologies like AI and blockchain in elections?	What strategies should be developed to enhance stakeholder engagement across all sectors involved in environmental governance?	What policies and frameworks are needed to regulate AI use in the workplace, particularly concerning data privacy and ethical standards?
How can Brazil ensure all electoral stakeholders have the knowledge and resources to handle AI-driven electoral processes?	What are the best practices for ensuring equitable and ethical use of AI in environmental management?	How can the government support and facilitate the reskilling and upskilling of the workforce to mitigate job displacement due to AI and automation?
What strategies should be adopted to maintain and enhance public trust in an AI-integrated electoral system?	How can funding and resources be optimized to support adopting and implementing AI technologies in environmental management?	What measures should be taken to ensure equitable access to the benefits of AI-driven labor markets, especially for marginalized and vulnerable groups?
How can the impact of AI on election security be continuously monitored and managed to prevent and respond to cyber threats?	What infrastructural improvements are necessary to support the widespread adoption of AI technologies in environmental practices?	How should Brazil adapt its social security systems to accommodate new forms of employment generated by AI, such as gig and remote work?
What ethical guidelines should be established to govern the use of AI in political campaigning and voter analysis?	How should Brazil prepare for the potential socioeconomic impacts of AI-driven changes in environmental management?	What are the potential economic impacts of AI on the labor market, and how can policy adjustments preemptively address these changes?
	What measures should be implemented to continuously monitor and evaluate the effectiveness of AI technologies in achieving environmental goals?	How can public-private partnerships be leveraged to foster innovation while ensuring that labor market transitions due to AI are smooth and inclusive?

Note. Content generated by ChatGPT 4 (Anticipatory Governance.AI).

Text mining over the discussion

To understand the overall development, GPT is tasked with analyzing the frequency of words to better understand the context of the conversation (Table 13). This can assess the case's overall coherence.

Table 13. Comparison of text mining over the discussion.

Electoral Case	Environmental Case	Labor Case
AI (Artificial Intelligence) - 28 times	AI: 23 occurrences	AI: 27 times
Elections - 24 times	Environmental: 17 occurrences	Labor: 25 times
Technology - 20 times	Brazil: 10 occurrences	Technologies: 18 times
Security - 18 times	Technologies: 16 occurrences	Laws: 16 times
Legislation - 17 times	Management: 12 occurrences	Regulations: 14 times
Blockchain - 14 times	Policy: 15 occurrences	Impact: 13 times
Cybersecurity - 12 times	Legislative: 9 occurrences	Brazil: 12 times
Voting - 11 times	Government: 8 occurrences	Work: 11 times
Policy - 10 times	Impact: 8 occurrences	Policy: 10 times
Integration - 8 times	Stakeholders: 7 occurrences	Automation: 9 times

Note. Content generated by ChatGPT 4 (Anticipatory Governance.AI).

OVERALL ANALYSIS OF THE PROTOTYPE AND PRACTICAL IMPLICATIONS FOR POLICY DESIGN

In the previous section, we described the GPT outputs. Then, we adopted, for this study, five criteria to analyze the results and better understand the level of self-adoption of this AI across different levels of assistance, including user autonomy without experts, users with experts, and support to experts.

(a) The first criterion is the existence of model collapse (Dohmatob et al., 2024; Shumailov et al., 2023), which means signs of degenerative response in the model that can compromise the output's credibility (none, low, medium, high);

(b) The second criterion is AI hallucinations (Athaluri et al., 2023; Emsley, 2023; Salvagno et al., 2023), where the AI generates incorrect output that cannot be supported by cross-checking (none, low, medium, high);

(c) The third criterion is AI bias (Mikalef et al., 2022), which involves identifying traces of algorithmic discrimination or other specific biases (none, low, medium, high);

(d) The fourth criterion is generative interference, where the generative nature of the AI provides differences in the response structure, compromising the tool's goals (none, low, medium, high);

(e) The fifth criterion is the credibility and usefulness of the information, considering a qualitative expert analysis.

Based on the overall analysis, it is observed that, due to AI limitations, the prototype is more suitable for assisting AG experts rather than automating AG for non-experts, as a way to mitigate risks from halluci-

nations, bias, and generative interference, adding specialized human knowledge to the process. That means technology, at this moment, can assist foresight and policy design experts to collaborate with policymakers in collective reasoning and intelligence. In the three scenarios, the real limitations of current Brazilian policies in interaction with AI became evident, which can provide a sense of alert and urgency for change. In order to promote necessary changes, AI provided specific points regarding specific legislation that, guided by the proposed questions, can foster the necessary debate to reach consensus on legal frameworks (Table 14).

Table 14. Overall analysis.

Stage of the Agent	Model Collapse	AI Hallucinations	AI Bias	Generative Interference	Credibility and Usefulness (comment)
1) Stakeholders Involved in the System	low	low	medium	low	Responses are coherent but can be limited and biased.
2) Main Classes of AI Technologies (Research-Level)	medium	high	medium	low	There are non-existent articles in the Labor Case and some biases regarding AI for voter control, such as 'voter behavior analysis.'
3) AI Innovations	low	medium	high	high	Responses centered in the USA system, incapable of contextualizing.
4) Regulations Affected By AI	low	low	low	medium	There is no mention of the laws discussed in the parliament.
5) Foresight Scenarios	low	medium	low	medium	Plausibility and probability are coherent. As an imaginative vision of the future, some outputs cannot be checked.
6) Roadmapping	low	low	low	high	Despite differences in the structure, the output is a valid roadmap.
7) Policy Implications and Suggested Law Updates	low	low	medium	high	Although all suggestions make sense in context, they are generally limited to adding elements to existing laws.
8) Policy AI Readiness	low	low	medium	low	Although the parameter is centered in the USA, there is consistency and transparency in the measures.
9) Questions To Policymakers	low	low	medium	low	The questions are coherent and valid but quite general, as they can be applied in any context.
10) Text Mining Over the Discussion	low	low	medium	low	Adequate account but reveals a lack of context and understanding.
Overall Analysis	almost everything low	predominantly low	predominantly medium	predominantly medium	The prototype is more suitable to assist experts rather than automating AG, both foresight and policy design, for non-experts.

Note. Elaborated by the authors.

Suggested prototype improvements and reflects on ai agency

This AI LLM was designed to serve as an agent to support AG in both foresight and policy design. The initial instructions for the LLM aimed to narrow its choices, making the responses more precise. These instructions included:

(a) Specific contextualization considering its role, function, and expected behavior, acting as a specific set of experts;

(b) Specific contextualization about the context, including country and domain;

(c) Specific stages to be followed;

(d) Specific instructions to validate each stage of the adopted AG flow with a human;

(e) Provision of objective measures at the end, based on the overall reasoning.

Despite including this contextualization and these features, the results — comparing the cases — pointed out specific behaviors that suggest improvements in GPT instruction:

(a) Providing feedback to users about the instruction and assumed setup: for each step, the AI reports what it understood, what will be developed, and critical concepts. This can help users identify if there is a mistake in how the AI is considering a given idea, which can lead to significant differences in the results;

(b) Reinforcing the context at each step: for instance, in the AI Innovation Case, startups were analyzed in the USA context, not the Brazilian one. Each step should reinforce the country and policy domain;

(c) AI policy readiness metric: despite attempts to keep it open, a specific set of criteria must be credited and previously disclosed to the user;

(d) Reminders: prompting the user to check the information (such as articles) and validate facts;

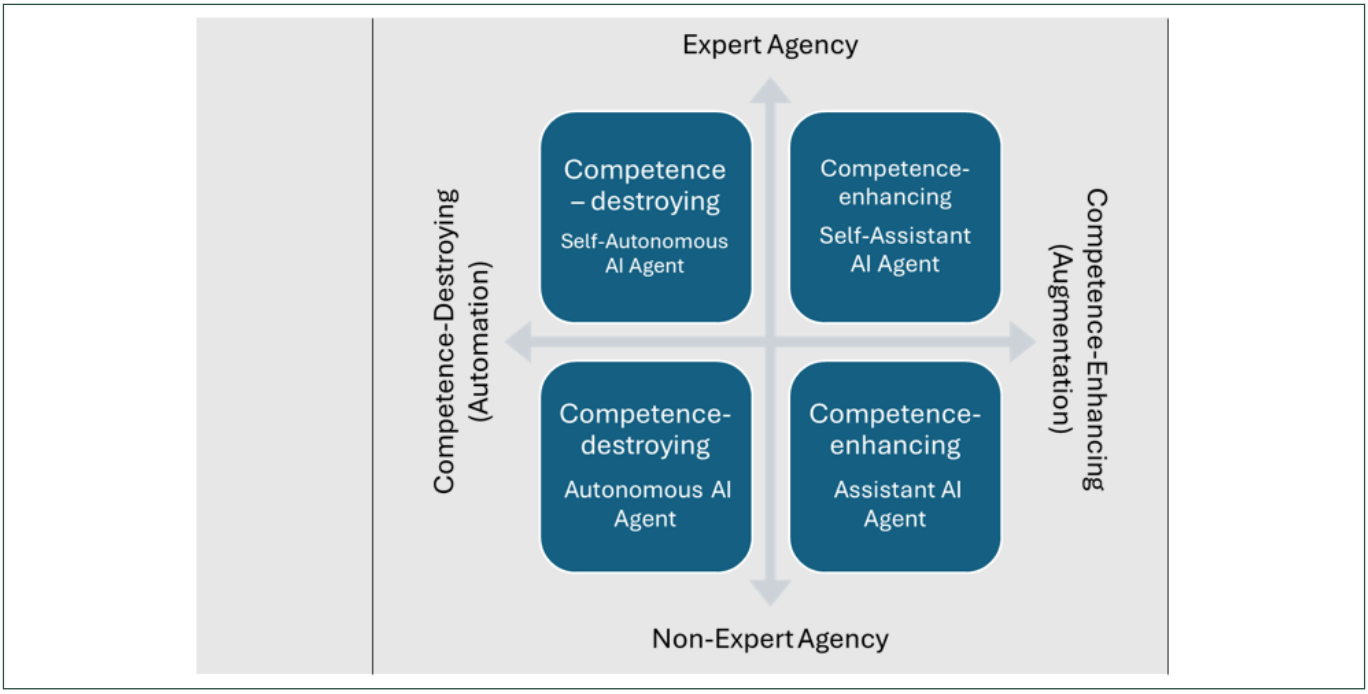
(e) Increasing RAG capabilities with anchors and counterfactual explanations, such as including a feature to 'explain the reasoning for this answer' and citing sources.

One key conclusion from this experiment is the necessity of including experts in designing and evaluating

such AI tools. In GPT instruction, there is a body of embedded knowledge, and this article demonstrates how even the adoption of best practices in GPT instruction by experts in the field can still result in a tool producing inaccurate results due to the nature of LLM models. Therefore, before deploying such tools in actual cases, the adoption of the best frameworks for design and assessment should be followed, which is not the case today.

Another aspect of this exercise is understanding that LLM generative AI primarily consists of statistical models that infer the most likely response based on inputs from past training data. LLMs are usually configured to respond according to the expected distribution or to reach the average 'mean' or standard of a given subject. However, in future scenarios, this might not be appropriate. Outliers, wild cards, weak signals, and intuition for highly creative insights are also essential in dealing with long-range scenario development. Therefore, LLMs provide a somewhat standardized set of responses, which can accelerate insights and narrow perspectives if taken as definitive answers. Another relevant aspect is the performance of this tool, which is impacted by the high number of tokens consumed due to the significant amount of data necessary to create knowledge about foresight, policy design, and assessment in a given context.

That means the prototype is more suitable for assisting experts rather than automating foresight and policy design AG for non-experts. These results align with the typology of Paschen et al. (2020) on AI innovation. Based on the competence and agency rationale, we propose an update of the typology for AI agents (Figure 5). We observed that, at this moment, this AI agent is more suitable as a self-assistant AI agent, depending upon exerted agency to increase AI trustability. This typology is essential for users to better understand when a given agent can perform its function autonomously or in an augmented fashion.

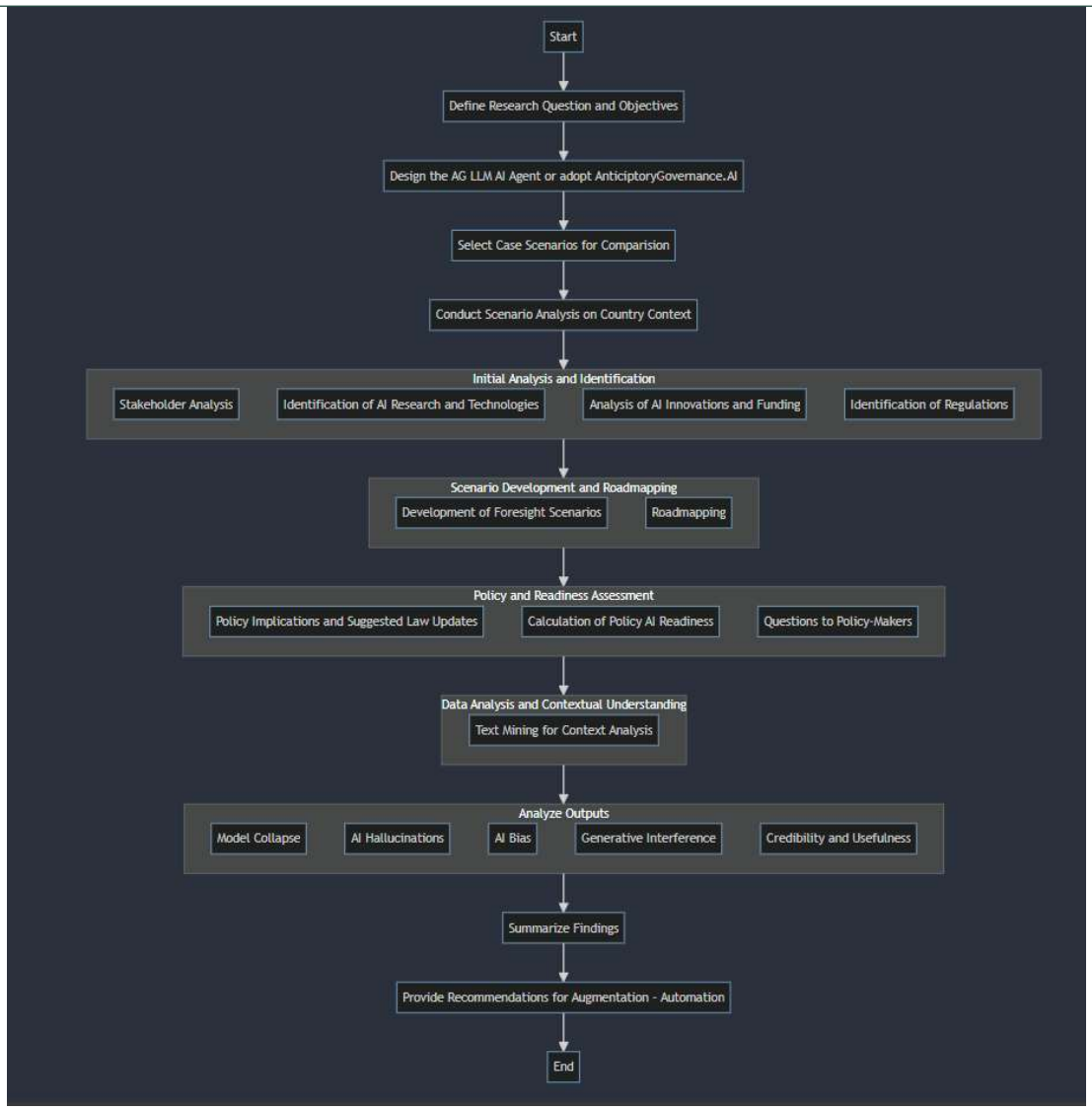


Source: Elaborated by the authors.

Figure 5. A typology for AI agents and their effects on competencies and expert agencies.

AI assessment method protocol emerged

This experiment provided the foundation for an assessment protocol for designing and validating the behavior of generative AI agents, specifically in foresight and roadmapping for policy analysis or integration into an anticipatory governance system. It is recommended that researchers validate these agents with different technologies and scenarios to increase trust in the adoption of LLM-based agents. The following steps are proposed to provide more clarity and structure for the LLM function, and the analysis of the outputs based on the five categories highlights points that expert users should focus on to better understand the AI's benefits and limitations (Figure 18, Table 18). This protocol outlines the steps for assessing large language model (LLM) AI agents in the context of anticipatory governance. The objective is to comprehensively analyze and evaluate LLM agents' effectiveness and readiness for governance applications, particularly in predicting and preparing for future scenarios. Eleven steps guide researchers through AG LLM validation (Figure 6, Table 15).



Source: Elaborated by the authors.

Figure 6. Assessment method for generative AI technology in foresight and policy design in public management.

Table 15. Assessment method: Stages and actions.

Stages	Actions
1. Define the Research Question and Objectives	Formulate the primary research question to guide the assessment. Define specific objectives that align with the research question.
2. Design the AG LLM AI Agent or Adopt AnticipatoryGovernance.AI	Develop and configure the AG LLM AI agent tailored to the research requirements. Ensure the design accommodates specific needs of Anticipatory Governance.
3. Select Case Scenarios for Comparison	Identify and select diverse case scenarios for comparison. Consider AG themes in Technology, Environment, Law, Health, etc. Ensure scenarios represent different contexts and challenges.
4. Conduct Scenario Analysis on Country's Context	Run selected scenarios within different countries' contexts. Instruct the agent with specific elements related to the country and theme.
5. Initial Analysis and Identification	Stakeholder Analysis: Identify critical stakeholders involved in each scenario. Identify existing AI research, technologies, innovations, funding, and regulations. Assess their significance and trustability.
6. Scenario Development and Roadmapping	Develop foresight scenarios based on initial analyses. Roadmapping: Create detailed roadmaps with key steps and milestones.
7. Policy and Readiness Assessment	Evaluate policy implications and suggest necessary updates. Assess AI readiness of current policies. Formulate questions to policymakers for input on readiness.
8. Data Analysis and Contextual Understanding	Use text mining techniques to extract and analyze contextual data. Cross-validate content for deeper context analysis.
9. Analyze Outputs from Previous Steps	Investigate model collapse, AI hallucinations, and biases. Assess generative interference, credibility, and usefulness of outputs.
10. Summarize Findings	Compile and summarize key findings from analyses. Include different perspectives if available.
11. Provide Recommendations for Augmentation and Automation	Develop recommendations for augmenting and automating governance processes. Focus on enhancing anticipatory governance capabilities.

Note. Elaborated by the authors.

This study aimed to investigate the perceptions of foresight and AI experts regarding the use of a generative AI (GEN_AI) assistant in anticipatory governance (AG), with a focus on refining a prototype tool and establishing an assessment protocol. The experts' perspectives validated the protocol's relevance and utility in addressing the complexities associated with integrating GEN_AI into AG processes. They recognized the potential of GEN_AI to enhance key AG activities, such as data collection, scenario development, and policy recommendations, while also identifying critical limitations inherent to current large language models (LLMs). These limitations include issues such as AI hallucinations, biases, lack of contextual understanding, and challenges related to model interpretability, which are particularly concerning in high-stakes domains like public policy. The proposed protocol offers a structured framework for evaluating GEN_AI tools, emphasizing the necessity of 'human approaches' — referred to here as human curacity — to mitigate risks and enhance the reliability of outputs. Experts acknowledged the protocol as a significant theoretical contribution, particularly due to its emphasis on explainable AI (xAI) and its applicability across diverse stakeholder roles in anticipatory governance. However, they emphasized the importance of continuous refinement and replication of assessments to address emerging challenges and ensure the sustainable adoption of GEN_AI in AG practices. The need for adaptation to each strategic monitoring process was also highlighted, considering the intrinsic requirement of alignment with each context and environmental observation cycle. Overall, the protocol was regarded as a crucial step toward bridging the gap between the technological capabilities of GEN_AI and the complex, context-sensitive demands of anticipatory governance.

CONCLUSIONS

The goal of this study was to address the research question *"How is the perception of foresight and AI experts over a generative AI assistant for anticipatory governance?"* by analyzing the perception of foresight and AI experts over a GEN_AI assistant for AG, in three cases, to improve a prototype tool and emerge an assessment protocol for GEN_AI in anticipatory governance. Further understanding of a GEN_AI behavior is crucial before scaling a service for non-experts in a context of public policy, mainly due to elements inherent to current LLMs such as model collapse, AI hallucinations, AI bias, and how the generative process can create interferences in human reasoning, which in some contexts are more favorable than others. Moreover, the xAI lens and institutional adoption in public management drive the need for auditing such tools from a scientific perspective.

Implications for AG theory and generative AI analysis

AG is a complex process that encompasses the domains of science, technology, and innovation, understanding emerging elements, designing scenarios, interpreting potential impacts, and relating to policy regulation assessment and implication — with the participation of several stakeholders in complex themes that range from geopolitical tensions (such as elections), climate change (such as environment and industry), and social inequality (such as AI technology impacts). Therefore, several activities, from foresight to policy design, can be considered complex, including framing, scanning, data collection, data organization, data sensemaking, scenario building, and scenario impact interpretation within a specific stakeholder system. Moreover, interpreting the legal framework and understanding possible improvements are also tasks that require a significant amount of expert knowledge. This means that policy enrichment by foresight needs to combine human-machine approaches to deal with the challenge of time in emerging technologies. Given the rapid pace of technological changes, AG processes that take months will be more vulnerable to address these issues. This study contributes to the field of applied anticipatory governance and the management of knowledge generated through an end-to-end process, from foresight to recommendations for policy design.

However, although LLMs do have great potential for AG support, they also have current limitations. Scientific assessments can help further advancement, especially to better understand tool capabilities for experts and non-experts. These limitations can be understood (Bender et al., 2021; Devlin et al., 2019; Shen et al., 2023) in terms of some degree of hallucination of information (information that seems credible but is incorrect), bias and lack of fairness (embedding bias that is present in the source material), lack of context understanding (since LLMs are probability models, they lack intrinsic understanding or consciousness necessary to interpret nuanced or ethical considerations that require high- and low-context reading), dependency on data quality and tuning (inaccuracies or exploitations of training data or tuning parameters can lead to several consequences in model outputs), security (adversarial attacks that can manipulate outcomes), and model interpretability (LLMs based on deep learning often operate as black boxes). These elements were observed at some level in the experiment. Limitations can be addressed over time, considering investments in scientific research and industry development. While these limitations are visible, LLM generative AI accelerates foresight and policy design data collection, sce-

nario development, impact analysis, and policy recommendations. This article, therefore, aims to address the gap in understanding, specifically focusing on an AG AI tool. Understanding how this tool can better serve the different stakeholders involved in this process is essential. To achieve this, we examine the viewpoints of both experts and non-expert users. That's why replicating these assessments is needed, and the protocol provided is a significant theoretical contribution. xAI is also considered for generative AI applications in anticipatory governance to understand how this tool can better serve the different stakeholders involved in this process. To achieve this, we examine the viewpoints of both experts and non-expert users. That's why replication of these assessments is needed, and the protocol provided is an important theoretical contribution, also considering xAI for generative AI applications in anticipatory governance.

At this stage, experts in AG can face more augmentation capabilities offered by adopting such AI agents because they can address the interpretability of these limitations, considering both prompting and fact-checking. These activities are critical for utilizing the outputs of an LLM as input for creativity, analysis, or decision-making. However, more limitations arise when adopting non-experts, especially when expecting some level of automation or an end-to-end conclusive and definite answer. Central to this is the notion discussed by Bender et al. (2021), where humans tend to attribute meaning where there is none, misleading themselves or others when taking synthetic text as meaningful. Both foresight, which deals with future-oriented information, and policymaking, which deals with society's beliefs and information, involve different levels of meaning, in which synthetic data can face limitations in its mathematical process throughout the end-to-end process.

Therefore, from the perspective of stakeholder integration, an LLM assistant can, observing this experiment, greatly benefit from a human-in-the-loop perspective; however, the level of trustworthiness of the initial response, mainly when a non-expert conducts the AI agent, is questionable. These results must be considered in the context of AG theory and processes and their relationship with generative AI. While generative AI can automate simpler processes, such as providing essential information in product and service consumption conversations, its application in anticipatory governance — especially given the complex concepts involved, ranging from technology to law — requires additional perspectives.

Implications for research, public management, and AG practice

The assessment protocol can foster a structured method for researchers replicating assessments in other generative AI agents, increasing the trustability of this technology. From a more specific perspective, AI adoption needs to be considered regarding learning, automation, augmentation, and innovation (Hassani et al., 2020; Srinivasan, 2022; Tyson & Zysman, 2022; Verma & Singh, 2022). Therefore, when considering an AG AI assistant, some recommendations can be made:

(a) Learning: since learning involves exploration and willingness to trial and error, AG can be an introductory tool for researchers and policymakers to explore topics. However, it is essential to be aware of its limitations;

(b) Automation: AG cannot fully automate end-to-end foresight or policymaker activities. As observed, human intervention is required at every stage of the process, and the human in the loop is necessary in each stage to refine reasoning, understand context, and validate outcomes;

(c) Augmentation: in an actual anticipatory governance process, properly conducted, the AI can enhance capabilities for knowledge discovery and understanding of the problem. It does not provide the final answer but contributes to its construction through the integration of human capabilities.

(d) Innovation: to better understand new tasks, it is essential to consider the typical roles in an anticipatory governance workshop or process.

A given AG can include a set of roles such as facilitator/animator, subject matter expert, foresight analyst, policymakers, stakeholders, risk managers, strategic planners, innovation officers, legal advisors, and technology specialists. AG deals with a high level of complexity, from understanding new science and technology to scenario creation, impact interpretation, and formulating requisites for new policy designs or updates. Therefore, this kind of LLM assistant can aid in learning, automation, and augmentation across different levels within these roles. These are highly contextual, considering factors such as the problem, team, LLM adopted, and other available resources. This perspective can guide AI adoption for public management and AG practice, expanding the role of technology in collaboration with humans among AG processes.

Study limitations and futures perspectives

This study was based on version 4 of OpenAI ChatGPT 4 (2024), and therefore, the performance achieved can change or is expected to change in the next few years. However, the basic probabilistic foundation of how LLMs through transformers operate and their known

limitations will remain, making it imperative for stakeholders in the anticipatory governance process to understand the current extent of these limitations and make more sustainable use of such technology for real-case scenarios. This leads to future research perspectives on how AG stakeholders, in different roles, can effectively utilize GEN_AI tools — especially future AutoAI and multi-agent systems — in real-case scenarios, and how human agency and AI agency collaborate and interact over time. Differences between human and machine outputs are also directions for research. Considering the global geopolitical, climate, and social challenges and the increasing need for AI technologies by stakeholders to better understand and solve complex problems, the integration of AG and AI adoption will provide essential solutions for humanity's problem-solving efforts.

REFERENCES

Andrade, N. N. G., & Kontschieder, V. (2021). AI impact assessment: A policy prototyping experiment. *SSRN Electronic Journal*. https://dx.doi.org/10.2139/ssrn.3772500

Athaluri, S. A., Manthena, S. V., Kesapragada, V. S. R. K. M., Yarlagadda, V., Dave, T., & Duddumpudi, R. T. S. (2023). Exploring the boundaries of reality: Investigating the phenomenon of artificial intelligence hallucination in scientific writing through ChatGPT references. *Cureus, 15*(4), e37432. http://doi.org/10.7759/cureus.37432

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In FAccT 2021 - *Proceedings* of the 2021 ACM Conference on Fairness, Accountability, and Transparency (pp. 610-623). https://doi.org/10.1145/3442188.3445922

Bevolo, M., & Amati, F. (2020). The potential role of AI in anticipating futures from a design process perspective: From the reflexive description of "design" to a discussion of influences by the inclusion of AI in the futures research process. *World Futures Review*, *12*(2), 198-218. https://doi.org/10.1177/1946756719897402

Boobier, T. (2018). The impact of AI on professions. In T. Boobier (Ed.), *Advanced Analytics and AI: Impact, Implementation, and the Future of Work* (pp. 117-154). Wiley.

Cao, L. (2022). Beyond AutoML: Mindful and actionable AI and AutoAI with mind and action. *IEEE Intelligent Systems*, *37*(5), 6-18. https://doi.org/10.1109/MIS.2022.3207860

Coeckelbergh, M. (2023). Democracy, epistemic agency, and AI: Political epistemology in times of artificial intelligence. *AI and Ethics*, *3*(4), 1341-1350. https://doi.org/10.1007/s43681-022-00239-4

Coeckelbergh, M., & Sætra, H. S. (2023). Climate change and the political pathways of AI: The technocracy-democracy dilemma in light of artificial intelligence and human agency. *Technology in Society*, *75,* 102406. https://doi.org/10.1016/j.techsoc.2023.102406

Complementary Law no. 135 of June 4, 2010. (2010). Lei da Ficha Limpa. *Diário Oficial da União*. https://www.planalto.gov.br/ccivil_03/leis/lcp/lcp135.htm

Decree no. 10,222 of February 5, 2020. (2020). Estratégia Nacional de Segurança Cibernética. *Diário Oficial da União*. https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2020/decreto/d10222.htm

Decree-Law no. 5,452 of May 1, 1943. (1943). Consolidação das Leis do Trabalho (CLT). *Diário Oficial da União*. https://www.planalto.gov.br/ccivil_03/decreto-lei/del5452.htm

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*. https://doi.org/10.48550/arXiv.1810.04805

Dohmatob, E., Feng, Y., Yang, P., Charton, F., & Kempe, J. (2024). A tale of tails: Model collapse as a change of scaling laws. *arXiv preprint arXiv:2402.07043*. https://doi.org/10.48550/arXiv.2402.07043

Emsley, R. (2023). ChatGPT: These are not hallucinations but fabrications and falsifications. *Schizophr, 9*, 52. https://doi.org/10.1038/s41537-023-00379-4

Farrow, E. (2020). Organisational artificial intelligence future scenarios: Futurists insights and implications for the organisational adaptation approach, leader and team. *Journal of Futures Studies*, *24*(3), 1-15. https://jfsdigital.org/wp-content/uploads/2020/03/01-Farrow-Organisational-Artificial-Intelligence-Future-Scenarios-Ed.9.pdf

Fuerth, L. S. (2009). Foresight and anticipatory governance. *Foresight*, *11*(4), 14-32. https://doi.org/10.1108/14636680910982412

Fuerth, L. S. (2012). The project on forward engagement. *Anticipatory Governance: The Project on Forward Engagement*. https://forwardengagement.org/

Georgieff, A., & Hyee, R. (2022). Artificial Intelligence and employment: New cross-country evidence. *Frontiers in Artificial Intelligence, 5*. https://doi.org/10.3389/frai.2022.832736

Geurts, A., Gutknecht, R., Warnke, P., Goetheer, A., Schirrmeister, E., Bakker, B., & Meissner, S. (2022). New perspectives for data-supported foresight: The hybrid AI-expert approach. *Futures & Foresight Science*, *4*(1), e99. https://doi.org/10.1002/ffo2.99

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, *63*(11), 139-144. https://doi.org/10.1145/3422622

Grønsund, T., & Aanestad, M. (2020). Augmenting the algorithm: Emerging human-in-the-loop work configurations. *Journal of Strategic Information Systems*, *29*(2), 101614. https://doi.org/10.1016/j.jsis.2020.101614

Guston, D. H. (2014). Understanding "anticipatory governance." *Social Studies of Science, 44*(2), 218-242. https://doi.org/10.1177/0306312713508669

Hassani, H., Silva, E. S., Unger, S., TajMazinani, M., & Mac Feely, S. (2020). Artificial Intelligence (AI) or Intelligence Augmentation (IA): What is the future? *AI (Switzerland), 1*(2), 143-155. https://doi.org/10.3390/ai1020008

Heo, K., & Seo, Y. (2021). Anticipatory governance for newcomers: Lessons learned from the UK, the Netherlands, Finland, and Korea. *European Journal of Futures Research*, *9*(1). https://doi.org/10.1186/s40309-021-00179-y

Hussain, M., Tapinos, E., & Knight, L. (2017). Scenario-driven roadmapping for technology foresight. *Technological Forecasting and Social Change*, *124,* 160-177. https://doi.org/10.1016/j.techfore.2017.05.005

Islam, S. R., Eberle, W., Ghafoor, S. K., & Ahmed, M. (2021). Explainable Artificial Intelligence approaches: A survey. arXiv preprint. https://arxiv.org/abs/2101.09429

Jokinen, L., Balcom Raleigh, N. A., & Heikkilä, K. (2023). Futures literacy in collaborative foresight networks: advancing sustainable shipbuilding. *European Journal of Futures Research, 11*(1). https://doi.org/10.1186/s40309-023-00221-1

Kolliarakis, G., & Hermann, I (2020). Towards European Anticipatory Governance for Artificial Intelligence. *DGAP*. https://dgap.org/en/research/publications/towards-european-anticipatory-governance-artificial-intelligence

Kunseler, E. M., Tuinstra, W., Vasileiadou, E., & Petersen, A. C. (2015). The reflective futures practitioner: Balancing salience, credibility and legitimacy in generating foresight knowledge with stakeholders. *Futures, 66,* 1-12. https://doi.org/10.1016/j.futures.2014.10.006

Laux, J., Wachter, S., & Mittelstadt, B. (2024). Trustworthy artificial intelligence and the European Union AI act: On the conflation of trustworthiness and acceptability of risk. *Regulation & Governance, 18*(1), 3-32. https://doi.org/10.1111/rego.12512

Law no. 4,737 of July 15, 1965. (1965). Código Eleitoral. *Diário Oficial da União*. https://www.planalto.gov.br/ccivil_03/leis/l4737.htm

Law no. 9,096 of September 19, 1995. (1995). Lei dos Partidos Políticos. *Diário Oficial da União*. https://www.planalto.gov.br/ccivil_03/leis/l9096.htm

Law no. 9,433 of January 8, 1997. (1997). Lei das Águas. *Diário Oficial da União*. https://www.planalto.gov.br/ccivil_03/leis/l9433.htm

Law no. 9,605 of February 12, 1998. (1998). Lei de Crimes Ambientais. *Diário Oficial da União*. https://www.planalto.gov.br/ccivil_03/leis/l9605.htm

Law no. 12,187 of December 29, 2009. (2009). Política Nacional sobre Mudança do Clima. *Diário Oficial da União*. https://www.planalto.gov.br/ccivil_03/_ato2007-2010/2009/lei/l12187.htm

Law no. 12,305 of August 2, 2010. (2010). Política Nacional de Resíduos Sólidos. *Diário Oficial da União*. https://www.planalto.gov.br/ccivil_03/_ato2007-2010/2010/lei/l12305.htm

Law no. 12,651 of May 25, 2012. (2012). Código Florestal. *Diário Oficial da União*. https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2012/lei/l12651.htm

Law no. 12,965 of April 23, 2014. (2014). Marco Civil da Internet. *Diário Oficial da União*. https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/l12965.htm

Law no. 13,467 July 13, 2017. (2017). Reforma Trabalhista. *Diário Oficial da União*. https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2017/lei/l13467.htm

Law no. 13,709 of August 14, 2018. (2018). Lei Geral de Proteção de Dados Pessoais (LGPD). *Diário Oficial da União*. https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm

Longo, L., Brcic, M., Cabitza, F., Choi, J., Confalonieri, R., Del Ser, J., Guidotti, R., Hayashi, Y., Herrera, F., Holzinger, A., Jiang, R., Khosravi, H., Lecue, F., Malgieri, A. P., Samek, W., Schneider, J., Speith, T., & Stumpf, S. (2024). Explainable Artificial Intelligence (XAI) 2.0: A manifesto of open challenges and interdisciplinary research directions. *Information Fusion, 106*, 102301. https://doi.org/10.1016/j.inffus.2024.102301

Maffei, S., Leoni, F., & Villari, B. (2020). Data-driven anticipatory governance. Emerging scenarios in data for policy practices. *Policy Design and Practice*, *3*(2), 123-134. https://doi.org/10.1080/25741292.2020.1763896

Mikalef, P., Conboy, K., Eriksson Lundström, J., & Popovič, A. (2022). Thinking responsibly about responsible AI and 'the dark side' of AI. *European Journal of Information Systems*, *31*(3), 257-268. https://doi.org/10.1080/0960085X.2022.2026621

Mishra, A. (2023). AI Alignment and social choice: Fundamental limitations and policy implications. *arXiv preprint*. https://doi.org/10.48550/arXiv.2310.16048

Muiderman, K., Zurek, M., Vervoort, J., Gupta, A., Hasnain, S., & Driessen, P. (2022). The anticipatory governance of sustainability transformations: Hybrid approaches and dominant perspectives. *Global Environmental Change*, *73*, 102452. https://doi.org/10.1016/j.gloenvcha.2021.102452

Myllyoja, J., Rilla, N., & Lima-Toivanen, M. (2022). Strengthening futures-oriented agenda for building innovation ecosystems. *European Journal of Futures Research*, *10*(1). https://doi.org/10.1186/s40309-022-00211-9

Nagahisarchoghaei, M., Nur, N., Cummins, L., Nur, N., Karimi, M. M., Nandanwar, S., Bhattacharyya, S., & Rahimi, S. (2023). An empirical survey on explainable AI technologies: Recent trends, cases, and categories from technical and application perspectives. *Electronics*, *12*(5), 1092. https://doi.org/10.3390/electronics12051092

Oehmichen, J., Schult, A., & Dong, J. Q. (2023). Successfully organizing AI innovation through collaboration with startups. *Mis Quarterly Executive*, *22*(1). https://aisel.aisnet.org/misqe/vol22/iss1/4

Ohta, H. (2020). The analysis of Japan's energy and climate policy from the aspect of anticipatory governance. *Energies*, *13*(19), 5153. https://doi.org/10.3390/en13195153

OpenAI. (2024). *ChatGPT 4* . https://openai.com/pt-BR/

Panizzon, M., & Janissek-Muniz, R. (2025). Theoretical dimensions for integrating research on anticipatory governance, scientific foresight, and sustainable S&T public policy design. *Technology in Society*, *80*, 102758. https://doi.org/10.1016/j.techsoc.2024.102758

Paschen, J., Wilson, M., & Ferreira, J. J. (2020). Collaborative intelligence: How human and artificial intelligence create value along the B2B sales funnel. *Business Horizons*, *63*(3), 403-414. https://doi.org/10.1016/j.bushor.2020.01.003

Pflanzer, M., Dubljević, V., Bauer, W. A., Orcutt, D., List, G., & Singh, M. P. (2023, August 1). Embedding AI in society: Ethics, policy, governance, and impacts. *AI and Society*. Springer Science and Business Media Deutschland GmbH. _

Puaschunder, J. M. (2019). On Artificial Intelligence's razor's edge: On the future of democracy and society in the artificial age. *Journal of Economics and Business*, *2*(1), 100-119. https://dx.doi.org/10.2139/ssrn.3297348

Radanliev, P., & Roure, D. (2023). Review of the state of the art in autonomous artificial intelligence. *AI and Ethics*, *3*(2), 497-504. https://doi.org/10.1007/s43681-022-00176-2

Rojas, A., & Tuomi, A. (2022). Reimagining the sustainable social development of AI for the service sector: the role of startups. *Journal of Ethics in Entrepreneurship and Technology*, *2*(1), 39-54. https://doi.org/10.1108/JEET-03-2022-0005

Saeed, W., & Omlin, C. (2023). Explainable AI (XAI): A systematic meta-survey of current challenges and future opportunities. *Knowledge-Based Systems*, *263*, 110273. https://doi.org/10.1016/j.knosys.2023.110273

Salvagno, M., Taccone, F. S., & Gerli, A. G. (2023). Artificial intelligence hallucinations. *Critical Care*, *27*(1), 180. https://doi.org/10.1186/s13054-023-04473-y

Shen, Y., Heacock, L., Elias, J., Hentel, K. D., Reig, B., Shih, G., & Moy, L. (2023). ChatGPT and other large language models are double-edged swords. *Radiology*, *307*(2). https://doi.org/10.1148/radiol.230163

Shumailov, I., Shumaylov, Z., Zhao, Y., Gal, Y., Papernot, N., & Anderson, R. (2023). The curse of recursion: Training on generated data makes models forget. *arXiv preprint arXiv:2305.17493*. https://doi.org/10.48550/arXiv.2305.17493

Srinivasan, V. (2022). AI & learning: A preferred future. *Computers and Education: Artificial Intelligence*, *3*, 100062. https://doi.org/10.1016/j.caeai.2022.100062

Störmer, E., Bontoux, L., Krzysztofowicz, M., Florescu, E., Bock, A.-K., & Scapolo, F. (2020). Foresight – Using science and evidence to anticipate and shape the future. In V. Šucha & M. Sienkiewicz (Eds.), *Science for Policy Handbook* (pp. 128-142). Elsevier.

Suresh, H., & Guttag, J. (2021). A framework for understanding sources of harm throughout the machine learning life cycle. In ACM International Conference *Proceeding*. Association for Computing Machinery. https://doi.org/10.1145/3465416.3483305

Theunissen, M., & Browning, J. (2022). Putting explainable AI in context: Institutional explanations for medical AI. *Ethics and Information Technology*, 23. https://doi.org/10.1007/s10676-022-09649-8

Tolan, S., Pesole, A., Martinez-Plumed, F., Fernandez-Macias, E., Hernandez-Orallo, J., & Gomez, E. (2020). Measuring the occupational impact of AI: Tasks, cognitive abilities and AI benchmarks. *Journal of Artificial Intelligence Research*, *71*, 191-236. https://doi.org/10.1613/jair.1.12647

Tyson, L. D., & Zysman, J. (2022). Automation, AI & Work. *Daedalus*, *151*(2), 256-271. https://doi.org/10.1162/daed_a_01914

Van Noordt, C., & Misuraca, G. (2022). Artificial intelligence for the public sector: results of landscaping the use of AI in government across the European Union. *Government Information Quarterly*, *39*(3), 101714. https://doi.org/10.1016/j.giq.2022.101714

Van Woensel, L. (2020). Scientific foresight: Considering the future of science and technology. In *St Antony's Series*. Palgrave Macmillan.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems* (Vol. 2017-Dec). Neural information processing systems foundation.

Verma, S., & Singh, V. (2022). Impact of artificial intelligence-enabled job characteristics and perceived substitution crisis on innovative work behavior of employees from high-tech firms. *Computers in Human Behavior*, *131*, 107215. https://doi.org/10.1016/j.chb.2022.107215

Wright, G., O'Brien, F., Meadows, M., Tapinos, E., & Pyper, N. (2020). Scenario planning and foresight: Advancing theory and improving practice. *Technological Forecasting and Social Change*, *159*, 120220. https://doi.org/10.1016/j.techfore.2020.120220

Yang, W., Wei, Y., Wei, H., Chen, Y., Huang, G., Li, X., Li, R., Yao, N., Wang, X., Gu, X., Amin, M. B., & Kang, B. (2023). Survey on explainable AI: From approaches, limitations, and applications aspects. *Human-Centric Intelligent Systems*, *3*, 161-188. https://doi.org/10.1007/s44230-023-00038-y

Authors

Mateus Panizzon
Universidade Caxias do Sul, Business School
Rua Francisco Getúlio Vargas, n. 1130, Petrópolis, CEP 95070-560, Caxias do Sul, RS, Brazil
Universidade Federal do Rio Grande do Sul, Escola de Administração
Rua Washington Luiz, n. 855, Centro Histórico, CEP 90010-460, Porto Alegre, RS, Brazil
mpanizzo@ucs.br

Raquel Janissek-Muniz
Universidade Federal do Rio Grande do Sul, Escola de Administração
Rua Washington Luiz, n. 855, Centro Histórico, CEP 90010-460, Porto Alegre, RS, Brazil
rjmuniz@ufrgs.br

Natália Marroni Borges
Universidade Federal do Rio Grande do Sul, Escola de Administração
Rua Washington Luiz, n. 855, Centro Histórico, CEP 90010-460, Porto Alegre, RS, Brazil
natalia.marroni@gmail.com

Amanda Cainelli
Universidade Federal do Rio Grande do Sul, Escola de Administração
Rua Washington Luiz, n. 855, Centro Histórico, CEP 90010-460, Porto Alegre, RS, Brazil
amandacainelli@gmail.com

Authors' contributions

1st author: conceptualization (lead), data curation (lead), formal analysis (lead), funding acquisition (supporting), investigation (lead), methodology (lead), resources (lead), software (lead), visualization (lead), writing - original draft (lead).

2nd author: funding acquisition (lead), project administration (lead), supervision (lead), validation (lead), writing - review & editing (equal).

3rd author: data curation (equal), formal analysis (equal), funding acquisition (supporting), investigation (supporting), methodology (equal), validation (equal), writing - review & editing (equal).

4thauthor: data curation (equal), formal analysis (equal), funding acquisition (supporting), investigation (supporting), methodology (equal), validation (equal), writing - review & editing (equal).