AIGOV

Implementing ethical, trustworthy and fair Artificial Intelligence Systems in Public Sector

D2.1 AIGOV Holistic Framework

Version-Status:	V1.0 Final
Date:	31/03/2024
Dissemination level:	PU

The research project was supported by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the **"2nd Call for H.F.R.I. Research Projects to support Faculty Members & Researchers"** (Project Number: 2412).

Deliverable factsheet

Project Number:	2412
Project Acronym:	AIGOV
Project Title:	Implementing ethical, trustworthy and fair Artificial Intelligence Systems in Public Sector
Principal Investigator:	Konstantinos Tarabanis
Scientific Area:	SA9. Management & Economics of Innovations
Scientific Field:	9.4 ICT enabled Innovation, Digitisation and Industrial Renewal
Host Institution:	University of Macedonia
Collaborating Organization:	Region of Central Macedonia, Greece

Deliverable title:	AIGOV Holistic Framework
Deliverable number:	D2.1
Official submission date:	31/03/2024
Actual submission date:	31/03/2024
Author(s):	Areti Karamanou, Dimitrios Zegkinis, Maria Zotou, Evangelos Kalampokis, Konstantinos Tarabanis

Abstract:	This document is the deliverable, entitled D2.1 “AIGOV Holistic Framework”, of the second work package of the AIGOV project. It reports the results of T2.1, T2.2, and T2.3 of WP2 in the form of a Holistic Framework for AI in public administration.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

The following table presents the acronyms used in the deliverable in alphabetical order.

Abbreviation	*Description*
AI	Artificial Intelligence
API	Application Programming Interface
CC	Creative Commons
CSV	Comma-Separated Values
DGA	Data Governance Act
EU	European Union
FAIR	Findable, Accessible, Interoperable, Reusable
GAN	Generative Adversarial Network
GDPR	General Data Protection Regulation
GPT	Generative Pre-trained Transformer
IoT	Internet of Things
KG	Knowledge Graph
KPI	Key Performance Indicator
LLM	Large Language Model
NLP	Natural Language Processing
OGD	Open Government Data
OECD	Organisation for Economic Co-operation and Development
PEFT	Parameter-Efficient Fine-Tuning
RAG	Retrieval-Augmented Generation
RL	Reinforcement Learning
RLHF	Reinforcement Learning from Human Feedback
SHAP	SHapley Additive exPlanations
SQL	Structured Query Language
WP	Work Package

Executive Summary

This deliverable presents the key outcomes of Work Package 2 (WP2) of the AIGOV project, which develops the methodological, ethical, and organisational foundations needed for the trustworthy, fair, and accountable adoption of Artificial Intelligence (AI) in the public sector, with a particular focus on Generative AI (GenAI) and Large Language Models (LLMs).

Building on the ecosystem analysis of WP1, WP2 translates international best practices and emerging European regulatory requirements, including the EU AI Act, the OECD AI Principles, and the EU data and interoperability frameworks, into three practical and future-oriented instruments for public administrations:

1. **AIGOV Government Data Value Cycle** (Task 2.1)

WP2 introduces a renewed seven-step data value cycle that reflects the needs of modern AI systems and the growing importance of unstructured and dynamic data. The cycle covers:

1. Data collection
2. Data curation
3. Data integration and linking
4. Data storage
5. Data dissemination
6. Data usage
7. Data value creation

For each step, the framework outlines requirements, challenges, and guidelines that help governments ensure data quality, privacy, interoperability, and readiness for AI-driven innovation.

2. **AIGOV Framework for Trustworthy, Fair, and Accountable AI** (Task 2.2)

WP2 defines a four-pillar framework to guide ethical AI design and deployment in the public sector:

1. Transparency and accountability
2. Responsible data management and access
3. Comprehensibility and multilingual support
4. Interoperability and reusability

The framework provides eight guidelines supported by concrete tools and assessment methods. These guidelines help public administrations develop explainable, inclusive, and safe AI systems aligned with democratic values and legal obligations.

3. **AIGOV Transformation and Adoption Framework** (Task 2.3)

To support practical implementation, WP2 delivers a structured five-phase transformation model enabling responsible AI adoption:

1. Assess readiness and context
2. Design ethical, value-aligned use cases
3. Build and test prototypes
4. Deploy and scale responsibly
5. Govern, evaluate, and sustain

This model integrates all WP2 outputs, supporting public administrations in redesigning services, ensuring human oversight, managing risks, and institutionalising long-term governance.

WP2 equips public administrations with actionable guidance to responsibly leverage AI while maintaining transparency, fairness, accountability, and multilingual inclusiveness. By addressing the full lifecycle of data and AI, from data collection to sustainable governance, WP2 enables public authorities to harness the benefits of Generative AI while safeguarding public trust and complying with European values and regulation.

1 Introduction

The aim of this section is to introduce the background the work pursued within Task2.1 "AIGOV Government Data Value Cycle", T2.2 "AIGOV Framework for trustworthy, fair and accountable AI", and Task2.3 "AIGOV Transformation and Adoption Framework" of the AIGOV project. The scope and the objective that the current document has set out to achieve are presented in sub-section 1.1. The intended audience for this document is described in sub-section 1.2 while sub-section 1.3 outlines the structure of the rest of the document.

1.1 Scope

The present document is the deliverable "AIGOV Holistic F ramework" (henceforth, referred to as D2.1) of the AIGOV project. The main objective of D2.1 is to document the results of Task2.1 "AIGOV Government Data Value Cycle", T2.2 "AIGOV Framework for trustworthy, fair and accountable AI", and Task2.3 "AIGOV Transformation and Adoption Framework" of WP2.

The specific objectives of WP 2 include:

- To enable governments to access robust, accurate data, in a manner that maintains privacy and conforms to societal and ethical norms by defining the AI Government Data Value Cycle
- To facilitate public authorities to assess the necessity of AI in solving a problem
- To define and deliver the AIGOV Framework for trustworthy, fair and accountable AI in the public sector that identifies trade-offs, mitigates risk and bias, and ensures an appropriate role for humans.
- To provide the required methods and tools to public administrations to transform existing processes, improve public servant skills, and enable service interoperability by delivering the AIGOV Transformation and Adoption Framework.

Although the original proposal did not specifically mention Generative AI, the unprecedented advancements in Large Language Models since 2022 have fundamentally reshaped the technological and governance landscape. To ensure that AIGOV remains future-proof and capable of addressing contemporary challenges, D2.1 expands its scope to analyse and incorporate Gennerative AI technologies, which now constitute a central component of AI adoption in the public sector.

1.2 Intended Audience of this Deliverable

The intended audience for this document is public administration, policy-makers, and anyone interested in deploying Artificial Intelligence in the public sector.

1.3 Relationship with Previous Deliverables

This deliverable builds upon and extends the work performed in Work Package 1 (WP1) of the AIGOV project. Deliverable D1.1 "State of Play Analysis" provided an extensive analysis of the

existing landscape of Artificial Intelligence in the public sector, identifying the ethical, legal, and governance challenges associated with AI adoption. In addition, Deliverable D1.2 ""The AIGOV Ecosystem." proposed the AIGOV Ecosystem, defining the foundational components and stakeholders involved in implementing ethical, trustworthy, and fair AI systems in public administrations. It introduced the four key pillars of the ecosystem, namely collection, construction, evaluation, and translation that describe the data and AI lifecycle within public organizations.

Building on these outcomes, the present deliverable under Work Package 2 (WP2) operationalizes the conceptual and analytical results of WP1 by developing methods, guidelines, and frameworks that enable the trustworthy, fair, and accountable deployment of AI systems in public governance.

In particular, this document introduces the three artefacts, (i) the AIGOV Government Data Value Cycle, (b) the AIGOV Framework for Trustworthy, Fair, and Accountable AI, and the (c) AIGOV Transformation and Adoption Framework, which translate the ecosystem principles established in WP1 into actionable methodologies, assessment tools, and transformation pathways for the public sector.

1.4 Structure

The structure of the document is as follows:

- Section 2 describes the details of the methods used to create this deliverable.
- Section 3 presents the main aspects of Generative Artificial Intelligence required to understand the contents of this deliverable.
- Section 4 presents the AIGOV Government Data Value Cycle, incorporating the steps that government data should go through in order to facilitate public value creation through AI. Towards this direction, the data types available in the public sector as well as the requirements, challenges, and guidelines for each step of the cycle are presented.
- Section 5 presents the AIGOV Framework for trustworthy, fair and accountable AI, its pillars and respective guidelines for the public sector.
- Section 6 describes the AIGOV Transformation and Adoption Framework.
- Finally, Section 7 concludes this deliverable.

2 Method

In this section we present the methodological approach followed in order to achieve the objectives of the three tasks of WP2. The approach is grounded in international best practices and aligned with emerging regulatory and governance requirements, particularly those stemming from the EU AI Act**,** the OECD AI Principles, and global public-sector digital transformation frameworks. These sources provide a normative and operational foundation ensuring that the three WP2 artefacts, namely:

- The AIGOV Data Value Cycle
- The AIGOV Framework for Trustworthy, Fair, and Accountable AI
- The AIGOV Transformation and Adoption Framework

are future-proof, compliant, and applicable across diverse public-administration contexts.

Figure 1 presents the relation of the three frameworks. The AIGOV Transformation & Adoption Framework incorporates and operationalises the AIGOV Data Value Cycle and the AIGOV Framework for Trustworthy, Fair & Accountable AI. Together, these frameworks provide the data foundations and ethical-governance requirements necessary for responsible and effective AI adoption in the public sector.

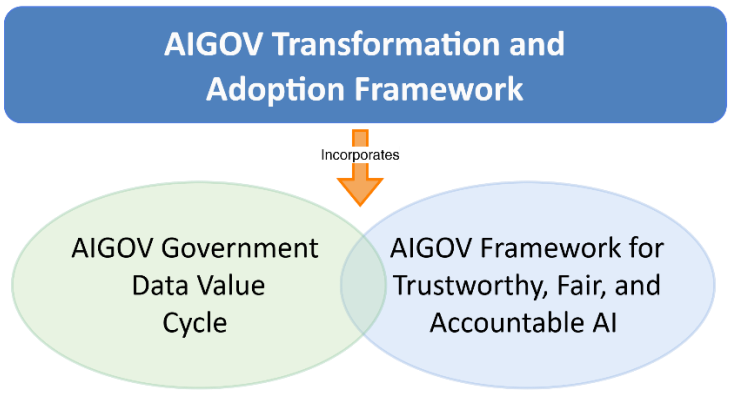

Figure 1 Integration of AIGOV Frameworks

2.1 Method for the AIGOV Data Value Cycle

This Section presents the method used to synthesize the AIGOV Data Value Cycle. The objective was to define a structured, ethically grounded process for managing government data in support of trustworthy Artificial Intelligence (AI) systems and public value creation.

The methodology is aligned with the OECD Recommendation on Open Government Data, the OECD Framework for Data Governance in the Public Sector, and the EU AI Act's provisions on data quality, representativeness, and governance for high-risk AI systems.

The methodological approach comprises the following steps:

Step 1. Identify best practices, standards, and challenges in public-sector data management. A comprehensive review of scientific, policy, and technical literature was conducted to identify best practices, standards, and challenges in public-sector data management. The two key sources included:

- OECD's Government Data Value Cycle [76], widely recognised as the global reference model for public-sector data governance.
- McKinsey & Company [65] industry analysis on data architectures for Generative AI, which provides up-to-date insights on data readiness for AI ecosystems.

Step 2. Identification and classification of public-sector data types relevant for AI scenarios. Structured, unstructured (text-rich), and dynamic sensor data were analysed to determine their availability, risks, quality challenges, and suitability for AI and LLM-based use cases. This step included assessing metadata needs, provenance requirements, and interoperability constraints. This step ensures alignment with both traditional data-governance principles and evolving requirements linked to Generative AI adoption.

Step 3. Stakeholder-informed refinement through participatory workshop. A co-creation workshop with public administration representatives and postgraduate students provided practical insight into current data practices, barriers, and organisational constraints. These insights contextualised technical standards with real administrative workflows.

Step 4. Synthesis of requirements, challenges, and guidelines for each stage of the Data Value Cycle. The cycle was expanded to seven steps (collection, curation, integration/linking, storing, dissemination, usage, and value creation). For each step, legal (GDPR, AI Act), organisational, ethical (bias, representativeness), and technical (APIs, semantic standards, vector databases) requirements were systematically documented.

Step 5. Incorporation of Generative AI-specific data considerations. Requirements for unstructured data mapping, embedding generation, semantic search, and RAG (retrieval-augmented generation) pipelines were integrated, ensuring that the cycle reflects the needs of modern LLM-based public-sector applications.

2.2 Method for the AIGOV Framework for Trustworthy, Fair, and Accountable AI

This section presents the method used to synthesize the AIGOV Framework for Trustworthy, Fair, and Accountable AI, with a particular emphasis on Generative AI.

The objective is to establish a structured, evidence-based, and ethically grounded framework that supports the trustworthy and responsible deployment of AI in the public sector. The framework seeks to address both the technical and governance dimensions of AI by providing practical guidelines that ensure transparency, fairness, accountability, and inclusiveness.

The methodological comprises the following steps:

Step 1. Identification of the main AI governance frameworks. The first step focuses on identifying and analyzing the main international and European policy, ethical, and technical sources, including frameworks that define principles of trustworthy and responsible AI (e.g., EU AI Act, OECD AI Principles, UNESCO AI Ethics Recommendation, EC Ethics Guidelines for Trustworthy AI).

Step 2. Extraction and alignment of core principles for trustworthy public-sector AI. From the reviewed sources, principles were grouped into four thematic pillars:

1. Transparency & accountability
2. Data management & access
3. Comprehensibility & multilingual support
4. Interoperability & reusability

Step 3. Analysis of Generative AI–specific risks and requirements. The methodology incorporated new issues relevant to LLMs: hallucinations, opacity, multilingual performance, bias propagation, explainability challenges, and dynamic safety evaluation. These risks informed the selection of guidelines and assessment tools.

Step 4. Synthesis of guidelines and mapping to assessment tools. Eight guidelines were defined, each supported by practical evaluation methods such as SHAP/LIME explainability, multilingual evaluation benchmarks, data quality assessment, API testing, and FAIR data compliance checks. This ensures actionable, operational guidance.

2.3 Method for the AIGOV Transformation and Adoption Framework

This section presents the method used to synthesize AIGOV Transformation and Adoption Framework.

The development of the framework followed a design science and evidence-based policy methodology, ensuring rigour, applicability, and alignment with global public-sector standards. The method consisted of four steps:

Step 1: Review of international digital-government transformation frameworks. In this step, key frameworks were analysed to identify best practices, structure, and applicability, including the OECD Digital Government Policy Framework[1], the UN Digital Government Model framework[2], the World Bank GovTech Maturity Index[3], and the OECD E-Leaders Handbook on

[1] https://www.oecd.org/en/publications/the-oecd-digital-government-policy-framework_f64fed2a-en.html
[2] https://www.un-ilibrary.org/content/books/9789211067286c006
[3] https://www.worldbank.org/en/programs/govtech/gtmi

the Governance of Digital Government[4]. This analysis ensured alignment with internationally recognised public-sector transformation principles.

Step 2: Identification of gaps in existing frameworks related to AI. While existing frameworks effectively support digital transformation, they do not fully address emerging challenges relating to Generative AI, such as multilingual human-AI interaction, algorithmic accountability, data governance for AI models, and human-in-the-loop decision making.

Step 3: Integration of AIGOV-developed components. The framework incorporates findings from (1) the AIGOV Data Value Cycle, ensuring data accessibility, quality, interoperability, and ethical use, and (2) the AIGOV Framework for Trustworthy, Fair and Accountable AI, establishing governance, assessment, and oversight principles to ensure public legitimacy, fairness, and transparency.

Step 4: Framework synthesis and structuring into transformation phases. The final framework was structured into sequential phases that public administrations can follow to assess readiness, design AI-enabled services, pilot and refine solutions, scale adoption, and continuously improve outcomes.

[4] https://www.oecd.org/en/publications/the-e-leaders-handbook-on-the-governance-of-digital-government_ac7f2531-en.html

3 Generative Artificial Intelligence in the Public Sector

Generative AI has a promising potential in a wide range of applications. In contrast to conventional AI systems that used to concentrate on tasks such as classification or prediction, Generative AI models are now capable of generating novel data, images, text, code, and even music that resembles content created by humans. Driven by advanced algorithms Generative AI these models generate realistic and unique outputs by learning from large datasets. Generative AI is currently transforming the way we create content, develop products, and even diagnose diseases, hence, has application in business, entertainment, and healthcare. This chapter is an introduction to Generative AI, its potential for the public sector, and its promising impact.

3.1 Generative Artificial Intelligence and Large Language Models

The enormous amounts of text data that are daily generated and stored can be exploited by AI and Natural Language Processing (NLP) technologies. Generative AI can generate text, images, videos, code, and other types of data in a way that mimics human actions. Large Language Models (LLMs) are at the core of Generative AI. LLMs are AI deep learning models trained on vast amounts of datasets that are capable of performing Natural Language Processing (NLP) tasks.

There are two main Generative AI techniques: Generative Adversarial Network (GAN) and Generative Pre-trained Transformer (GPT). GAN employs a dual-network system (generator and discriminator) to produce synthetic data and authenticate its genuineness. GAN is usually used for video and voice generation. GPT models leverage publicly available digital content to generate human-like text in various languages and contexts.

Over the last few years, Large Language Models (LLMs) have been established as the premier approach for a plethora of NLP tasks, including chatbots and virtual assistants. Central to the development of LLMs is the use of neural-network-based statistical models which assign probabilities to sequences of words, coupled with the transformer architecture introduced in 2017[5]. LLMs are pre-trained on massive text datasets to understand, interpret, and generate human-like text, acquiring knowledge about language structure, semantics, facts, and even limited reasoning abilities.

Ever since the invention of the revolutionary transformer architecture [18] [105], and the attention mechanism [112], LLMs have advanced at a tremendous speed. The creation of models with the ability to capture contextual information regardless of orientation has been

[5] Vaswani A. et al, (2017) Attention is All you Need, Advances in Neural Information Processing Systems 30 (NIPS 2017)

a game changer in terms of model performance, allowing for the creation of systems reliant on their capacities. LLMs can perform various tasks, from simple text classification [21] and named entity recognition [65] to summarization [57] and translation [135]. However, the largest potential of LLMs stems from their ability to understand [8] and generate human-like texts, enabling countless applications. LLMs with such capacities are becoming more and more available to the public.

At the same time, it is widely believed that the U.S. industry is ahead of European and open-source competitors when it comes to LLM capabilities[6]. However, during the last years several initiatives aimed at improving Europe's standing on Language Technologies (LT) including LLMs. The recently started European Language Data Space aims at facilitating the sharing of language data and models for multilingual and multimodal LTs. The European Language Grid (ELG) provides a scalable and powerful infrastructure using a grid architecture for European LTs. The European Language Equality (ELE) project developed strategies, implementation plans, and roadmaps towards digital language equality in Europe. Moreover, the High-Performance Language Technologies (HPLT) project develops multilingual training materials and train language models that support European languages. OpenGPT-X, a collaborative project between science, business and technology funded by the German government, builds and trains LLMs for the EU economy and intends to offer open-source versions of its models.

The emergence of large language models (LLM) has led to the development of novel approaches that aim to uncover information located within natural language data. Based on the ever-evolving capacities of the models particularly regarding the tasks of natural language understanding and generation, applications have been created that harness this ability, in a variety of fields like medicine [93] [52], education [45], and finance [122].

3.2 Foundation Large Language Models

The term foundation models, in essence, refers to large, pre-trained models that establish the groundwork for deploying subsequent models using various approaches. Foundation LLMs undergo extensive training on massive datasets for extended periods of time, demanding substantial computational resources, and consequently achieving state-of-the-art performance levels.

In the field of NLP, ever since the invention of the revolutionary transformer architecture [112] and in recent years, many natural language foundation models have been created. However, the majority of LLMs are either closed source, non-European, or both. For example, the majority of LLMs are developed, owned, and operated by a few dominant technology giants, primarily in the U.S. and China, including Google with PaLM and Bard, OpenAI with

[6] Council of the European Union (2023), ChatGPT in the Public Sector – overhyped or overlooked?

ChatGPT, and Baidu with Ernie 3.0 Titan. Amongst these models are the GPT [131] LLM family (GPT-1 [80], GPT-2 [95], GPT-3 and GPT-3.5 [129], and GPT-4 [1]) released by OpenAI, the Llama LLM family (Llama [110] and Llama2 [111]) released by Meta, Google's BARD that is based on LaMDA [107], Mistral AI's Mistral [43] and Mixtral 8x7B [41] and many more. These foundation models, nowadays widely available, either through APIs or locally hosted solutions, offer new ways to manipulate natural language textual information.

OpenAI Foundation Large Language Models.

The models of the GPT series are among the most important introductions of Open AI, and American research organization. According to OpenAI, OpenAI's large language models are developed using three primary sources of information: (1) information that is publicly available on the internet, (2) information that we license from third parties, and (3) information that our users or our human trainers provide.

As their name states, the GPT models are based on the Transformer architecture. The GPT models are pre-trained on a vast amounts of text data from various sources making them aware of grammar and facts. They also have some reasoning capabilities. In order to enable GPT models to adapt to specific domains, they can be fine-tuned using domain-specific datasets. GPT models take a prompt as an input and generate coherent and contextually relevant text based on it. GPT models are able to effectively and accurately perform a variety of tasks, including text completion, translation, and summarization.

The first version of the GPT model, introduced in 2018, demonstrated the potential of the Transformer architecture for natural language processing tasks. The second version of GPT is GPT-2. GPT-2 was released in 2019. It had 1.5 billion parameters and was initially withheld due to concerns about potential misuse. However, GPT-2's capabilities were significantly improved related to the previous version. The next version was GPT-3, which was launched in 2020. GPT-3 follows the trend of increasing the size of language models having 175 billion parameters, which makes it one of the largest and most powerful language models at that time. Its main advantage is that it requires little or no fine tuning to perform a wide range of tasks. The GPT- 3.5 variant is the next version of the GPT-3 models released on 2022. This variant focused more on making significant refinements of its predecessor rather than just scaling up its size. Finally, the latest version introduced by Open AI in March 2023 is GPT-4, which is a significant milestone in the history of AI. Apart from having more parameters and training data, GPT-4 differentiates previous models because of its superior natural language creation and comprehension. Even more varied and comprehensive datasets are used to train this model. In addition, this version advances previous versions in many critical areas including enhancements in understanding of the context variables, which results in more coherent and contextually accurate content. GPT-4 also is more effective in reducing bias and, hence, generate more fair, trustworthy and respectful content. GPT-4 is also capable of processing image and text inputs and producing text outputs.

ChatGPT, is an LLM-based system offering the capabilities of a chatbot service. The current free version of ChatGPT is based on GPT-3.5, while the current paid version of GPT, namely ChatGPT Plus, is based on GPT-4. Finally, ChatGPT provides access to ChatGPTs that allow discovering and creating custom versions of ChatGPT that combine instructions, extra knowledge, and any combination of skills.

Llama family of Large Language Models by Meta.

Llama was firstly released on February 2023. Llama focuses on advancing the performance capabilities of smaller models, rather than through increasing the number of parameters. Specifically, Llama is available with a size of 65B, 33B, 13B, and 7B parameters. LLaMA 65B and LLaMA 33B are trained on 1.4 trillion tokens, while LLaMA 7B on one trillion tokens. LLaMA takes as an input a sequence of words to predict the next word and, recursively, generate the final text. To this direction, text from the 20 most commonly spoken languages were used with a main focus on languages that use the Latin or Cyrillic alphabets. Llama is released under a non-commercial license.

On July 2023 Llama 2, the next version of Llama LLM was released, which is free for both research and commercial purposes. Llama 2's model weights and starting code for the pre-trained model and conversational fine-tuned versions are also available. In this version of Llama, Meta also gives much attention to transparency. Comparing to Llama and apart from being open to everyone, Llama2 offers double the context length of Llama (4,096 tokens) allowing for greater complexity and a more coherent, fluent exchange of natural language. It is pre-trained on more data increasing its knowledge base and contextual understanding. Finally, the fine-tuning of the model used Reinforcement Learning from Human Feedback (RLHF), making its responses more aligned with the human responses.

Google's Bard, Gemini, and Gemma Large Language Models.

On February 2023, Google announced Bard. Bard was initially trained with an LLM called Google LaMDA. In May 2023, Bard was re-trained with a new LLM called Pathways Language Model 2 (PaLM 2). PaLM 2 can process information up to 500 times faster than LaMDA and is up to 10 times more accurate.

Subsequently, on December 2023 Google released the descendant of Bard, Gemini 1.0, with capabilities of image, audio, video, and text understanding. The Gemini family consists of three models with different sizes; Gemini Ultra serving high complex tasks, Gemini Pro serving a wide range of tasks, and Gemini Nano serving "on-device" tasks. Gemini Ultra was the first LLM to outperform human experts on a specific benchmark, obtaining a score of 90%.

The next release of Gemini, Gemini 1.5 was introduced in February 2024. This LLM is expected to be more powerful and capable model than 1.0 Ultra due to a number of technical advancements, including its new architecture and a larger one-million-token context window,

which equates to roughly an hour of silent video, 11 hours of audio, 30,000 lines of code, or 700,000 words.

In parallel with the release of Gemini 1.5, Google also released Gemma. Gemma are free and open-source LLMs and lightweight versions of Gemini. The Gemma family includes two models, with 2B and 7B parameters, respectively.

Mistral AI's Large Language Models.

Mistral includes two types of models; open-weights models, i.e., Mistral 7B, Mixtral 8x7B, and Mixtral 8x22B, which are high efficient models and optimized commercial models, i.e., Mistral Small, Mistral Medium, Mistral Large, and Mistral Embeddings. The first type is available under a fully permissive Apache 2 license. Optimized commercial models are designed for high performance and are available through flexible deployment options. Mistral versions are also available through an API. Mistral ranks second among all models available through an API.

3.2.1 Fine tuning of Large Language Models

Although Large Language Models have proven to be extremely capable, they are also extremely costly to train. For this reason, efficient ways to harness the power of LLMs have been explored. Towards this direction, four main approaches have been adopted that focus on adapting existing LLMs; (1) prompt engineering, (2) prompt learning, (3) adapter fine tuning, and (4) full parameter fine tuning. The second and third approaches belong to the wider group of parameter efficient fine tuning (PEFT), while only the first approach keep the weights of the LLM stable.

- *Prompt engineering* refers to the process of designing effective prompts to elicit desired responses from the LLM without altering the model's weights. Individual techniques of prompt engineering include chain of thought reasoning [116] and in-context learning [80] [6].
- *Prompt tuning* [54] adds trainable parts to the input layer of the LLM and trains them in order to have them act as conceptual prompts for the model by guiding its predictive abilities to a certain direction.
- *Adapter fine tuning* adds trainable components to the inner network. This allows for greater adaptation of the initial model to new tasks, and has the upside of minimizing the chances of catastrophic forgetting taking place in the network. By employing adapters, the fine-tuning process becomes much more computationally efficient with a fraction of the training data required for full-parameter fine tuning. Amongst the state of the art adapters present today are IA^3 [58] and LoRa [39].
- *Full-parameter fine tuning*, where all model weights are retrained with specific data. Supervised fine tuning is commonly employed, while Reinforcement Learning (RL) methods also exist, such as RL from Human Feedback (RLHF) [93] and Reinforcement

Learning from AI Feedback (RLAIF). Both methods use feedback to iteratively improve the model, but provide different levels of balance between quality of evaluation and cost. These RL methods are commonly used in model alignment.

3.2.2 Language adaptation for Large Language Models

Training large language models from scratch requires vast amounts of data. For this reason, most LLMs are created in resource rich languages such as English, Chinese, and Spanish etc. In order to create models of equal quality in resource-constrained languages, it is necessary to repurpose the aforementioned models in an efficient way. Language adaptation in this context, refers to the process of fine tuning a language model that was originally trained for natural language understanding and generation in a specific language (source language) to another one (target language).

Research efforts have been made regarding language adaptation in resource-constrained settings. In [131], researchers employed an adapter based PEFT method of IA^3. Their approach resulted in successfully adapting the BLOOM multilingual model in eight new languages, including Greek, using as training data 100K samples from the OSCAR multilingual dataset per language, run on an RTX 3090 GPU of 24GB memory. In [78], progressive cross-lingual transfer learning is presented. In this method, a large LLM, trained on a source language and a smaller LLM, trained on the target language, are used jointly to initialize the parameters of the embeddings of a new model, which is of equal size to the source model, but in the target language. The newly initialized model is then fine-tuned on datasets of the target language. This significantly reduces the need for computational resources while still maintaining a high level of quality regarding the output model, as shown in the paper, where state of the art scores were achieved in zero-shot downstream tasks.

3.2.3 Conversational adaptation

The training data for conversational generative LLMs should mimic the structure of dialogue inputs and outputs in order to produce responsive models that are conversational and generative. Additionally, the model should be further aligned based on specific values in order to produce a model that better aligns with end tasks and user preferences. Alignment is intended to change the model's behaviour so that it will respond with human values like helpfulness, safety, and dependability.

In terms of the current state of the art landscape for conversational adaptation, the LLama2-Chat series of models, which achieved state of the art metrics and even outperformed ChatGPT-were created by adapting the LLama2 foundational LLM [111] in a way that combines conversational alignment and fine tuning. On the other hand, a model learns practically all of its skills and knowledge during pre-training, while alignment instructs it on which sub distribution of formats to employ while dealing with users, according to the Superficial Alignment Hypothesis put forth in [138]. This means that a model could be

efficiently modified for discussion utilizing PEFT approaches, instead of requiring expensive full parameter supervised fine tuning, or RLHF, with a small number of well-chosen training samples.

3.2.4 Evaluation of Large Language Models

The evaluation of LLMs is a field of active research. In order to effectively evaluate the LLMs' capacities across a wide array of tasks, various benchmarks have been developed. With regards to multilinguality, the XNLI [8] benchmark is commonly used. It evaluates the level of cross-lingual language understanding in 15 languages, including Greek. Concerning the aspect of dialogue, the SQuAD [90] question answering benchmark, with XQuAD [1] being its multilingual counterpart, have been created. Lastly, in order to evaluate the model's factual reliability, FEVER [109], as well as the multilingual TRUE [18] benchmarks have been developed, that examine this aspect of the models' responses.

3.2.5 Retrieval Augmented Generation

As a consequence of their first training, LLMs possess an extensive amount of world knowledge [80], but they are not able to answer all questions with accuracy. Consequently, LLMs frequently cause "hallucinations" [140]. LLM responses are called "hallucinations" when factually false but are presented as true by the LLM are called hallucinations.

Retrieval Augmented Generation (RAG) is a technique employed in generative LLMs to improve the factual capabilities of the LLMs [56], i.e., improve the factuality of their output. More precisely, the original input is used to obtain context from a pool of factually valid and externally provided texts based on a criterion, typically embedding similarity. The augmented prompt is given to the generator once the pertinent text or texts have been acquired and added to the original question as context. When selected carefully, context has been demonstrated to have positive effects on the factuality of the generated outputs [80]. On the other hand, if the context is inappropriate, it could result in poorer performance [95], hence choosing the proper context is extremely important.

RAG systems rely on large, curated textual datasets with accurate factual information. After dividing these datasets into chunks, embeddings are made for every chunk. The result is more closely aligned with reality by storing both in vector stores, retrieving them using similarity techniques, and feeding them into the model as extra input.

3.2.6 Instruction Learning

The enormous potential of Large Language Models originates from the generative models' capacity to learn from examples and instruction on a variety of natural language tasks that they were not trained on initially, an ability that frequently surpasses fine-tuning approaches [6]. In particular, it has been demonstrated that instruction-fine-tuned as well as pretrained models' zero-shot task generalization performance can be improved through in-context

instruction learning [45]. Conversational LLMs are trained to do novel, unseen tasks as humans would through training and example-based learning. The LLM is typically able to complete the new natural language job with excellent performance by comprehending it when it is presented in a descriptive way. Applications of this type of human-oriented instruction [59] include summarization [80], text classification [104], and named entity recognition [116]; among others. These are activities that were previously completed by specialized models that had been trained or adjusted for these uses.

3.2.7 Reasoning in Large Language Models

Reasoning in Large Language Models pertains to their ability to form thought patterns resembling human thought processes. They have been shown to display strong capacity for abstract pattern induction in analogical tasks [107], and can do so in zero-shot scenarios effectively [105]. Reasoning plays a crucial part in planning and following instructions towards a specific and predetermined goal. These instructions can range from simple directions (zero-shot), to example based learning (one-shot and few-shot learning), and result in systems adept at several downstream tasks. However, the level to which these emergent abilities are utilized, largely relies in the way that the model's input is formulated.

3.2.8 Agents

The concept of agents revolves around the notion of self-adjusting computational systems performing operations [29] [122]. Harnessing the aforementioned emergent capabilities of LLMs, mainly reasoning through LLM-generated prompt engineering, has shown to lead to advancements in autonomous, or semi-autonomous agents that can utilize tools and adapt to diverse scenarios more adequately [137].

These agents generate plans towards the actualization of a specific goal, execute the individual steps of the plan with usage of additional tools and capabilities, and have advanced self-correcting capacities enabled by memory modules allowing for flexible and efficient operation.

Autonomous and semi-autonomous agent-based systems have already been used across various domains, such as medicine [49] and the public sector [61], showing great promise in the future.

3.3 The potential of Generative Artificial Intelligence for the Public Sector

The public sector, including both governments and parliaments, generates and stores enormous amounts of text data, which can be harnessed by Artificial Intelligence (AI) and Natural Language Processing (NLP) technologies. For example, Open Government Data (OGD), offer a way for information to be easily accessible to everyone. In particular, OGD (often published as linked data) offer the added benefits of semantically structured information, allowing for complex queries and retrieval of factually correct and frequently

updated information. By combining the natural language understanding and generation capacities of LLM with the widely available information contained within OGD through techniques such as Retrieval Augmented Generation, a system could overcome the LLM's inability to retain factual information while at the same time offer an easier way to access and retrieve OGD.

In addition, traditionally public decisions have been taken so far mostly by human beings [2]. Generative AI and LLMs however are now able to produce, under certain conditions and within certain limits, decisions similar to human ones, and, in some cases, it can extract information hidden in data that is not identifiable to humans and hence, even exceed the cognitive possibilities of humans [3] [4] .

In D1.1 State of Play Analysis, a list of cases from the public sector that leverage AI was presented. Since then, public sector bodies are already beginning to deploy Generative AI to enhance their services. For example, recently, the Greek Government has deployed a beta version of the "mAIgov"[7], a chatbot powered by Microsoft Azure OpenAI technologies, that has been trained using open data sourced from gov.gr, the Greek Open Government Data portal, and various other public entity websites. Additional public sector cases that use Generative AI are presented in Table 1.

Table 1 Cases of using Generative AI in the Public Sector

	Aim	GenAI Task
1	Assist social service delivery for people experiencing homelessness [67]	Text summarisation
2	Using a Large Language Model to Choose Effective Climate Change Messages [6]	Generation and evaluation of climate-change messages using a Large Language Model
3	OIE4PA, our latest study on extracting and classifying relations from tenders of the Public Administration [98]	Text Classification
4	Enhancing public access to digital governmental data [112]	Chatbot
5	Generative AI for street-level bureaucracy [95]	Chatbot
6	Topic Classification in the Domain of Public Affairs [84]	Text classification

[7] Digital assistant gov.gr, "mAIGov"

7	Legal assistant tailored for interacting with legal resources [63]	Question Answering
8	Improve citizen understanding of complex government policies [134]	Policy explanation and Q&A
9	Provide accessible explanations of municipal budgets and enable participatory budgeting [126]	LLM-powered chatbot generating natural-language budget explanations

Relevant LLM-based applications with huge potential for the public sector include the deployment chatbots and virtual assistants that address basic questions or recommend specific government services to the citizens (e.g., Passport Issuance procedure); document analysis for identifying key information in complex documents such as legal contracts; summarisation of large volumes of text; assisting in decision-making by generating reports and evaluating applications and grants; screening CVs and matching candidates for recruitment; and providing advanced internal knowledge retrieval and search services, utilizing the wealth of information provided by the public sector, from documents across different departments, ministries and local authorities to Open Government Data (OGD) portals.

These proprietary LLMs that can be used in the public sector follow a black-box approach and thus impede accountability, transparency, impartiality, or reliability, which are core principles of the public sector. Recently, the use of knowledge graphs has been proposed as a strategy to technically mitigate many of these issues. Moreover, these models mainly focus on widely used languages (e.g., English) and fall short of supporting under-represented ones, including many of the 24 official languages of the European Union.

3.4 Challenges of adopting Large Language Models

The majority of LLMs are either closed source, non-European (Llama 2), or both (GPT-3.5 and GPT-4). Such models pose significant concerns regarding accountability, sufficient multilingual capacity and transparency by raising concerns about biases, security vulnerabilities, and the potential exploitation of user data. To address these issues, fostering the adoption of open source European LLMs (e.g., BLOOM, Mistral 7B, Mixtral 8x7B) becomes crucial for promoting linguistic and cultural diversity, privacy, and innovation within Europe's digital landscape.

This sub-section presents the main challenges of adopting Large Language Models.

3.4.1 Multilingualism in Generative Artificial Intelligence

Large Language Models are typically trained on widely used languages (e.g., English) and fall short of supporting under-represented ones. Consequently, their performance is **limited for**

less widely spoken languages and dialects [91], including many of the 24 official languages of the European Union (e.g., Greek).

3.4.2 Privacy and Security

Additionally, the use of systems relying on another application programming interface (API), such as integrating Open AI's GPT-4 API into a government service, poses significant challenges related to data **privacy and security.** For example, in a public sector setting, LLMs may have access to sensitive government data or they could be trained in massive personal data from the Web without the consent of the users violating General Data Protection Regulation (GDPR) and similar data protection laws.

3.4.3 Accuracy

The accuracy of content that is generated by Generative AI is very important since potential mistakes may lead to liability issues. For example, in translation tasks, common error include terminology mistakes, inconsistent translation of homonyms, and omissions in otherwise grammatically correct translations [100]. Especially in public administration, accuracy mistakes may lead, for example, to legal consequences [21].

3.4.4 Hallucinations

While LLM hold great potential for the public sector, generating grammatically correct and convincingly fluent text, they can inadvertently propagate inaccuracies present in their training data and generate text that is factually incorrect. Indeed, even though LLMs hold considerable amounts of world knowledge, thanks to their initial training, they still lack the ability to be factually correct in all their responses, which often results in "hallucinations" [140]: factually incorrect responses presented by the LLM as correct ones. In order to tackle this weakness, several methods have been developed, mainly utilizing prompt engineering. Amongst those methods is the supply of factually correct context along with the input. The main, state of the art, proposed architecture supporting this is retrieval augmented generation (RAG). In RAG, specialized components retrieve information relevant to the original input from corpora and supply it to the model along with it. This has been shown to boost the factual capacities of the models significantly. In cases where no external corpora are available, techniques like chain of thought prompting [116] and tree of thought prompting [125] are usually employed. These techniques enhance the reasoning capabilities of the LLMs by providing examples of the correct flow of thought that the model should follow, something that allows for more effective access and utilization of its already existing knowledge.

As a result, public officials should focus on reviewing and validating the outputs of the LLMs, since it it's incumbent upon humans to discern the veracity and applicability of AI-generated content. Thus, the challenge lies on seamlessly merging the capacities of AI systems with expert human judgments.

3.4.5 Explainability of Large Language Models

Proprietary LLMs follow a **black-box approach** and thus impede accountability, transparency, impartiality, or reliability, which are core principles of the public sector. LLMs are regarded as sophisticated and intricate systems, hence they are still criticized for their **lack of interpretability.** Composed of billions of parameters, including layers, neurons, and transformer blocks, the inner workings of these models remain largely opaque. This considerable complexity significantly **complicates the interpretation and understanding** of their mechanisms.

Towards addressing this issue, a range of explainability techniques for LLMs, including but not limited to methods such as layer-wise relevance propagation, attention visualization, and feature attribution, help in tracing model decisions back to input data. Additionally, the use of interpretable (surrogate) models and human-in-the-loop systems can provide greater transparency and understanding of LLM outputs.

Neuro-symbolic approaches have also emerged towards this end, where a combination of neural networks and knowledge bases are paired, to result in explainable, factually reliable systems that combine the predictive capabilities of the state of the art machine learning models with externally provided information [65]. The systems usually employ knowledge graphs in order to store and access the external information, since graph structures provide an efficient way to represent entities and their connections [79] [44]. This further enhances the intrinsic explainability potential of the system, since predictions are influenced by concrete facts.

3.4.6 Footprint

Large Language Models, due to their extensive computational requirements and energy required for training and inference, can have a significant negative impact on the environment. Taking into consideration that governments need to meet their net zero targets and other environmental commitments, they shall focus on understanding LLMs functionalities and propose new approaches, such as developing smaller, domain or task specific language models.

3.4.7 Cost

Although LLMs have proven to be extremely capable in a plethora of Natural Language Processing tasks, they are also usually tremendously costly to train and may cost thousands of euros per month. To this end, two main pricing strategies exist depending on the type of the model; the first is for LLMs provided by API where the cost is calculated according to the amount of tokens (e.g., words) that have been processed by the model, while the second one is for hosting LLMs in local or cloud infrastructure, where the cost is calculated based on the type and time of computing sources usage (e.g., GPUs).

3.4.8 Data

The rapid ascent of (1) data daily collected and/or generated by the public sector and (2) digital technologies including AI is reshaping economies and societies, presenting governments with profound challenges and opportunities in their daily operations. In the 21st century, governments face mounting pressure to meet citizen expectations, manage constrained budgets, and address emerging policy issues. Failure to adapt to this evolving landscape risks undermining public trust and exposing governments to significant risks. While data holds promise for societal advancement, its transformational potential remains largely untapped due to obstacles such as legacy technologies, skills gaps, and legal constraints in the public sector. Despite some progress in harnessing data strategically to enhance policy making and service delivery, its utilization is not yet universally recognized or adequately resourced as a fundamental driver of public value creation. Hence, the positive effects of applying state-of-the-art technologies on public sector data will not be possible if data is not ready [57].

According to a recent report of McKinsey [65], cost effectively managing and scaling data are among the most challenging tasks that prevent leading organizations applying Generative AI in real use cases. Furthermore, a Salesforce survey [96], found there is a lot of concern over the quality of data used in Generative AI. Towards this direction, advances in AI are now making data management a high priority.

3.4.9 Ethical Concerns

Potential misuse of generated content, biased, and unfair LLMs are among ethical concerns that prohibit the wide deployment and use of Generative AI models and applications in real life settings, especially when it comes to the public sector.

3.5 Scientific and/or social impact of Generative Artificial Intelligence for the public sector

It is estimated that Generative AI could elevate the global GDP by 7% over the next decade[8] and, at the same time, bring forth automation to approximately 300 million jobs worldwide, contributing to the global economy $2.6 *trillion* to $4.4 *trillion* annually[9]. The impact of using LLM in the public sector will be unprecedented. Applications include the deployment of chatbots and virtual assistants that address basic questions or recommend specific government services to the citizens (e.g., Passport Issuance procedure); document analysis for identifying key information in complex documents such as legal contracts; summarisation of large volumes of text; assisting in decision-making by generating reports and evaluating

[8] Goldman Sachs (2023) "Generative AI could raise global GDP by 7%"

[9] McKinsey (2023) "The economic potential of generative AI: The next productivity frontier."

applications and grants; screening CVs and matching candidates for recruitment; and providing advanced internal knowledge retrieval and search services, utilizing the wealth of information provided by the public sector, from documents across different departments, ministries and local authorities to Open Government Data (OGD) portals. They have the potential to streamline bureaucratic procedures, reduce administrative burdens, and make public services more transparent and accountable. The introduction of such models can also lead to more proactive and data-driven public interventions, improving outcomes in areas such as public health, urban planning, and disaster response.

The broader implications of advancing LLMs are profound in terms of societal, economic, and educational benefits. From an economic perspective, the evolution of LLMs can lead to considerable savings by automating and enhancing various tasks, thereby reducing the time and resources spent on them. Moreover, as LLMs evolve gaining advanced capabilities, they necessitate specialized research and development, leading to employment prospects in areas like artificial intelligence, machine learning, and natural language processing. Socially, the introduction of LLMs fosters a culture of continuous learning and adaptation. The democratization of knowledge that LLMs offer can lead to broader public engagement and bridge the informational divide. Furthermore, they have the potential to significantly elevate the quality of life for individuals by providing personalized education, aiding in medical diagnosis, and offering solutions to everyday problems, enhancing overall well-being.

4 Unlocking Public Sector Potential: Harnessing the Power of Data

Recent AI advancements make the secure generation, sharing, and usage of data more possible than ever before. At the same time, concerns over data integrity pose substantial risks for AI and other emerging technologies, necessitating close attention to cybersecurity threats and national security implications.

In this context, information resilience has been recently defined as "*The capacity of organisations to create, protect, and sustain agile data pipelines, that are capable of detecting and responding to failures and risks across their associated value chains in which the data is sourced, shared, transformed, analysed, and consumed*" [93]. Data produced and consumed by the public sector should, hence, be handled in such a way that information resilience is preserved.

Towards this direction, the Organisation for Economic Co-operation and Development (OECD) presented the idea of a government data value cycle [76] presented in Figure 2. The data cycle comprises four distinct phases of data management within government: data collection and generation; data storage, security, and processing; data sharing, curation, and publication; data utilization and reutilization. The first two phases primarily focus on the public sector's responsibility in managing and safeguarding the data it generates, collects, and retains, with significant implications for data rights and the preservation of public value. Conversely, the latter two stages present opportunities for generating novel public value, which will be explored further in the latter part of this chapter.

1) Data collection and generation; this initial stage of data application within government encompasses accessing diverse data sources, including those generated by government activities, third-party datasets, IoT devices, service designs, government contracts, and collaborations with private sector entities, emphasizing the importance of universal data standards across sectors, while laying the groundwork for future data reuse and shaping citizen experiences of government services.
2) Data storage, security, and processing; this critical phase of the data cycle involves storing, securing, and processing data, which not only ensures public trust in the public sector's data handling capabilities, but also lays the foundation for subsequent phases, with a focus on internal decisions regarding data governance and infrastructure, particularly pertinent for those managing personally identifiable information.
3) Data sharing, curation, and publication; the third phase of the government data value cycle involves the sharing, curation, and publication of stored, secured, and processed data, with considerations for legal constraints and the importance of data interoperability platforms and licensing to ensure data quality and accessibility.
4) Data utilization and reutilization; The final phase of the government data value cycle, focusing on the use and reuse of data, presents the most visible opportunity for generating public value, supported by a robust data governance ecosystem, which is crucial for ensuring data quality and accessibility throughout the cycle.

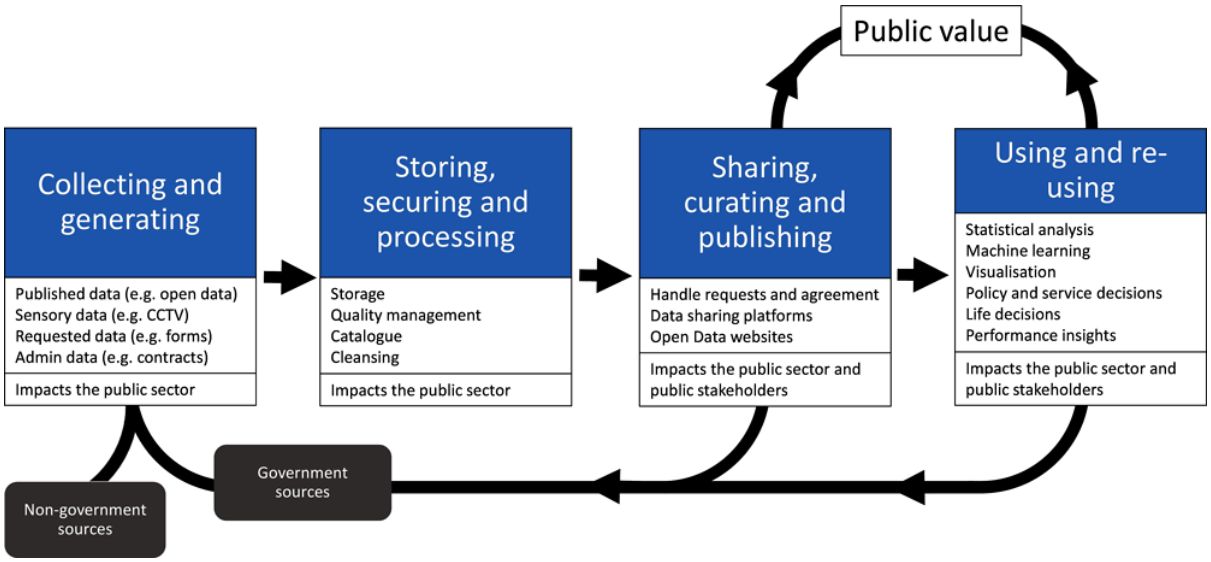

Figure 2 The government data value cycle (Source: [76])

In addition, the recent advances in Generative AI require a shift in the strategy of data management in order to be able to adapt to Generative AI needs. McKinsey & Company [65] recently documented the upgrades needed within existing traditional data architectures to enable Generative AI (Figure 3). Specifically, the key upgrades include:

- *Mapping and tagging unstructured data sources.* Since Large language models (LLMs) primarily require unstructured data (e.g., text data) there is a need to map out all unstructured data sources and establish metadata tagging standards so models can process the data and also data can be found when needed.
- *Establishing Vector databases* in data repositories in order to store prioritised content of data as embeddings.
- *Implementing data pre-processing pipelines* (e.g., convert in proper formats) so as to prepare data for deploying and/or dine tuning LLMs.
- *Integrating LLMs* to develop more sophisticated applications of Generative AI.
- *Developing prompt-engineering capabilities* for enhanced data services. Prompt engineering should be effective enough so as to be able to structure LLM questions in a way that elicits the best response from generative AI models

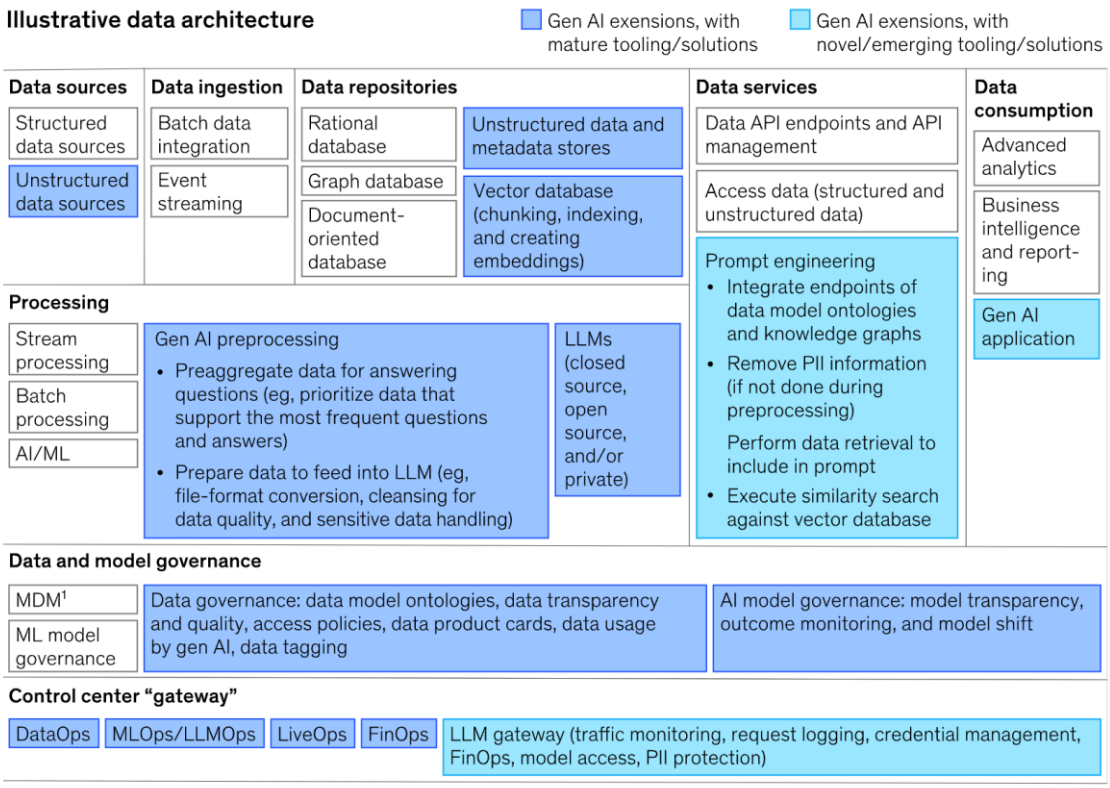

Figure 3 Fuelling Generative AI (Source: McKinsey [65])

Given these developments, it becomes essential to update the government data value cycle so as to reflect the new demands of Generative AI. In the next sub-section the renewed government data value cycle, i.e., the *AIGOV Government Data Value Cycle*, is presented.

4.1 The AIGOV Government Data Value Cycle

This sub-section presents the AIGOV Government Data Value Cycle. The cycle specifies in detail the steps that government data should go through in order to facilitate public value creation through AI. Towards this direction, the data types available in the public sector and which are important in AI scenarios are identified. Thereafter, the requirements, challenges, and guidelines for each step are presented.

4.1.1 Data types

The data types which are available in the public sector and are important for AI scenarios include may be classified into three categories namely, dynamic data, textual/unstructured data, structured data.

- Textual/Unstructured data. These include unstructured or semi-structured data including clinical notes and Electronic Health Records (EHR), police records and reports, regulatory and compliance reports like inspection reports, and social media

data. They are usually available in pdf or doc formats. Textual data are important for training and fine-tuning Large Language Models.

- Structured data often stored in databases or spreadsheets. These include, for example, demographic and census data, education data, e.g., student enrolments, historical weather data, and geospatial data.
- Dynamic data including environmental, traffic, satellite, meteorological, and sensor generated data. Dynamic data have been recently recognized by the European Commission as an important part of Open Government Data presenting huge potential economic value [23]. It is indicative that the majority of the national OGD portals disseminate dynamic data [71]. The immediate availability and regular updates of these data are crucial for the creation of added value data-driven services and applications [23].

4.1.2 The AIGOV Government Data Value Cycle

The AIGOV Government Data Value Cycle (Figure 4) comprises seven steps, namely

(1) data collection;

(2) data curation;

(3) data integration and linking;

(4) data storing;

(5) data dissemination;

(6) data usage; and

(7) data value creation.

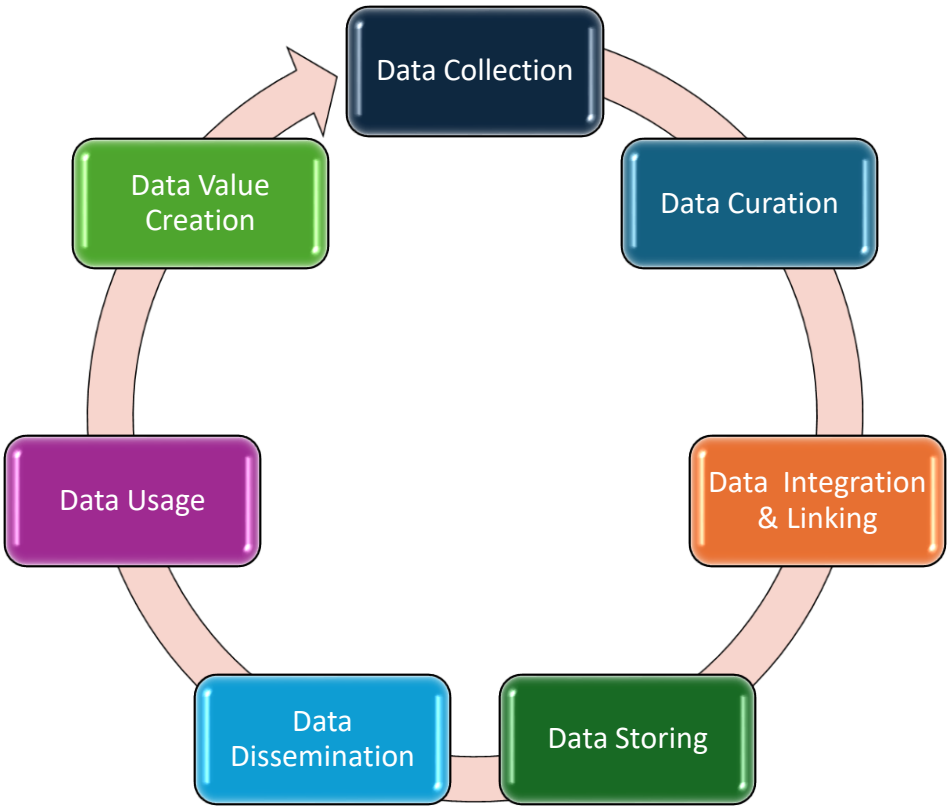

Figure 4 The AIGOV Data Value Cycle

4.1.3 Data Collection

This step is responsible for managing access to diverse and secure data sources and ensuring a **social licence** for data access and reuse.

Requirements:

- R1.1 Access to diverse data. Government and other data may be isolated in data silos as open government data published in various official data portal, or as closed data in various public services. Governments need to be able to access these data.
- R1.2 Access to unstructured data. The big change that Generative AI has brought when it comes to data is that the scope of value has gotten much bigger because of Generative AI's ability to work with unstructured data, including chats, videos, and code. This represents a significant shift because public organizations have traditionally had capabilities to work with only structured data, such as data in tables. This gives the chance to exploit, apart from structured data (e.g., data in tables), also unstructured data stored mainly in text documents that are currently unused. Thus, it should be possible for public organizations to publish and manage qualitative structured and unstructured data (such as text).
- R1.3 Access to dynamic data. Dynamic data (including environmental, traffic, and sensor data) were recently recognized as an important part of OGD [8]. For example,

dynamic (or real-time) data with traffic-related information (e.g., counted number of vehicles, average speed) that are generated by sensors have only recently started being provided as OGD available for free access and reuse.

Challenges:

- Ch1.1 Difficult access to data. Accessing and collecting isolated government (including dynamic) and other data of different types and formats is challenging.
- Ch1.2 High variability and rapid obsolescence of dynamic data. Dynamic data are characterized by their high variability and rapid obsolescence, making their immediate availability and regular updates crucial for the creation of added-value services and applications. Hence, collecting and reusing this type of data is challenging.
- Ch1.3 Low quality of dynamic data (also a challenge in data curation). Sensors are prone to malfunctions caused by, e.g., bad weather and temperature conditions, resulting in anomalous observations [58] and, hence, low quality of this kind of data.
- Ch1.4 Data protection and legal concerns. Generative AI can affect data privacy, though, especially when its models' training involves massive datasets that typically contain personal information. Inadequate management can expose sensitive information, resulting in serious privacy concerns.

Guidelines:

- G1.1 Use standards to structure government data used in AI.
- G1.2 Use Application Programming Interfaces (APIs) to provide access to government data including dynamic data.
- G1.3 Implement dynamic data pipelines in order to enable continuous data ingestion and processing.

4.1.4 Data Curation

Data curation refers to the set of activities required to ensure that data are accurate, reliable, representative, and ready for reuse in AI applications. It encompasses cleaning, validation, annotation, enrichment, and documentation, and it underpins the transparency and trustworthiness of AI systems deployed in public administration.

High-quality curation is essential to reduce bias, guarantee compliance with data protection regulations, and sustain interoperability across domains. In the AIGOV context, curation combines automated techniques (e.g., data profiling, imputation, augmentation) with human oversight to maintain both efficiency and accountability.

Requirements:

- R2.1 Ensure data quality and completeness.
- R2.2 Understanding data characteristics, limitations, and potential biases.
- R2.3 Provide unbiased and representative datasets.

- R2.4 sDocument provenance and transformations.

Challenges:

- Ch2.1 Low or incosnistent quality of dynamic data. Sensors are prone to malfunctions caused by, e.g., bad weather and temperature conditions, resulting in anomalous observations [58] and, hence, low quality of this kind of data.
- Ch2.2 Scarcity of large, high-quality textual datasets. Generative AI requires substantial and diverse corpora; public sector data availability remains limited compared to the private domain.
- Ch2.4 Bias detection and mitigation remain complex. Biases may stem from historical data collection practices or social inequalities embedded in the data.
- Ch2.5 Resource constraints and lack of skilled personnel. Data curation often competes with operational priorities, and specialized AI data-engineering skills are scarce in the public sector.

Guidelines:

- G2.1 Establish a data-quality management framework. Define measurable quality indicators (accuracy, completeness, timeliness, consistency) and monitor them periodically.
- G2.2 Combine automation with human-in-the-loop validation. Use automated cleaning and anomaly-detection tools while maintaining expert supervision for critical datasets.
- G2.3 Apply data-imputation and augmentation techniques. Address incomplete or small datasets using statistical imputation, synthetic data generation, or controlled data augmentation for textual corpora.
- G2.4 Implement verifiable and repeatable curation workflows. Maintain logs and metadata describing each modification, ensuring reproducibility and auditability.
- G2.5 Promote ethical and unbiased data practices. Integrate fairness checks, bias detection algorithms, and human review throughout the curation pipeline.
- G2.6 Invest in data-literacy and capacity-building. Train data stewards, AI engineers, and public servants to understand data provenance, bias, and quality assurance processes.
- G2.7 Leverage scalable, interoperable tools. Employ machine-learning-based curation, crowdsourcing, and collaborative platforms consistent with European interoperability standards (e.g., SEMIC, W3C).

4.1.5 Data Integration & Linking

Data integration and linking enable public administrations to break down information silos and to obtain a unified, cross-sectoral view of public sector operations. By connecting

heterogeneous datasets governments can generate deeper analytical insights, support evidence-based policymaking, and enhance interoperability.

In the AIGOV context, integration is understood as both a technical process (aligning formats, schemas, and identifiers) and an institutional process (establishing governance, trust, and accountability in data sharing). This stage is critical for realizing the promise of AI, which depends on comprehensive, consistent, and semantically rich data sources.

Requirements:

- R3.1 Interoperability of heterogeneous data sources. Integrate open and closed government datasets, as well as relevant private-sector or research data, through harmonized standards and metadata models.
- R3.2 Adoption of common data models and ontologies. Apply shared vocabularies such as SEMIC Core Vocabularies, DCAT-AP, and W3C RDF/OWL to ensure machine-readable, semantically aligned datasets.
- R3.3 Trusted data-sharing frameworks. Establish clear legal, organizational, and technical agreements for sharing data across institutions in compliance with GDPR, the Data Governance Act, and national data-sharing legislation.
- R3.4 Privacy-preserving integration. Implement data-linkage methods (e.g., pseudonymisation, federated queries) that allow analytical use of sensitive data without compromising confidentiality.

Challenges:

- Ch3.1 Heterogeneity of data formats and legacy systems. Disparate systems often use incompatible structures, making automated linkage complex and resource-intensive.
- Ch3.2 Data ownership and stewardship ambiguities. Institutional reluctance or unclear mandates can hinder cross-agency data exchange.
- Ch3.3 Legal and ethical constraints on sensitive data. Integrating personal or confidential information must respect strict privacy safeguards and accountability mechanisms.
- Ch3.4 Limited technical capacity. Many administrations lack semantic-integration tools and skilled personnel to maintain linked data infrastructures.

Guidelines:

- G3.1 Adopt semantic-interoperability standards. Use the European Interoperability Framework (EIF) principles to guide integration strategies.
- G3.2 Use Linked Data and API-based architectures. Implement RESTful APIs and Linked Data principles (URIs, RDF, SPARQL) to enable discoverable, machine-readable, and reusable data connections.

- G3.3 Establish data-governance agreements. Define roles (data owner, provider, consumer), access rights, and quality assurance responsibilities in formal memoranda or data-sharing contracts.
- G3.4 Ensure privacy-preserving linkage. Apply techniques such as federated learning, homomorphic encryption, or synthetic data generation to analyze sensitive data without direct exposure.
- G3.5 Foster trusted partnerships. Build collaborative ecosystems with academia, industry, and civic organizations to co-develop interoperable datasets and promote open innovation.
- G3.6 Monitor integration performance. Regularly evaluate interoperability metrics (including latency, completeness, and consistency) to improve data-sharing efficiency and reliability.

4.1.6 Data Storing

The Data Storing stage of the AIGOV Government Data Value Cycle focuses on the secure, sustainable, and efficient preservation of data assets. It ensures that government data are accessible, reliable, and interoperable over time, supporting both operational continuity and the training of AI systems.

This stage involves selecting appropriate storage architectures, implementing robust security and access controls, maintaining metadata and provenance records, and ensuring compliance with data protection regulations such as the GDPR.

As AI technologies evolve, particularly with the advent of Generative AI and Large Language Models (LLMs), traditional storage architectures must adapt. In addition to conventional relational or document databases, vector databases are increasingly required to store data embeddings, allowing semantic search and efficient retrieval of contextually relevant information.

Requirements:

- R4.1 Long-term, secure, and scalable storage. Public administrations must adopt storage infrastructures capable of handling growing data volumes while ensuring security, redundancy, and continuity of access.
- R4.2 Compliance with legal and ethical obligations. Storage solutions must conform to data protection and confidentiality requirements, including data retention and deletion policies aligned with the GDPR and national laws.
- R4.3 Metadata and provenance management. Each dataset must be accompanied by detailed metadata describing its source, structure, updates, and transformations to ensure traceability and reusability.

- R4.4 Support for advanced AI storage architectures. Systems must be designed to accommodate AI-ready data formats, including vector representations, embeddings, and multimodal data required by LLMs.

Challenges:

- Ch4.1 Evolving infrastructure needs. Traditional data warehouses are often inadequate for AI-driven analytics, requiring the integration of hybrid architectures combining relational, graph, and vector databases.
- Ch4.2 Managing sensitive or classified data. Government data may include personal, financial, or security-related information that demands tiered access and encryption mechanisms.
- Ch4.3 Ensuring interoperability and portability. Data stored in heterogeneous systems can become siloed, hindering cross-agency analytics and reuse.
- Ch4.4 Implementing vector databases for generative AI. Deploying and maintaining vector databases tailored to LLMs presents new technical challenges related to storage efficiency, indexing, and privacy of embeddings.
- Ch4.5 Balancing accessibility with sustainability. Continuous data replication and high availability increase costs and environmental footprint, requiring energy-efficient solutions.

Guidelines:

- G4.1 Prioritize privacy, security, and compliance. Adopt privacy-by-design principles; enforce role-based access control (RBAC), data encryption, and regular security audits to protect stakeholders' trust.
- G4.2 Implement efficient data discovery mechanisms. Maintain centralized metadata catalogues and data registries using open standards (e.g., DCAT-AP) to enable users to locate and access datasets easily.
- G4.3 Track and document data provenance. Record data origin, transformations, and lineage within metadata systems to ensure accountability and reproducibility in AI applications.
- G4.4 Strengthen data governance frameworks. Define clear responsibilities for data stewardship, backup management, retention, and archival processes under an overarching data governance policy.
- G4.5 Integrate advanced storage solutions for AI. Employ vector databases and embedding indexes to support semantic search and retrieval, enabling generative AI systems to access only the most relevant information (e.g., retrieving key passages from lengthy documents).
- G4.6 Ensure interoperability and sustainability. Favor open, cloud-agnostic storage solutions compatible with the European Data Spaces initiative, and monitor their environmental impact.

4.1.7 Data Dissemination

Data dissemination represents the stage where public sector data transitions from being an internal administrative resource to a shared public asset. It focuses on making data discoverable, accessible, and reusable by stakeholders, including citizens, researchers, private organizations, and other administrations.

In the AIGOV framework, dissemination must balance openness with responsibility, ensuring transparency and innovation while safeguarding privacy, security, and ethical integrity. Well-designed dissemination practices enable Generative AI systems and other advanced analytics tools to leverage high-quality public data responsibly, contributing to more efficient and evidence-based governance.

Requirements:

- R5.1 Enable broad yet controlled data access. Governments should ensure that datasets are accessible to authorized users and the public, following the principle "as open as possible, as closed as necessary."
- R5.2 Ensure data interoperability and discoverability. Disseminated data must adhere to standardized metadata schemas (e.g., DCAT-AP) and interoperability guidelines (e.g., SEMIC, W3C) to facilitate cross-platform reuse.
- R5.3 Provide clear licensing and usage rights. Data releases should specify usage conditions through open licenses such as Creative Commons or Open Data Commons, ensuring legal clarity for reuse.
- R5.4 Maintain accessibility and inclusiveness. Dissemination platforms should comply with accessibility standards (WCAG 2.1) and support multilingual and machine-readable formats.
- R5.5 Ensure ethical publication of data for AI. Public administrations must avoid releasing sensitive or identifiable data that could be misused for discriminatory or unethical AI applications.

Challenges:

- Ch5.1 Balancing openness and protection. Determining what can be safely shared without compromising privacy, intellectual property, or national security remains a persistent challenge.
- Ch5.2 Data fragmentation across portals and platforms. Many administrations host datasets across disparate systems, making them difficult to discover or aggregate.
- Ch5.3 Inconsistent metadata quality. Insufficient or non-standardized metadata reduces the usability and visibility of public data.
- Ch5.4 Limited capacity for real-time or dynamic data publishing. Disseminating live data streams (e.g., environmental or traffic data) requires scalable infrastructures and automated validation pipelines.

- Ch5.5 Trust and provenance. Users need to be confident that disseminated data is authentic, up to date, and traceable to its source.

Guidelines:

- G5.1 Apply open data principles systematically. Adopt the Open Data Charter and the EU Directive on Open Data (2019/1024) to guide publication policies and ensure consistency across administrations.
- G5.2 Use standardized metadata and identifiers. Describe datasets using DCAT-AP, include persistent identifiers (PIDs/DOIs), and register them in national and European data catalogues.
- G5.3 Provide machine-readable, multilingual, and API-enabled access. Publish data in open formats (e.g., CSV, JSON, RDF) and enable retrieval via APIs to facilitate reuse by AI and data analytics systems.
- G5.4 Establish data release workflows with quality assurance. Implement pre-publication checks for accuracy, privacy, and format compliance. Maintain version control and update logs.
- G5.5 Ensure security and ethical oversight. Conduct data disclosure risk assessments and, where applicable, use differential privacy or anonymization techniques before release.
- G5.6 Promote awareness and capacity building. Provide guidance, documentation, and training to help civil servants, developers, and researchers use public data responsibly.
- G5.7 Encourage feedback and collaboration. Incorporate mechanisms for users to report issues, suggest improvements, or contribute to data enrichment - fostering a participatory open data ecosystem.

4.1.8 Data Usage

The Data Usage stage represents the point where government data are transformed into actionable insights and public value. This stage connects the technical dimensions of data management with the societal, ethical, and political context in which data-driven decisions are made.

In the public sector, using data for AI and analytics requires more than technical readiness; it depends on earning and maintaining a social licence, the approval and trust of stakeholders, citizens, and policymakers. Achieving this trust involves demonstrating that data are used transparently, ethically, and for legitimate public-interest purposes.

The AIGOV framework promotes a human-centric, participatory, and responsible approach to data usage, ensuring that AI systems and analytics serve collective wellbeing, uphold democratic accountability, and maintain public confidence.

Requirements:

- R6.1 Secure a social licence for data-driven initiatives. Public administrations must ensure that citizens, civil society, and other stakeholders understand and consent to how their data are being used for policy design, service delivery, or AI decision-making.
- R6.2 Ensure purposeful and ethical analytics. Data use should be aligned with clearly defined objectives that serve the public interest, avoiding unnecessary or intrusive data processing.
- R6.3 Guarantee transparency and accountability in decision-making. AI-assisted analytics must be explainable, traceable, and auditable, enabling oversight by internal and external authorities.
- R6.4 Foster human oversight and interpretability. Decisions supported by AI systems should always allow human review and intervention to safeguard fairness and responsibility.
- R6.5 Promote workforce competence and data literacy. Civil servants and decision-makers should be equipped with the skills to interpret, question, and responsibly apply data insights.

Challenges:

- Ch6.1 Engaging diverse stakeholders. Establishing trust requires proactive communication and participatory governance, particularly when using data that affect citizens directly.
- Ch6.2 Reconciling conflicting interests. Data usage may raise tensions between efficiency, privacy, and equity; balancing these priorities demands clear ethical frameworks and political commitment.
- Ch6.3 Limited transparency of AI models. Complex algorithms, especially large language models, can obscure decision rationales, complicating accountability.
- Ch6.4 Insufficient organizational culture of data ethics. Many public institutions lack formal mechanisms for ethical review, continuous monitoring, or redress in AI-driven decision processes.
- Ch6.5 Uneven levels of data literacy. Without adequate skills, civil servants may misinterpret data or overly rely on automated recommendations.

Guidelines:

- G6.1 Obtain and sustain social licence. Engage stakeholders early, communicate the purpose and benefits of data use, and provide channels for feedback, complaints, and dialogue.

- G6.2 Pursue purposeful and proportionate analytics. Clearly define the societal problem each analytics or AI application aims to address, ensuring proportionality between data use and public benefit.
- G6.3 Embed ethics and accountability frameworks. Implement ethical impact assessments, algorithmic transparency registers, and oversight committees for AI deployments in government.
- G6.4 Promote human oversight. Maintain "human-in-the-loop" systems to verify, interpret, and approve outputs of AI models, particularly in high-stakes contexts (e.g., healthcare, justice, welfare).
- G6.5 Build data literacy and analytical capacity. Train public sector staff in interpreting data responsibly, understanding AI limitations, and recognizing bias and uncertainty.
- G6.6 Communicate results transparently. Publish methods, findings, and impacts of data use in accessible formats to strengthen public trust and accountability.
- G6.7 Encourage cross-sector collaboration. Work with academia, civil society, and industry to co-create ethical frameworks and share lessons learned from data-driven projects.

4.1.9 Data Value Creation

The final stage of the AIGOV Government Data Value Cycle focuses on the creation of public value from data and AI. It represents the point where well-governed, interoperable, and ethically managed data are transformed into tangible outcomes including improved public services, informed policymaking, increased administrative efficiency, and enhanced citizen trust.

In the context of Generative AI and Large Language Models, this phase involves not only extracting insights from data but also developing and deploying AI systems that can generate content, automate knowledge discovery, and support decision-making. Realizing this potential requires strategic planning, ethical safeguards, and institutional capacity building to ensure that value creation remains aligned with democratic principles, fairness, and societal goals.

Requirements:

- R7.1 Evolve organizational data architecture. Public administrations must periodically assess and modernize their data infrastructure to enable efficient, secure, and scalable support for AI and Generative AI use cases.
- R7.2 Deploy AI and LLM solutions tailored to the public sector. AI systems must be fine-tuned to reflect the linguistic, legal, and cultural context of public administration while ensuring explainability and accountability.
- R7.3 Ensure multilingual and inclusive AI systems. Given Europe's linguistic diversity, AI models should support multiple languages and be inclusive of different cultural and social contexts.

- R7.4 Integrate fairness, transparency, and explainability. Data-driven value creation must include mechanisms to detect and mitigate bias, document decision logic, and provide transparent explanations of algorithmic outputs.
- R7.5 Build human and institutional capacity. Establish specialized roles (e.g., Head of AI, AI Ethics Officer, Prompt Engineer) and develop training programs that strengthen data literacy and ethical awareness across the public sector.
- R7.6 Foster cross-sector collaboration and knowledge transfer. Value creation is maximized when governments collaborate with academia, startups, and civic organizations to co-develop AI applications addressing shared challenges.

Challenges:

- Ch7.1 Identifying high-value and low-risk AI use cases. Determining where AI truly adds value requires careful prioritization.
- Ch7.2 Managing infrastructure and cost constraints. Deploying and maintaining AI models, particularly LLMs, demands significant computational and financial resources, including data storage, GPUs, and secure cloud environments.
- Ch7.3 Ensuring fairness and transparency of algorithms. “Black-box” AI models risk producing opaque or biased results that undermine accountability and citizen trust.
- Ch7.4 Limited institutional readiness. Many administrations lack the technical and organizational frameworks to manage AI at scale, including governance, procurement, and lifecycle monitoring mechanisms.
- Ch7.5 Regulatory uncertainty. The evolving European AI Act and related policies introduce compliance challenges that must be addressed in design and deployment phases.

Guidelines:

- G7.1 Assess and evolve data architecture. Evaluate current data systems and plan necessary upgrades (e.g., data pipelines, APIs, storage formats) to support AI integration and scalability.
- G7.2 Identify and prioritize use cases strategically. Not all challenges require AI solutions. Apply frameworks (such as McKinsey’s 4Cs: complexity, criticality, creativity, and compliance) to prioritize applications based on potential impact, feasibility, and risk.
- G7.3 Deploy and fine-tune LLMs for public-sector needs. Use open or proprietary models fine-tuned with government-specific datasets. Ensure adherence to data sovereignty requirements and minimize dependence on private-sector vendors when possible.

- G7.4 Support multilingualism and inclusivity. Implement models capable of understanding and generating content in multiple languages to ensure accessibility for all citizens.
- G7.5 Facilitate LLM integration through frameworks. Use integration libraries such as LangChain or LlamaIndex to connect models with structured datasets, internal document repositories, and public services.
- G7.6 Maintain human oversight and accountability. Keep “humans in the loop” to validate AI outputs, especially in sensitive policy or service contexts. Assign clear accountability for AI decisions.
- G7.7 Establish comprehensive risk management. Define organizational risk posture, conduct AI impact assessments, and implement mitigation measures covering ethical, operational, and security risks.
- G7.8 Build organizational skills and roles. Create dedicated positions for AI strategy, engineering, and ethics, ensuring alignment between technical and governance domains.
- G7.9 Develop applications jointly with end users. Involve public servants and citizens early in co-design processes to improve model performance, contextual accuracy, and social acceptability.
- G7.10 Communicate transparently. Develop communication plans clarifying the scope, limitations, and safeguards of AI systems to foster informed trust among stakeholders.
- G7.11 Start small and scale responsibly. Pilot projects in controlled settings before broad deployment. Evaluate impacts, refine models, and expand only once benefits and risks are fully understood.

5 The AIGOV Framework for Trustworthy, Fair, and Accountable AI

This Section presents the AIGOV Framework for Trustworthy, Fair, and Accountable Artificial Intelligence (AI) with emphasis on Generative AI. The framework is based on four main pillars of trustworthy, fair, and accountable AI, namely

1. Transparency and accountability,
2. Data management and access,
3. Comprehensibility and multilingual support, and
4. Interoperability and reusability.

These four pillars were derived through a review of international policy frameworks, ethical guidelines, and academic literature addressing the responsible use of AI in the public sector.

The rest of this section presents guidelines (eight in total) for each of the four pillars. Specifically, the first pillar includes two guidelines, the second pillar three guidelines, the third pillar two guidelines, and the fourth pillar one guideline.

5.1 Pillar 1: Transparency and accountability

Transparency and accountability are widely recognized as foundational principles for trustworthy AI. The European Commission's Ethics Guidelines for Trustworthy AI [23], the OECD Principles on AI[10], and leading ethical frameworks [27] [41] emphasize that AI systems must be traceable, explainable, and subject to oversight.

In order to build trust in AI and, especially in Generative AI, the created models need to be must be created assuring their transparency and explainability in order to mitigate AI-associated risks like bias. In addition, to enhance transparency, AI models should be open, allowing to everyone access to the model's architecture, data sources, and methodologies used during its creation, training, or fine-tuning. Together, these principles mitigate risks of bias, error, and misuse. Accordingly, this pillar focuses on developing explainable and open AI models, reinforcing human oversight and institutional responsibility.

Guideline 1. AI models should be explainable.

Methods and Assessment tools for Guideline 1: Tools and libraries that enable understand, interpret, and explain the results of AI models. These include, for example, the SHapley Additive exPlanations (SHAP) framework for tree-based models (e.g., XGBoost, LightGBM), deep learning models (e.g., GNNExplainer), and general machine learning models, and the Local Interpretable Model-agnostic Explanations (LIME).

[10] https://www.oecd.org/en/topics/ai-principles.html

Guideline 2. AI models and systems, their architecture, data sources, and methodologies used during their creation, training, or fine-tuning should be open to all stakeholders.

Methods and Assessment tools for Guideline 2: AI Models and datasets used should be publicly available in public hubs such as GitHub, Hugging Face, open data portals.

5.2 Pillar 2: Responsible data management and access

Responsible data management is central to the development of reliable and equitable AI. The OECD Digital Government Policy Framework [74], the European Data Strategy, and the EU Data Governance Act[11] highlight that AI systems depend on data that are accurate**,** timely**,** governed, and accessible. In addition, the World Bank's *Data for Better Lives* [123] underline that ethical data practices underpin public trust in AI. This pillar therefore translates policy principles into practical guidance on data quality assurance, granularity, and timely access to dynamic datasets through interoperable APIs.

Guideline 3. Data used in AI should be accurate and timely.

Methods and Assessment tools for Guideline 3: In order to assess the quality of data include used in AI statistical testing (e.g., using Python or R) can be employed to calculate key metrics like mean, variance, standard deviation, skewness etc., and also calculate null values, find outliers, data duplicates and inconsistencies. Machine learning anomaly detection techniques such as isolation forest have also been used in literature to explore the quality including the accuracy and timeliness of sensor data [47]. Finally, the timeliness of the data can also be accessed directly from the metadata, such as timestamps of the latest updates and by monitoring the frequency and consistency of data updates.

Guideline 4. Data used in AI should be provided at various levels of granularity.

Methods and Assessment tools for Guideline 4: The granularity of data can be evaluated through the exploration of the data using tools like Tableau and Power BI at various levels of aggregations and the application of Online AnaLytic Processing (OLAP) functions like drill down and roll up to verify the availability of granular data. The same exploration can be also made through Python and related packages like pandas. In case that data are provided through an API, API monitoring tools like Postman can be also employed to evaluate their granularity.

Guideline 5. Access on dynamic data (e.g., sensor data) should be enabled through an Application Programming Interface (API).

Methods and Assessment tools for Guideline 5: In order to assess the functionality of provided APIs, tools like Postman and Swagger can be used, for example, to test endpoints with

[11] https://digital-strategy.ec.europa.eu/en/policies/data-governance-act

different parameters and ensure they deliver dynamic data correctly. Real-Time API Monitoring Tools like APImetrics can be also used to monitor API performance and functionality.

5.3 Pillar 3: Comprehensibility and multilingual support

AI technologies must serve all citizens, regardless of language, culture, or educational background. The UNESCO Recommendation on the Ethics of Artificial Intelligence [114] emphasizes inclusiveness, linguistic diversity, and accessibility as ethical imperatives. This pillar therefore ensures that Generative AI outputs are understandable, linguistically inclusive, and culturally sensitive, supporting equitable access to AI-enabled public services and deliberative democratic processes.

Guideline 6. Responses of (Generative) AI should be comprehensible to all stakeholders speaking multiple languages and with different cultural and educational backgrounds.

Methods and Assessment tools for Guideline 6: Evaluate the accuracy and overall quality of multilingual responses when training and fine-tuning Large Language Models in various languages. In addition, employ results of recent studies (e.g., [103]) and recent research projects' results (e.g., TrustLLM[12]) and adapt them to each case's unique requirements and challenges in order to evaluate and analyze the trustworthiness of LLMs.

Guideline 7. AI tools should facilitate the retrieval of information from vast document collections to support evidence-based AI decision-making.

Methods and Assessment tools for Guideline 7: XAI frameworks and methods could be used to ensure the transparency of the decision-making process, making it easier to trace how the retrieved information is used to support evidence-based decisions.

5.4 Pillar 4: Data interoperability and reusability

Interoperability and reusability are essential for ensuring that AI systems are sustainable**,** efficient, and cross-sectorally connected. The European Interoperability Framework[13] stresses the need for data to be Findable, Accessible, Interoperable, and Reusable (also following the FAIR principles). By promoting open standards, metadata harmonization, and linked-data models, public administrations can enhance transparency, reduce duplication, and facilitate collaboration across agencies. This pillar embeds these principles into AI design, ensuring that LLMs and AI systems can access, share, and reuse data across administrative boundaries**,** thereby increasing both efficiency and accountability.

[12] https://trustllm.eu/
[13] https://ec.europa.eu/isa2/eif_en/

Guideline 8. AI models, and especially Large Language Models (LLMs), should be built on interoperable datasets, enabling seamless integration, sharing, and reuse across various applications and systems.

Methods and Assessment tools for Guideline 8: Data can be assessed using the FAIR Data Principles (Findable, Accessible, Interoperable, Reusable)[14]. In addition, AI models can be trained on already interoperable datasets such as, for example, datasets modelled following the linked data principles.

[14] FAIRsharing.org

6 The AIGOV Transformation and Adoption Framework

The AIGOV Transformation and Adoption Framework aims to facilitate public authorities to explore how public services will need to be redesigned to leverage the impact of AI. It supports public administrations to determine current strengths and weaknesses, set achievable goals, and construct transformation plans by taking into account all the AIGOV methods, guidelines, and tools in order to achieve ethical, trustworthy, and fair adoption of AI technologies, with particular emphasis on Generative AI.

The AIGOV Transformation and Adoption Framework integrates the AIGOV Government Data Value Cycle (T2.1) and the AIGOV Framework for Trustworthy, Fair, and Accountable AI (T2.2), translating these into actionable steps to guide ethical, fair, and sustainable AI adoption in the public sector.

6.1 Background: Public Service Provision and the Role of AI

Public service provision in government typically consists of two complementary phases [83]:

(1) the informative phase, during which citizens seek information about a service (such as eligibility criteria, required documents, or procedural steps), and

(2) the performative phase, where the request is processed, validated, and ultimately executed by the public administration.

Traditionally, the informative phase is considered more straightforward to automate, as reflected in maturity models of e-government where the provision of online information represents the earliest and most attainable stage of sophistication. However, modern public service ecosystems are characterized by high levels of variation: many services exist in hundreds of context-specific versions, depending on citizen profiles, administrative jurisdiction, legal conditions, and situational triggers. As a result, delivering accurate and personalized information, even in the informative phase, remains a significant challenge for public administrations.

The concept of targetization, often associated with the advanced stages of digital government, refers to the ability of administrations to tailor information and service execution to the specific circumstances of a citizen or business. While historically linked to personalization during the performative phase, advances in Generative AI and natural language understanding now enable meaningful personalization in the informative phase as well. This shift demonstrates how AI technologies, particularly Large Language Models, can reshape not only service execution but also how citizens understand, navigate, and engage with public services.

Recognizing these dynamics is essential for the AIGOV Transformation and Adoption Framework, as the redesign of public services must consider how AI affects and enables both

phases of service delivery, while ensuring accessibility, fairness, trustworthiness, multilingual support, and compliance with legal and ethical standards.

6.2 The AIGOV Transformation and Adoption Framework

The AIGOV Transformation and Adoption Framework consists of five phases, each supported by tools, guidelines, and decision checkpoints:

Phase 1: Assess Readiness and Context. The first phase of the AIGOV Transformation and Adoption Framework focuses on determining whether the introduction of AI is justified, feasible, and aligned with public values and legal requirements. Rather than assuming AI as a default solution, this phase ensures that public authorities clearly understand the problem space, their organisational maturity, and the risks and prerequisites associated with AI deployment. This prevents premature or inappropriate adoption while enabling informed decision-making and responsible planning.

The assessment considers four core dimensions:

1. **Strategic Context and Public Value** to evaluate whether AI is the most appropriate approach to address the identified need, considering purpose, expected benefits, proportionality, and alignment with public service priorities.
2. **Data Readiness** that analyses the availability, quality, accessibility, interoperability, and legal compliance of the data required to support AI, drawing on the principles of the AIGOV Government Data Value Cycle.
3. **Governance, Ethics, and Legal Compliance**: ensuring alignment with the principles of trustworthy, fair, and accountable AI, including transparency, explainability, inclusion, privacy protection, and clear allocation of accountability.
4. **Organisational and Capability Capacity:** assessing whether the administration possesses the necessary technical infrastructure, skills, leadership support, and change-readiness to adopt and manage AI sustainably.

The output of this phase is a structured AI Readiness Assessment, summarising strengths, gaps, risks, and feasibility. Based on this assessment, the administration determines whether the initiative should proceed to the next phase, be postponed until prerequisites are met, or be redirected toward an alternative non-AI solution.

Phase 2: Design Ethical and Value-Aligned AI Use Cases

Once readiness has been established, Phase 2 of the framework focuses on defining how AI can be applied to redesign or improve public services in a way that aligns with public values, legal frameworks, and societal expectations. This phase ensures that AI is not only technically feasible but also meaningful, equitable, and designed with citizens and public servants in mind.

During this phase, public administrations clarify the scope, purpose, and expected outcomes of the AI intervention. Rather than focusing solely on efficiency or automation, the design process emphasises human-centric and public value-oriented service redesign. This includes prioritising accessibility, fairness, multilingual support, transparency, and the right level of human oversight.

The design process incorporates three key activities:

- Problem and Service Redesign Definition: reframing the service challenge based on user needs, service delivery realities, and the insights generated from Phase 1. Tools such as service blueprints, process mapping, and problem framing help ensure clarity and shared understanding among stakeholders.
- Responsible AI Use Case Design: translating the identified needs into concrete AI-enabled use cases while applying the principles outlined in the AIGOV Framework for Trustworthy, Fair, and Accountable AI. This includes defining explainability requirements, expected risks and safeguards, ethical constraints, human-in-the-loop roles, and inclusiveness criteria such as multilingual accessibility.
- Success Metrics and Impact Definition: establishing clear evaluation criteria, including Key Performance Indicators (KPIs) for fairness, transparency, performance, accessibility, quality of service, and citizen experience. These metrics will later guide testing, monitoring, and iterative improvement.

Phase 2 ensures that AI-driven innovation is intentional, meaningful, and aligned with societal values. By anchoring the design in public sector responsibilities and ethical requirements, administrations can enter the development phase with a clear and responsible roadmap.

Phase 3: Build and Test

Phase 3 focuses on transforming the designed AI use case into a functional prototype and evaluating its performance, usability, and compliance with ethical and legal standards. This phase enables public administrations to explore the behaviour and real-world implications of AI before committing to full-scale deployment. It also ensures that any risks identified during earlier stages are proactively mitigated through testing, evaluation, and refinement.

Development in this phase follows an iterative approach, informed by user feedback, legal compliance, and continuous validation of expected benefits. Human oversight remains central: public servants, domain experts, and affected stakeholders engage throughout the testing and validation process to ensure the system performs reliably, transparently, and fairly.

Key activities in this phase include:

- Prototype Development and Technical Implementation: creating an initial functional version of the AI system using available data, technological components, and integration requirements identified in earlier phases. Reusable assets, interoperable components, and open standards should be prioritised to ensure future scalability.
- Ethical, Legal, and Safety Testing: validating the system against the AIGOV Framework for Trustworthy, Fair, and Accountable AI. This includes evaluating explainability, identifying potential bias, ensuring data governance compliance (including GDPR), and verifying accountability and human-in-the-loop mechanisms.
- User Testing and Feedback Integration: engaging civil servants, citizens, and relevant stakeholder groups in testing to assess usability, clarity, multilingual accessibility, perceived fairness, and alignment with service expectations. Insights gathered during testing inform refinement cycles.

The outcome of Phase 3 is a validated pilot-ready AI solution accompanied by a comprehensive evaluation report. The report documents system performance, user feedback, risk mitigation measures, ethical compliance, and identified limitations. This documentation ensures transparency and provides evidence to support decisions in later phases.

Phase 3 is essential for ensuring that the proposed AI solution is not only technically sound but also socially acceptable, lawful, understandable, and aligned with public sector values. By testing the system in controlled environments, administrations gain confidence and insights before moving toward full deployment and operational integration.

Phase 4: Deploy and Scale Responsibly

Phase 4 focuses on transitioning the validated AI solution from a controlled testing environment into operational use within the public administration. This phase ensures that deployment is deliberate, monitored, and supported by appropriate organisational, legal, and technical structures. Scaling is approached incrementally to minimise risk, preserve public trust, and ensure continuous ethical alignment.

Deployment is not only a technical task but also an organisational change process. It requires updating workflows, providing training, supporting end users, and ensuring that safeguards and accountability mechanisms remain active throughout the lifecycle of the solution. Public administrations must also ensure that the deployment complies with evolving regulatory requirements and interoperates effectively with existing systems and processes.

Key activities in this phase include:

- Operational Integration: embedding the AI solution into existing administrative processes, information systems, and service delivery channels. This includes ensuring interoperability, integration with data sources, and alignment with standardised public sector architectures and procurement requirements.

- Controlled Rollout and Monitoring: implementing the solution in stages, starting with limited deployment, monitoring real-world behaviour, and making iterative refinements. Monitoring includes performance metrics, fairness indicators, user satisfaction, explainability compliance, and error handling.
- Capacity Building and Support: equipping public servants, administrators, and end users with the skills required to operate and oversee the system. This may include training on human-AI collaboration, interpretation of outputs, escalation procedures, and ethical oversight.

The output of Phase 4 is a fully deployed AI-enabled public service operating with clear governance structures, continuous monitoring mechanisms, and established accountability pathways. A deployment record is maintained to document decisions, safeguards, system changes, and compliance measures throughout implementation.

Phase 4 ensures that AI deployment in the public sector is responsible, transparent, and aligned with long-term societal and institutional expectations. It enables administrations not only to adopt AI technologies but to operationalise them in a way that is trustworthy, fair, and sustainable.

Phase 5: Govern, Evaluate, and Sustain

Phase 5 ensures that the AI solution remains trustworthy, effective, and aligned with public values throughout its operational life. Artificial Intelligence is not a static technology; models evolve, data changes, regulations develop, and societal expectations shift. Therefore, ongoing governance, evaluation, and continuous improvement are necessary to maintain accountability, transparency, safety, and public trust.

This phase establishes the structures and processes needed to monitor system performance, manage risks, respond to feedback, and adapt the AI solution as context and requirements evolve. It also supports long-term organisational learning and capability building, ensuring that AI becomes a responsible and sustainable component of public sector operations.

The phase focuses on three core activity streams:

- Lifecycle Governance and Compliance: maintaining mechanisms that ensure continued alignment with legal frameworks, ethical principles, and accountability provisions. This includes periodic audits, model documentation updates, version control, and validation against emerging AI regulations and national standards.
- Monitoring, Evaluation, and Improvement: continuously tracking system performance, equity, accuracy, explainability, and user experience using indicators defined earlier in the process. Evaluation findings inform corrective actions, retraining, model adaptation, or, when necessary, decommissioning.
- Organisational Learning and Sustainability: supporting long-term adoption through capability development, knowledge retention, and updated operating practices. This

includes refining guidance, updating training programmes, sharing lessons learned, and contributing to reusable assets and interoperable building blocks across public administration.

The output of Phase 5 is a sustained and responsibly managed AI-enabled public service supported by a structured governance model, documented oversight processes, and continuous alignment with ethical, societal, legal, and operational requirements.

Phase 5 ensures that AI deployment does not end with implementation. Instead, it establishes an ongoing commitment to stewardship, transparency, and public value, ensuring that AI remains a tool for strengthening democratic governance and high-quality public service delivery.

7 Conclusions

This deliverable presents the methodological, technical, and governance foundations required for the ethical, trustworthy, and effective adoption of Artificial Intelligence (AI), and particularly Generative AI, in the public sector. Building on international best practices, the evolving European regulatory landscape, and extensive research on AI, data management, and public administration, WP2 delivers three complementary artefacts that together form a comprehensive framework for responsible AI-enabled transformation.

First, the AIGOV Government Data Value Cycle provides a renewed, seven-step model for managing public sector data in ways that ensure quality, interoperability, accessibility, and readiness for AI. By integrating emerging requirements such as support for unstructured data, vector databases, dynamic data ingestion, and transparent curation workflows, the cycle establishes a robust foundation for data-driven public value creation.

Second, the AIGOV Framework for Trustworthy, Fair, and Accountable AI translates high-level policy and ethical principles, such as those from the EU AI Act, the OECD AI Principles, and the UNESCO Recommendation on AI, into eight practical guidelines covering transparency, data governance, multilingual support, explainability, interoperability, and responsible access. These guidelines help public administrations design AI systems that are reliable, understandable, and aligned with democratic values.

Third, the AIGOV Transformation and Adoption Framework offers a structured, five-phase model supporting public administrations throughout the AI adoption lifecycle, from assessing organisational readiness and designing value-aligned use cases to developing, deploying, and sustainably governing AI solutions. This framework ensures that AI is introduced only where appropriate, implemented safely, monitored continuously, and evaluated with respect to ethics, fairness, and societal impact.

Collectively, the outputs of WP2 position public administrations to take advantage of AI opportunities while managing associated risks. They emphasise the importance of high-quality data, human oversight, transparency, multilingual inclusiveness, ethical safeguards, and organisational capability building. These principles are essential for maintaining public trust and securing the legitimacy of AI-enabled public services.

WP2 therefore lays the groundwork for the practical experimentation and pilot deployment activities of WP3. By providing methodological clarity, actionable guidelines, and governance structures, this deliverable equips public authorities with the tools needed to implement AI responsibly, ensuring that new technologies strengthen, not compromise, public value, fairness, accountability, and democratic integrity.

References

1. Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., ... & McGrew, B. (2023). Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.
2. Anastasopoulos, A., Barrault, L., Bentivogli, L., et al. (2022) Findings of the iwslt 2022 evaluation campaign, in: International Workshop on Spoken Language Translation.
3. Artetxe, M., Ruder, S., & Yogatama, D. (2019). On the cross-lingual transferability of monolingual representations. arXiv preprint arXiv:1910.11856.
4. Azamfirei, R., Kudchadkar, S. R., & Fackler, J. (2023). Large language models and the perils of their hallucinations. *Critical Care*, *27*(1), 1-2.
5. Bapna, A., Arivazhagan, N., & Firat, O. (2019). Simple, scalable adaptation for neural machine translation. arXiv preprint arXiv:1909.08478.
6. Benchetrit, T., Kremer, I., Hemberg, E., Sankaranarayanan, A., & O'Reilly, U. M. (2024, February). Using a Large Language Model to Choose Effective Climate Change Messages. In *AAAI-2024 Workshop on Public Sector LLMs: Algorithmic and Sociotechnical Design*.
7. Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. Advances in neural information processing systems, 33, 1877-1901.
8. Brimos, P.; Karamanou, A.; Kalampokis, E.; Tarabanis, K. Graph Neural Networks and Open-Government Data to Forecast Traffic Flow. Information 2023, 14, 228. https://doi.org/10.3390/info14040228
9. Bruce, D., et al., Unlocking the potential of generative AI: Three key questions for government agencies (Dec. 7, 2023), https://www.mckinsey.com/industries/public-sector/our-insights/unlocking-the-potential-of-generative-ai-three-key-questions-for-government-agencies .
10. Cabinet Office. Generative ai framework for hm government, Jan 2024.
11. Cabinet Office and Innovation & Technology Department for Science, Mar 2024.
12. Caserta, H. Harreis, K. Rowshankish, N. Srinidhi, A. Tavakoli (2023). The data dividend: Fueling generative AI. Retrieved from https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-data-dividend-fueling-generative-ai
13. Chesnevar, C., Modgil, S., Rahwan, I., Reed, C., Simari, G., South, M., Vreeswijk, G., Willmott, S., et al. (2006) Towards an argument interchange format, The knowledge engineering review 21 293–316
14. Chen, X., Ye, J., Zu, C., Xu, N., Zheng, R., Peng, M., ... & Huang, X. (2023). How Robust is GPT-3.5 to Predecessors? A Comprehensive Study on Language Understanding Tasks. *arXiv e-prints*, arXiv-2303.
15. Chiarcos, C. (2012) Ontologies of linguistic annotation: Survey and perspectives., in: LREC, Citeseer, pp. 303–310.
16. Colombo, P., Pires, T. P., Boudiaf, M., Culver, D., Melo, R., Corro, C., Martins, A. FT, Esposito, F., Raposo, V. L., Morgado, S. et al. (2024) Saullm-7b: A pioneering large language model for law. arXiv preprint arXiv:2403.03883.
17. Conneau, A., Lample, G., Rinott, R., Williams, A., Bowman, S. R., Schwenk, H., & Stoyanov, V. (2018). XNLI: Evaluating cross-lingual sentence representations. arXiv preprint arXiv:1809.05053.
18. Dahl, M., Magesh, V., Suzgun, V., and Ho, D.E. (2024) Large legal fictions: Profiling legal hallucinations in large language models. arXiv preprint arXiv:2401.01301.

19. Davitti, E. (2013) Dialogue interpreting as intercultural mediation: Interpreters' use of upgrading moves in parent–teacher meetings, Interpreting 15 168–199.
20. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
21. de-Dios-Flores, I., Pichel Campos, J. R., Ioana Vladu, A., & Gamallo Otero, P. (2023). LANGUAGE TECHNOLOGIES FOR A MULTILINGUAL PUBLIC ADMINISTRATION IN SPAIN. *Journal of Language & Law/Revista de Llengua i Dret*, (79).
22. Doerr, N. (2012) Translating democracy: how activists in the european social forum practice multilingual deliberation, European Political Science Review 4 361–384.
23. European Commission: Directorate-General for Communications Networks, Content and Technology and Grupa ekspertów wysokiego szczebla ds. sztucznej inteligencji, Ethics guidelines for trustworthy AI, Publications Office, 2019, https://data.europa.eu/doi/10.2759/346720
24. European Parliament and European Council. 2019. Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast). Off. J. Eur. Union 172 (2019), 56–83.
25. Fitsilis, F. (2021). Artificial Intelligence (AI) in parliaments – preliminary analysis of the eduskunta experiment. The Journal of Legislative Studies, 27(4):621–633.
26. Fitsilis, F. and Theodorakopoulos, G. (2024). Better regulation and its evolution in the hellenic legislative and parliamentary system. Statute Law Review, 45(1):hmae003.
27. Floridi, L., & Cowls, J. (2022). A unified framework of five principles for AI in society. *Machine learning and the city: Applications in architecture and urban design*, 535-545.
28. Garrido-Merchan, E. C., Gozalo-Brizuela, R., & Gonzalez-Carvajal, S. (2023). Comparing BERT against traditional machine learning models in text classification. *Journal of Computational and Cognitive Engineering*, *2*(4), 352-356.
29. George, S.; Santra, A.K. Traffic Prediction Using Multifaceted Techniques: A Survey. *Wirel. Pers. Commun.* **2020**, *115*, 1047–1106.
30. Green, S., Hurst, L., Nangle, B., Cunningham, P., Somers, P, and Evans, R. Software agents: A review. Department of Computer Science, Trinity College Dublin, Tech. Rep. TCSCS-1997-06, 1997.
31. Gruske, C. (2023) Alberta courts caution against using unverified citations generated by AI or large language models.
32. Habermas, J. (1991) Thes tructural transformation of the public sphere: An inquiry into a category of bourgeois society, MIT press, 1991.
33. Harris, M. and Wilson, A. Representative bodies in the AI era: Insights for legislatures.
34. He Ke, Y., Yang, R., Lie, S. A., Xin Yi, T., Abdullah, H. R., Wei, D. S., and Liu, N. (2024) Enhancing diagnostic accuracy through multi-agent conversations: Using large language models to mitigate cognitive bias. arXiv preprint arXiv:2401.14589.
35. Helberger, N. and Diakopoulos, N. (2023). Chatgpt and the AI Act. Internet Policy Review, 12(1).
36. Hevner, A., Chatterjee, S. (2010) Design research in information systems: theory and practice, volume 22, Springer Science & Business Media.
37. Honovich, O., Aharoni, R., Herzig, J., Taitelbaum, H., Kukliansy, D., Cohen, V., ... & Matias, Y. (2022). TRUE: Re-evaluating factual consistency evaluation. arXiv preprint arXiv:2204.04991.

38. Houlsby, N., Giurgiu, A., Jastrzebski, S., Morrone, B., De Laroussilhe, Q., Gesmundo, A., ... & Gelly, S. (2019, May). Parameter-efficient transfer learning for NLP. In International Conference on Machine Learning (pp. 2790-2799). PMLR.
39. Hsieh, E. (2006) Conflicts in how interpreters manage their roles in provider–patient interactions, Social Science & Medicine 62 721–730.
40. Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., ... & Chen, W. (2021). Lora: Low-rank adaptation of large language models. arXiv preprint arXiv:2106.09685.
41. Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature machine intelligence*, *1*(9), 389-399.
42. Jiang, A. Q., Sablayrolles, A., Roux, A., Mensch, A., Savary, B., Bamford, C., ... & Sayed, W. E. (2024). Mixtral of Experts. arXiv preprint arXiv:2401.04088.
43. Jiang, A. Q., Sablayrolles, A., Mensch, A., Bamford, C., Chaplot, D. S., Casas, D. D. L., ... & Sayed, W. E. (2023). Mistral 7B. arXiv preprint arXiv:2310.06825.
44. Jovanović, M., & Campbell, M. (2023). Connecting AI: Merging Large Language Models and Knowledge Graph. Computer, 56(11), 103-108.
45. Kadrić, M., Rennert, S., Schäffner, C. (2021) Diplomatic and political interpreting explained, Taylor & Francis Group.
46. Kalampokis, E., Karacapilidis, N., Karamanou, A., Tarabanis, K. (2024) Fostering Multilingual Deliberation through Generative Artificial Intelligence , IFIP EGOV-CeDEM-ePart2024 (EGOV2024), CEUR, Vol.3737.
47. Karamanou A, Brimos P, Kalampokis E, Tarabanis K. Exploring the Quality of Dynamic Open Government Data Using Statistical and Machine Learning Methods. *Sensors*. 2022; 22(24):9684. https://doi.org/10.3390/s22249684.
48. Karan Singhal, Shekoofeh Azizi, Tao Tu, S Sara Mahdavi, Jason Wei, Hyung Won Chung, Nathan Scales, Ajay Tanwani, Heather Cole-Lewis, Stephen Pfohl, et al. 2022. Large language models encode clinical knowledge. arXiv preprint arXiv:2212.13138 (2022).
49. Ke, Y. H., Yang, R., Lie, S. A., Lim, T. X. Y., Abdullah, H., R., Ting, D. S. W., and Liu, N.. Enhancing diagnostic accuracy through multi-agent conversations: Using large language models to mitigate cognitive bias. arXiv preprint arXiv:2401.14589, 2024.
50. Koehn, P., Knowles, R. (2017) Six challenges for neural machine translation, in: First Workshop on Neural Machine Translation, Association for Computational Linguistics, pp. 28–39.
51. Kung, T. H., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., ... & Tseng, V. (2023). Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLoS digital health*, *2*(2), e0000198.
52. Lai, J., Gan, W., Wu, J., Qi, Z., and Yu, P. S. (2023) Large language models in law: A survey. arXiv preprint arXiv:2312.03718.
53. Laios, A., Theophilou, G., Jong, D. De., Kalampokis, E. (2023a) The Future of AI in Ovarian Cancer Research: The Large Language Models Perspective. *Cancer Control*;30.
54. Le Scao, T., Fan, A., Akiki, C., Pavlick, E., Ilić, S., Hesslow, D., Castagné, R., Luccioni, A. S., Yvon, F., Gallé, M. et al. (2022) Bloom: A 176b-parameter open-access multilingual language model.
55. Lester, B., Al-Rfou, R., & Constant, N. (2021). The power of scale for parameter-efficient prompt tuning. arXiv preprint arXiv:2104.08691.
56. Li, H., Su, Y., Cai, D., Wang, Y., & Liu, L. (2022). A survey on retrieval-augmented text generation. *arXiv preprint arXiv:2202.01110*.

57. Liu, Y. (2019). Fine-tune BERT for extractive summarization. *arXiv preprint arXiv:1903.10318*.
58. Liu, H., Tam, D., Muqeeth, M., Mohta, J., Huang, T., Bansal, M., & Raffel, C. A. (2022). Few-shot parameter-efficient fine-tuning is better and cheaper than in-context learning. Advances in Neural Information Processing Systems, 35, 1950-1965.
59. Listorti, G., Basyte-Ferrari, E., Acs, S., and Smits, P. (2020). Towards an evidence-based and integrated policy cycle in the EU: A review of the debate on the better regulation agenda. JCMS: Journal of Common Market Studies, 58(6):1558–1577.
60. Lou, R., Zhang, K., & Yin, W. (2023). Is prompt all you need? no. A comprehensive and broader view of instruction learning. *arXiv preprint arXiv:2303.10475*.
61. Longo. E., The european citizens' initiative: too much democracy for eu polity?, German Law Journal 20 (2019) 181–200. doi:10.1017/glj.2019.12.
62. von Lucke, J., Fitsilis,F., and Etscheid, J. (2023) Research and development agenda for the use of ai in parliaments. In Proceedings of the 24th Annual International Conference on Digital Government Research, pages 423–433.
63. Mamalis, M.E., Kalampokis, E., Fitsilis, F., Theodorakopoulos, G., Tarabanis, K. (2024). A Large Language Model Agent Based Legal Assistant for Governance Applications. In: Janssen, M., et al. Electronic Government. EGOV 2024. Lecture Notes in Computer Science, vol 14841. Springer, Cham. https://doi.org/10.1007/978-3-031-70274-7_18
64. Mamalis, M., Kalampokis, E. , A. Karamanou, P. Brimos, K. Tarabanis (2024) Can Large Language Models Revolutionalize Open Government Data Portals? A Case of Using ChatGPT in statistics.gov.scot *27th Panhellenic Conference on Progress in Computing and Informatics (PCI 2023), ACM*, pp.53-59.
65. McKinsey Global Institute (2022). McKinsey Data & AI Summit.
66. Mehta, R., & Varma, V. (2023). LLM-RM at SemEval-2023 Task 2: Multilingual Complex NER using XLM-RoBERTa. *arXiv preprint arXiv:2305.03300*.
67. Momotko M., Izdebski W., Tambouris E., Tarabanis K. and Vintar M. An Architecture of Active Life Event Portals: Generic Workflow Approach. Electronic Government, LNCS # 4656, Springer Verlag, (2007), 104-115.
68. Morales-Gálvez, S. (2017) Living together as equals: Linguistic justice and sharing the public sphere in multilingual settings, Ethnicities 17 646–666.
69. National Archives Office of the Federal Register and Records Administration (2023) 88 fr 75191 - safe, secure, and trustworthy development and use of artificial intelligence. [government]. Federal Register.
70. Nelson, W., Lee, M. K., Choi, E., & Wang, V. (2024). Designing LLM-Based Support for Homelessness Caseworkers. In *AAAI-2024 Workshop on Public Sector LLMs: Algorithmic and Sociotechnical Design*.
71. Nikiforova, A. (2021). Smarter Open Government Data for Society 5.0: are your open data smart enough? Sensors 21, 15 (2021), 5204.
72. Nitta, Y. (1986) Problems of machine translation system- effect of cultural differences on sentence structure, Future Gener. Comput. Syst. 2 101–115. doi:10.1016/0167-739X(86)90004-X.
73. Van Noordt, C., and Misuraca, G (2022). Exploratory insights on Artificial Intelligence for government in Europe. Social Science Computer Review, 40(2):426–444.

74. OECD (2020), "The OECD Digital Government Policy Framework: Six dimensions of a Digital Government", *OECD Public Governance Policy Papers*, No. 2, OECD Publishing, Paris, https://doi.org/10.1787/f64fed2a-en .
75. Oltramari, A., Francis, J., Henson, C., Ma, K., & Wickramarachchi, R. (2020). Neuro-symbolic architectures for context understanding. arXiv preprint arXiv:2003.04707.
76. van Ooijen, C., B. Ubaldi and B. Welby (2019), "A data-driven public sector: Enabling the strategic use of data for productive, inclusive and trustworthy governance", *OECD Working Papers on Public Governance*, No. 33, OECD Publishing, Paris, https://doi.org/10.1787/09ab162c-en.
77. OpenAI, R. (2023). GPT-4 technical report. arXiv, 2303-08774.
78. Ostendorff, M., & Rehm, G. (2023). Efficient language model training through cross-lingual and progressive transfer learning. arXiv preprint arXiv:2301.09626.
79. Pan, S., Luo, L., Wang, Y., Chen, C., Wang, J., & Wu, X. (2023). Unifying Large Language Models and Knowledge Graphs: A Roadmap. arXiv preprint arXiv:2306.08302.
80. van Parijs, P. (2004) Europe's linguistic challenge, European Journal of Sociology 45 113–154. doi:10.1017/S0003975604001407.
81. Patten, A. (2007) Theoretical foundations of european language debates, in: D. Castiglione, C. Longman (Eds.), The Language Question in Europe and Diverse Societies: Political, Legal and Social Perspectives, Hart Publishing, London.
82. Peffers, K., Tuunanen, T. , Rothenberger, M. A., Chatterjee, S. (2007) A design science research methodology for information systems research, Journal of management information systems 24, 45–77
83. Peristeras, V. The Governance Enterprise Architecture-GEA for Reengineering Public Administration. PhD Thesis, University of Macedonia, Greece, 2006.
84. Peña, A. et al. (2023). Leveraging Large Language Models for Topic Classification in the Domain of Public Affairs. In: Coustaty, M., Fornés, A. (eds) Document Analysis and Recognition – ICDAR 2023 Workshops. ICDAR 2023. Lecture Notes in Computer Science, vol 14193. Springer, Cham. https://doi.org/10.1007/978-3-031-41498-5_2
85. Petroni, F., Lewis, P., Piktus, A., Rocktäschel, T., Wu, Y., Miller, A. H., & Riedel, S. (2020). How context affects language models' factual predictions. *arXiv preprint arXiv:2005.04611*.
86. Piller, I. (2016) Linguistic diversity and social justice: An introduction to applied sociolinguistics, Oxford University Press.
87. Popel, M., Tomkova, M., Tomek, J., Kaiser, Ł., Uszkoreit, J., Bojar, O., and Žabokrtský, Z. (2020) Transforming machine translation: a deep learning system reaches news translation quality comparable to human professionals, Nature Communications 11 4381.
88. Radford, A., & Narasimhan, K. (2018). Improving Language Understanding by Generative Pre-Training. https://blog.openai.com/ language-unsupervised. Accessed 7 March 2024.
89. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. OpenAI blog, 1(8), 9.
90. Rajpurkar, P., Zhang, J., Lopyrev, K., & Liang, P. (2016). Squad: 100,000+ questions for machine comprehension of text. arXiv preprint arXiv:1606.05250.
91. Rapoport, R. N. (1970) Three dilemmas in action research: with special reference to the tavistock experience, Human relations 23, 499–513.
92. Rehm, G. and Way, A. (2023). European Language Equality: A Strategic Agenda for Digital Language Equality. Cognitive Technologies. Springer.

93. Ringe, N. (2022) The Language(s) of Politics: Multilingual Policy-Making in the European Union, University of Michigan Press.
94. Sadiq, S., Aryani, A., Demartini, G. *et al.* Information Resilience: the nexus of responsible and agile approaches to information use. *The VLDB Journal* **31**, 1059–1084 (2022). https://doi.org/10.1007/s00778-021-00720-2
95. Salah, M., Abdelfattah, F. & Al Halbusi, H. (2023). Generative Artificial Intelligence (ChatGPT & Bard) in Public Administration Research: A Double-Edged Sword for Street-Level Bureaucracy Studies, International Journal of Public Administration, DOI: 10.1080/01900692.2023.2274801
96. Saleforce (2024), The State of Data and Analytics Report. https://salesforce.com/resources/research-reports/state-of-data-analytics/?d=cta-body-promo-8
97. Shi, F., Chen, X., Misra, K., Scales, N., Dohan, D., Chi, E. H., ... & Zhou, D. (2023, July). Large language models can be easily distracted by irrelevant context. In *International Conference on Machine Learning* (pp. 31210-31227). PMLR.
98. Siciliani, L., Ghizzota, E., Basile, P., & Lops, P. (2023). OIE4PA: open information extraction for the public administration. *Journal of Intelligent Information Systems*, 1-22.
99. Singhal, K., Azizi, S., Tu, T., Mahdavi, S. S., Wei, J., Chung, H. W., ... & Natarajan, V. (2023). Large language models encode clinical knowledge. *Nature*, *620*(7972), 172-180.
100. Solove, D. J. and Schwartz, P. M. (2023) EU Data Protection and the GDPR. Aspen Publishing.
101. Stefaniak, K. (2020). Evaluating the usefulness of neural machine translation for the Polish translators in the European Commission. In André Martins, Helena Moniz, Sara Fumega, Bruno Martins, Fernando Batista, Luisa Coheur, Carla Parra, Isabel Trancoso, Marco Turchi, Arianna Bisazza, Joss Moorkens, Ana Guerberof, Mary Nurminen, Lena Marg, & Mikel L. Forcada (Eds.),Proceedings of the 22nd annual conference of the European Association for Machine Translation (pp. 263–269). European Association for Machine Translation.
102. Stiennon, N., Ouyang, L., Wu, J., Ziegler, D., Lowe, R., Voss, C., ... & Christiano, P. F. (2020). Learning to summarize with human feedback. Advances in Neural Information Processing Systems, 33, 3008-3021.
103. Sun, L., et al., 2024. Trustllm: Trustworthiness in large language models. arXiv preprint arXiv:2401.05561.
104. Sun, X., Li, X., Li, J., Wu, F., Guo, S., Zhang, T., & Wang, G. (2023). Text Classification via Large Language Models. *arXiv preprint arXiv:2305.08377*.
105. Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. Large language models are zero-shot reasoners. Advances in neural information processing systems, 35:22199–22213, 2022.
106. Tay, Y., Dehghani, M., Bahri, D., and Metzler, D. (2022). Efficient transformers: A survey. arXiv preprint cs.LG/2009.06732.
107. Taylor Webb, Keith J Holyoak, and Hongjing Lu. Emergent analogical reasoning in large language models. Nature Human Behaviour, 7(9):1526–1541, 2023.
108. Thoppilan, R., De Freitas, D., Hall, J., Shazeer, N., Kulshreshtha, A., Cheng, H. T., ... & Le, Q. (2022). Lamda: Language models for dialog applications. arXiv preprint arXiv:2201.08239.

109. Thorne, J., Vlachos, A., Christodoulopoulos, C., & Mittal, A. (2018). FEVER: a large-scale dataset for fact extraction and VERification. arXiv preprint arXiv:1803.05355.
110. Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M. A., Lacroix, T., ... & Lample, G. (2023a). Llama: Open and efficient foundation language models. arXiv preprint arXiv:2302.13971.
111. Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., ... & Scialom, T. (2023b). Llama 2: Open foundation and fine-tuned chat models. arXiv preprint arXiv:2307.09288.
112. Trust, P., Omala, K., Minghim, R. et al. (2024) Augmenting Large Language Models for Enhanced Interaction with Government Data Repositories, PREPRINT (Version 1) available at Research Square, https://doi.org/10.21203/rs.3.rs-3897706/v1
113. Uchiyama, A. (2018) The politics of translation in meiji japan, in: The Routledge Handbook of Translation and Politics, Routledge, pp. 455–466.
114. Unesco. (2022). *Recommendation on the ethics of artificial intelligence*. United Nations Educational, Scientific and Cultural Organization.
115. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30.
116. Volkmer, I. (2014) The global public sphere: Public communication in the age of reflective interdependence, John Wiley & Sons.
117. Wang, Q., Li, B., Xiao, T., Zhu, J., Li, C., Wong, D. F., Chao, L. S. (2019) Learning deep transformer models for machine translation, in: Annual Meeting of the Association for Computational Linguistics.
118. Wang, S., Sun, X., Li, X., Ouyang, R., Wu, F., Zhang, T., ... & Wang, G. (2023). Gpt-ner: Named entity recognition via large language models. *arXiv preprint arXiv:2304.10428*.
119. Wei, J., Tay, Y., Bommasani, T., Raffel, C., Zoph, B., Borgeaud, S., Yogatama, D., Bosma, M., Zhou, D., Metzler, D., et al. (2022) Emergent abilities of large language models. arXiv preprint arXiv:2206.07682.
120. Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., ... & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. Advances in Neural Information Processing Systems, 35, 24824-24837.
121. Wojciechowska, M. (2019) Towards intersectional democratic innovations, Political Studies 67 895–911.
122. Wooldridge M. and Jennings, N. R., Intelligent agents: Theory and practice. The knowledge engineering review, 10(2):115–152, 1995.
123. World Bank. World development report 2021: Data for better lives. The World Bank. 2021.
124. Wu, S., Irsoy, O., Lu, S., Dabravolski, V., Dredze, M., Gehrmann, S., ... & Mann, G. (2023). Bloomberggpt: A large language model for finance. *arXiv preprint arXiv:2303.17564*.
125. Xi, Z., Chen, W., Guo, X., He, W., Ding, Y., Hong, B., Zhang, M., Wang, J., Jin, S., Zhou, E., et al. (2023) The rise and potential of large language model based agents: A survey. arXiv preprint arXiv:2309.07864.
126. Xu, J., Wang, J., Leung, J., & Gu, J. (2024, December). GRASP: Municipal Budget AI Chatbots for Enhancing Civic Engagement. In 2024 IEEE International Conference on Big Data (BigData) (pp. 7438-7442). IEEE.

127. Yang, Y. (2010) Working with assumptions –the dialogue interpreter as communication facilitator in medical encounters, Comparative Literature: East & West 12 (2010) 162–171.
128. Yao, S., Yu, D., Zhao, J., Shafran, I., Griffiths, T. L., Cao, Y., & Narasimhan, K. (2023). Tree of thoughts: Deliberate problem solving with large language models. arXiv preprint arXiv:2305.10601.
129. Ye, J., Chen, X., Xu, N., Zu, C., Shao, Z., Liu, S., ... & Huang, X. (2023). A comprehensive capability analysis of gpt-3 and gpt-3.5 series models. arXiv preprint arXiv:2303.10420.
130. Ye, S., Hwang, H., Yang, S., Yun, H., Kim, Y., & Seo, M. (2023). In-context instruction learning. *arXiv preprint arXiv:2302.14691*.
131. Yenduri, G., Srivastava, G., Maddikunta, P. K. R., Jhaveri, R. H., Wang, W., Vasilakos, A. V., & Gadekallu, T. R. (2023). Generative Pre-trained Transformer: A Comprehensive Review on Enabling Technologies, Potential Applications, Emerging Challenges, and Future Directions. *arXiv preprint arXiv:2305.10435*.
132. Yin, R. K. (2009) Case study research: Design and methods, volume 5. sage.
133. Yong, Z. X., Schoelkopf, H., Muennighoff, N., Aji, A. F., Adelani, D. I.,, ... & Nikoulina, V. (2022). Bloom+ 1: Adding language support to bloom for zero-shot prompting. arXiv preprint arXiv:2212.09535.
134. Yun, L. et al. (2024). *Improving citizen–government interactions with generative AI*. PLOS ONE. https://doi.org/10.1371/journal.pone.0311410
135. Zhao, A., Huang, D., Xu, Q., Lin, M., Liu, Y. and Huang, G. (2023). ExpeL: LLM Agents Are Experiential Learners. arXiv:2308.10144 [cs.LG]
136. Zhang, Z., Wu, S., Jiang, D., & Chen, G. (2021). BERT-JAM: Maximizing the utilization of BERT for neural machine translation. *Neurocomputing*, *460*, 84-94.
137. Zhiheng Xi, Wenxiang Chen, Xin Guo, Wei He, Yiwen Ding, Boyang Hong, Ming Zhang, Junzhe Wang, Senjie Jin, Enyu Zhou, et al. The rise and potential of large language model based agents: A survey. arXiv preprint arXiv:2309.07864, 2023.
138. Zhou, Y., Muresanu, A. I., Han, Z., Paster, K., Pitis, S., Chan, H., and Ba, J. (2022) Large language models are human-level prompt engineers. arXiv preprint arXiv:2211.01910.
139. Zhou, Chunting, et al. (2023). Lima: Less is more for alignment. arXiv preprint arXiv:2305.11206.
140. Zhou, Chunting, et al. (2020) Detecting hallucinated content in conditional neural sequence generation. arXiv preprint arXiv:2011.02593.