KG.GOV: Knowledge Graphs as the Backbone of Data Governance in AI

Albert Meroño-Peñuela[a], Elena Simperl[a], Anelia Kurteva[a], Ioannis Reklos[a]

[a]*Department of Informatics, King's College London, United Kingdom*

Abstract

As (generative) Artificial Intelligence continues to evolve, so do the challenges associated with governing the data that powers it. Ensuring data quality, privacy, security, and ethical use become more and more challenging due to the increasing volume and variety of the data, the complexity of AI models, and the rapid pace of technological advancement. Knowledge graphs have the potential to play a significant role in enabling data governance in AI, as we move beyond their traditional use as data organisational systems. To address this, we present KG.GOV, a framework that positions KGs at a higher abstraction level within AI workflows, and enables them as a backbone of AI data governance. We illustrate the three dimensions of KG.GOV: modelling data, alternative representations, and describing behaviour; and describe the insights and challenges of three use cases implementing them: Croissant, a vocabulary to model and document ML datasets; WikiPrompts, a collaborative KG of prompts and prompt workflows to study their behaviour at scale; and Multimodal transformations, an approach for multimodal KGs harmonisation and completion aiming at broadening access to knowledge.

Keywords: Knowledge graphs, AI, Governance

1. Introduction

Recent breakthroughs in generative AI (GenAI) offer immense potential for societal change, with the AI market expected to grow significantly from $540 billion in 2023 to $1.27 trillion over the next five years [60]. However, alongside this potential, the concerns about safety and reliability persist [10]. GenAI and machine learning (ML) heavily rely on vast amounts of high-quality and diverse data to learn patterns, make predictions, and generate new content. One of the main challenges the field faces is the lack of transparency regrading data: how it is collected, how it is used for training, and how it is subsequently repurposed within AI systems [29]. Establishing a strong data governance model (i.e. a set of agreed upon processes that encourage and ensure that data is accurately, securely, and ethically managed across enterprises and software systems[1]) is crucial for cultivating more trustworthy AI [36]. This involves making data, ML algorithms and GenAI models available and accessible (e.g. through clear access controls) for review (e.g. audits), and providing documentation of model inputs and outputs [36]. All data assets (e.g. datasets, documents and images with different formats) and their lineage and provenance in such settings need to be recorded and catalogued in a format that eases their interoperability across machines and humans.

Semantic technologies such as ontologies and Knowledge Graphs (KGs) have been known for many years as effective tools for data management (e.g. representation, organisation, sharing, reuse) and for supporting interoperability by design [21, 19, 23, 32]. Vocabularies such as DCAT[2] capture metadata about other datasets and facilitate their management and stewardship. Recently, KGs have also started to emerge as a more explicit data governance solution (e.g. [7, 47]) due to their ability and flexibility in modelling diverse intricate data dependencies and capture provenance information essential for improving AI system transparency and explainability [63]. Research has already shown that KGs can not only help model valuable data, its dynamic relationships, and the (legal) knowledge to which it abides (e.g. [9, 40]); but they can also enable AI to operationalise the data by making it actionable through semantics (e.g.[6, 71]). This positions KGs as tools that can go beyond data modelling, and potentially operationalising and automating data and AI governance in dynamic scenarios (e.g. data lifecycle in AI systems).

Considering this and years of research showcasing the usefulness of semantic technology for data and knowledge management in various domains and software applications [32], we believe that knowledge graphs could become powerful tools for the governance of data in AI, and especially towards the automation of such governance. In this way, KGs can be used for more than just modelling domains of knowledge: they can be used to also capture dynamic, meaningful metadata such as its lineage, security, accuracy, ethical management, etc; which are fundamental to enable, automate, and scale their governance. In this pa-

[1] Adapted from `https://csrc.nist.gov/glossary/term/data_governance`

[2] `https://www.w3.org/TR/vocab-dcat-3/`

per, we propose the KG.GOV framework (section 2) for data governance in AI, which focuses on three dimensions: *data modelling* and description; the various *multimodal representations* that data can take; and *agent behaviour*, both considering humans and machines. We provide a description of each dimension in section 2, and examples of our most recent work for each of the dimensions from the Croissant [3] vocabulary and the WikiPrompts and MuseIT projects in section 3. As a conclusion, in section 4, we reflect on commonalities of these applications and forecast trends for the future of KGs as tools that help operationalise and automate data governance in AI.

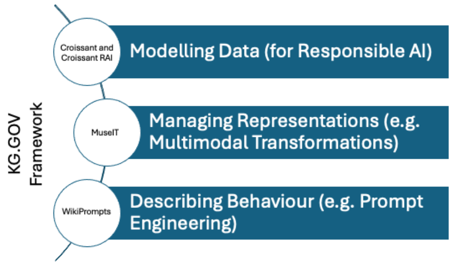


Figure 1: The KG.GOV Framework and its Dimensions

2. The KG.GOV Framework

The KG.GOV framework (Figure 1) aims to highlight the need for better understanding and modelling of the different aspects (or knowledge dimensions) of data to better support its governance in AI. In particular, dimensions we should consider:

1. **Modelling Data** in a format that is easily interoperable across machines and humans is essential for effective data governance at scale. Knowledge graphs and their semantics are known to support data's interoperability by enabling one to represent dynamic and complex contexts with ease. To do so effectively, however, one needs to first have a solid knowledge of (i) types of data needed improve AI's decision making at each stage through its lifecycle (e.g. training, refinement, deployment, maintenance), (ii) what data needs to be recorded for purposes such as legal audits of the system, (iii) the data access rights and special licences protecting ones intellectual property (IP). Section 3.1 presents our work on this in the scope of responsible AI (RAI).
2. **Managing Representations.** Beyond managing metadata relevant for AI systems, a fundamental part of governing AI consists on identifying what actual AI data objects, pieces of information, and knowledge need to be managed, and their possible representations. In traditional knowledge bases representations are based on the notion of *symbols*, but with the advent of machine learning these have been extended to *vectors* (i.e. for geometric computation and reasoning), but also text (e.g. with LLMs) and images (for text-to-image generative AI systems) as valid input and output of AI systems. GenAI is becoming increasingly adept at handling, understanding and reasoning over multimodal data such as images, sound, music, 3D geometry, and so on. Knowledge in modern AI has, now more than ever, polyglot and heterogeneous representations [35, 52, 58]. In section 3.2 we present insights to this focusing on the problem of knowledge gaps, bias, misrepresentation and multimodality.
3. **Describing Behaviour.** The rapid advancement of AI poses questions about the governance of behaviour in a world of "agentic AI", i.e. AI systems that "can adaptably achieve complex goals in complex environments with limited direct supervision" [59]. Reminiscent of other complex AI task-solving workflows like Semantic Web Services [20], this would put AI agents in a varied landscape, in which they interact with humans, data, services and applications in workflows that we do not yet fully understand. In this scenario, rather than just human-AI conversations and prompting we need to think about auto-generated prompts, LLMs talking to other LLMs, prompt engineering, etc; and about documenting the inputs and outputs of these AI processes and workflows. This can also be understood as simultaneous descriptions of both *retrospective provenance* of these workflows (capturing past workflow workflow execution and results) and their *prospective provenance* (configurations, inputs, outputs and pipelines for and with AI) [43]. Section 3.3 presents insights from our work on describing behaviour by modelling prompt workflows and interactions based on knowledge graphs.

3. Applications

In the following, we describe the insights and challenges of implementing the KG.GOV Framework in three different use cases, one for each of its dimensions. Some of these use cases aim at covering the dimensions completely (Croissant), whereas some others are more examples of partial solutions (Multimodal Transformations, Wikiprompts).

3.1. Modelling Data: Croissant

The rapid AI advancements AI and prominent regulatory frameworks (e.g. the AI Act [16] and previously the GDPR [17]), have highlighted the need for more accountability and responsibility in data's governance. This has also shifted the focus towards the need for a more data-centric approach to responsible AI [29]. The provision of structured summaries of various aspects of data's lifecycle in ML, for example, with the help of data cards[3] [53] and data nutrition labels [61] has become well-known approach for promoting transparency around data in AI.

A more recent effort focused on developing data-driven mechanisms for more accurate, safer, faster, and efficient AI has been put forward by the ML Commons[4] community, which brings together experts from industry and academia to openly and collaboratively develop standards for semantically modelling and documenting datasets' lineage and provenance in AI. However, determining what types of data to model, the required level of detail that needs to be captured, and how to effectively promote its responsible use and reuse within AI systems is a significant challenge with limited guidance provided by laws like the AI Act. Progress is being made in this direction by the Croissant working group[5] (part of ML Commons), which has been collaborating on a harmonised metadata format called **Croissant** (see Fig. 2) that builds on schema.org's Dataset[6] vocabulary and aims to support and ease ML dataset's discoverability, portability, reproducibility, and interoperability by enabling machine-readable documentation [3] at the following four layers:

- **Dataset Metadata Layer** describes general metadata about a dataset such as its name, description and applicable license.
- **Resources Layer** describes metadata about a dataset's content such as comprising files and the different formats are in.
- **Structure Layer** supports the description and organisation of different resources in a dataset.
- **Semantic Layer** supports the preservation of dataset's contextual integrity in various domain specific applications (e.g. geospatial analysis)

The research on Croissant's integration with existing data repositories such as HuggingFace[7], Kaggle[8], OpenML[9] has also shown promising results (over 400,000 datasets in the Croissant format can be now retrieved across these repositories) [3]. An interface (the Croissant Editor[10]) aimed at aiding users wanting to validate, annotate and create Croissant datasets has also been developed and is available openly for use. A full specification of Croissant is available publicly online[11].

The need for more descriptive metadata documentation of datasets to better support responsible AI (RAI)

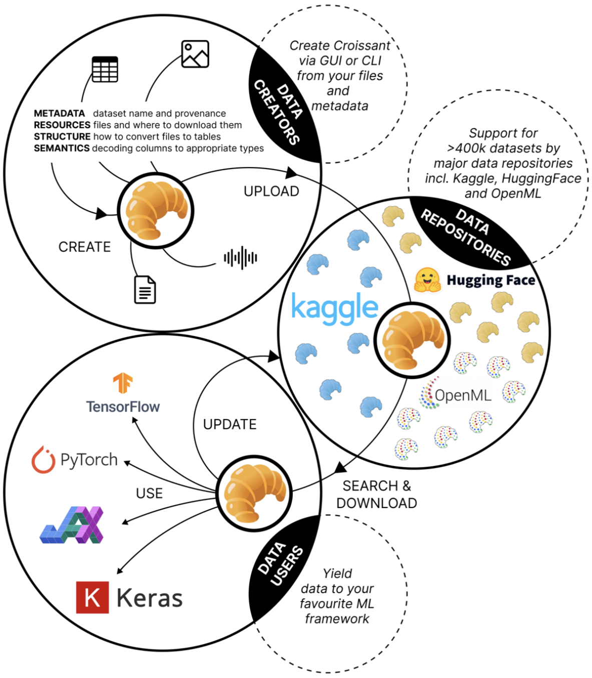


Figure 2: Croissant Lifecycle and Ecosystem (From [3]). Data creators can use Croissant to semantically enrich their datasets making them more interoperable and discoverable on data repositories. Data users can then search for ML-ready (Croissant) datasets and use them on various ML frameworks.

principles (e.g. transparency, accountability, privacy and security and reliability as defined in [46, 22]) and contribute to regulatory compliance verifications and evaluations of AI systems with regards to datasets' use, has also motivated the work on Croissant RAI[12]. Specific use cases that are currently investigated based on their relevancy and impact on real-world AI adoption are the semantic modelling of data's lifecycle in AI, processes such as data labelling and participatory data provision, regulatory compliance verification and AI safety and fairness evaluation. The work has already begun for the geospatial and healthcare (viewed as possibly high-risk AI application) domains, where the reuse, integration and alignment with other widely used and relevant vocabularies such as Data Use Vocabulary (DUO)[13], Data Privacy Vocabulary (DPV)[14], Data Catalogue Vocabulary (DCAT)[15], AIRO [28] for representing AI risks and languages such as the Open Digital Rights Language (ODRL)[12] is currently being investigated.

With the work on Croissant, the AI and knowledge engineering communities have come to a common understanding that modelling and documenting information on data's provenance plays a key role in enabling more transparency and explainability in AI [29]. The documentation

[3] ttps://sites.research.google/datacardsplaybook/
[4] https://mlcommons.org
[5] https://mlcommons.org/working-groups/data/croissant/
[6] https://schema.org/Dataset
[7] https://huggingface.co
[8] https://www.kaggle.com
[9] https://www.openml.org
[10] https://huggingface.co/spaces/MLCommons/croissant-editor
[11] https://docs.mlcommons.org/croissant/docs/croissant-spec.html
[12] https://docs.mlcommons.org/croissant/docs/croissant-rai-spec.html
[13] https://github.com/EBISPOT/DUO
[14] https://w3c.github.io/dpv/2.0/dpv/
[15] https://www.w3.org/TR/vocab-dcat-3/

of data through its lifecycle in AI (e.g. from collection through web scraping, through processing to form specialised datasets to use for training and testing) not only support its interoperability and wider reuse between communities but can also help establish accountability and responsibility with regards to the identities of different agents (software, people, organisations) handling the data and their roles [33, 51]. Further, the availability and accessibility to such information, especially in a machine-interoperable format, can significantly help in implementing robust data governance and verifying legal compliance with regards to how data is used and how AI systems are built. Prior to the AI Act and the rise of generative AI, a significant number of work has already been done on successfully utilising semantically modelled data to support semi- and fully-automated regulatory compliance verification with regards to (personal) data's governance [9, 39, 18, 62].

By considering existing work and building on it through an AI perspective, the ongoing work on Croissant RAI paves the way towards a standard for machine-readable documentation of datasets that ease and support their responsible use for AI. However, the multi-faced nature of RAI (covering technical, legal and social aspects of AI) presents us with different challenges in terms of what specific data (and provenance about it) needs to be modelled, at what level of detail and how exactly. Interoperability is an inherent limitation of the (social aspects of) KGs here, as there are already various existing vocabularies for data use and regulatory compliance in different domains competing in a similar space. An inherent limitation of KGs here may be one of implementation, as dealing with multimodal representations is generally less well supported than symbols/text in KG triplestores and query engines–new standards may help in addressing this.

3.2. Alternative Representations: Multimodal transformations

Developments such as the Transformer architecture [64] and the availability of large, high-quality datasets [24] have given rise to powerful Generative AI models that can generate human-like texts that resemble their training data while still being somewhat novel. In addition to text-only models there have been significant advances to state-of-the-art Multimodal Generative AI models enabling them to process and generate content across multiple data modalities, such as text, image, video, and audio. Leading models like OpenAI's GPT-4 Vision [70], which can accept text and image inputs and Google DeepMind's Gemini [4], which can accept text, images and video, are capable of sophisticated cross-modal interactions, where a model might generate a realistic image from a text description or produce audio based on visual inputs. Similarly, Meta's ImageBind [26] model is unique in its ability to create representations that bind information from six modalities—text, image, video, audio, depth, and IMU sensor data (motion data)—which enables it to operate in real-world multi-sensor environments.

These generative models leverage transformer architectures or large-scale diffusion models. Transformers [64], initially designed for natural language processing, have proven adaptable and highly effective for multimodal tasks due to their attention mechanisms that capture intricate relationships between different data types. Diffusion models [56], like those used in DALL-E, Midjourney and Stable Diffusion [15], gradually transform noise into coherent data (such as images). They are known for their high fidelity in visual generation tasks and are specialized in generating highly coherent images from text prompts with an advanced understanding of detailed prompts and context. Together, these models mark a significant leap in AI's ability to understand and generate across diverse modalities, opening up new possibilities in applications related to Multimodal Knowledge Graphs. In particular, Multimodal generative AI models can utilize Knowledge Graphs to improve their outputs in downstream tasks, which include understanding and reasoning, classification, content generation and retrieval tasks. At the time of writing, Hugging Face model repository[16] contains more than 1M models, with 150K+ models for text generation, 43K+ text-to-image models, 2K+ text-to-speech models, 1.5K text-to-audio (including music) models, and even a handful of text-to-video, image-to-video, image and text-to-3d models, etc. Furthermore, Multimodal generative AI models can be used for tasks such as acquisition, fusion and inference on Multimodal Knowledge Graphs [8].

AI's capability to transform data across different modalities is easy to dismiss as a "gimmick" or a nice-to-have (rather than a must-have) requirement, but increasingly plays a more important role when considering the impact of AI in society and the ways (often disadvantaged) people access knowledge. Key data and knowledge repositories that feed our everyday Web searches have well-known content gaps [1, 55] that can magnify gender and socioeconomic biases and favour discrimination. In KGs such as Wikidata [66] these gaps do not only exist in terms of properties and value literals but also in what we could call a multimedia gap, with central items missing e.g. images [2]. Missing alternative representations for public data is problematic for e.g. people with disabilities: it is estimated that one billion people (15% of the world's population)[17] experience some form of disability, with one-fifth of the estimated global total (110-190 million people) experiencing significant disabilities. Facilitating equal access to knowledge, via e.g. modality alternatives (images for those who cannot hear, sound for those who cannot see, etc.) is a right granted by the UN's Universal Declaration of Human Rights[18]. However, the provision of dynamic

[16]https://huggingface.co/models
[17]https://www.worldbank.org/en/topic/disability
[18]https://www.un.org/en/about-us/universal-declaration-of-human-rights

combinations of such knowledge modalities requires large amounts of labour that are hard to obtain in volunteer-driven projects.

The Horizon Europe project *Multisensory, User-centred, Shared cultural Experiences through Interactive Technologies* (MuseIT[19]) proposes technologies that facilitate and widen access to cultural assets. With **Multimodal Transformations**, we study ways to leverage modern generative AI to automatically provide representations of data that can cover different user needs according to their sensory capabilities. To address this, we propose an extension to the CUltural BEnchmark for Text-to-Image models (CUBE) [37]. CUBE contains 300K cultural artifacts across 8 countries (Brazil, France, India, Italy, Japan, Nigeria, Turkey, and USA) and 3 domains (cuisine, landmarks, art) extracted from Wikidata; and 1K text-to-image generation prompts that enable evaluation of cultural awareness of generative AI models. These prompts are automatically generated from the Wikidata KG properties directly, and thus the KG plays the key role of being the central and unique source of authoritative knowledge. We take this work as a basis to propose CUBE-MT, which extends CUBE in various ways:

- We extend the *modalities* supported by the benchmark, originally just images, to include also include 6 additional modalities: text, Braille, speech, music, video, and 3D—modalities that are relevant for the provision of audio, haptics, etc.
- We extend the *prompts* in the benchmark to account for the cultural awareness of generating those modalities
- We *run* the benchmark to generate a dataset with instances of those modalities, using publicly available models in Hugging Face (Stable Diffusion, Phi3, FastSpeech, MusicGen)

All CUBE-MT resources are available online[20], including the extended benchmark, dataset, and links to the original Wikidata items. Figure 3 shows an example Wikidata item and its corresponding multimodal representations produced while running the benchmark. Braille and video representations are also generated, but algorithmically.

Many opportunities and challenges are open for the future. Despite its importance for knowledge equity [55], we are conscious of the difficulty in evaluating generative AI outputs and we share current views that more adequate metrics and evaluation frameworks are needed [2]. This is tangential to policy and AI accountability, as users will inevitably raise concerns about trust in, and the ethics of, AI-generated content.

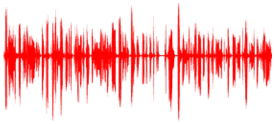
[19] https://www.muse-it.eu/

[20] https://github.com/albertmeronyo/CUBE-MT

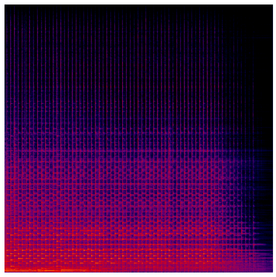
(a) Image for the prompt: *A panoramic view of Kojinyama Fortress in Japan, realistic*

Kojinyama Fortress is an ancient hilltop castle located at the summit of Mount Kojin near the Yamashiro province in present-day Kyoto Prefecture, Japan, offering a rich historical tapestry amid the tranquility of the Japanese countryside.

(b) Text for the prompt: *A one sentence textual description of Kojinyama Fortress from Japanese landscapes*



(c) Waveform representation for the speech reading the text of (b).



(d) Spectrogram representation of the music for the prompt: *A short song representing Kojinyama Fortress from Japanese landscapes*

Figure 3: Multimodal transformations for the Wikidata item Q71053154, *Kojinyama Fortress* and the CUBE-MT prompts used to generate them.

3.3. Describing Behaviour: WikiPrompts

To think about the governance of AI behaviour, we use the notion of *agenticness* in AI systems, i.e. the idea that AI systems that "can adaptably achieve complex goals in complex environments with limited direct supervision" [59]. Deploying multimodal, polyglot generative AI in complex environments suggests a varied landscape, in which various foundational, fine-tuned, few-shot instructed, etc models will need co-exist and cooperate to complete tasks. Such a cooperation necessarily entails not just human-AI interactions (i.e. prompting and conversations [30]), but also autonomous AI-data, AI-AI, AI-services (e.g. function calling [38]) interfacing that will require orchestration and modelling of data and process flows in a similar way to Semantic Web Services [20]. Therefore, key questions here are: what complex workflows does agentic AI generate and how can they be documented? How can their inputs and outputs be mapped to specific processes and models? Various proposals exist related to the latter, including AI cards [27] and AI usage cards [67], model cards for model reporting [49], taxonomies for AI incident reporting [50], as well as safety testing (e.g. AI Safety Benchmark [65]). However, none of these approaches cover the documentation and modelling of agentic *workflows* and *prompts*.

In order to implement these interactions and workflows, generative AI models use *prompts*, structured instructions in natural language that can be interpreted by a generative AI model [14]. Generating and tuning these prompts in such a way that the models generate the desired outcomes (with e.g. prompt engineering [57]) has become central to controlling their outputs, ensuring their safety, and studying their behaviour. For example, the Adversarial Nibbler competition is a "data-centric AI competition that aims to construct a diverse set of insightful examples of long tail problems for text-to-image models" [54]. Prompt engineering has also facilitated the emergence of some data management practices around prompts, e.g. by sharing them as datasets [37], on GitHub[21], Hugging Face[22], etc. These practices show that prompt management is becoming increasingly important. However, sharing prompts as ordinary datasets may not be sufficient to address the various challenges around them: How can prompts be adequately documented? How can a prompt be updated with a better version? What is the effect of a prompt in its generated output across different models? What is the provenance of a prompt? How can prompt datasets be integrated in a machine interoperable and executable way? How can prompts be searched in a systematic manner? How can prompt sharing be organised in a scalable way? How are prompts automatically generated by agentic AI systems? How are prompt workflows laid out and executed across different AI systems? Given the importance of prompts in modelling AI behaviour, these questions are especially important in the context of RAI, explainability and transparency.

To address these questions on *AI workflows* and *prompt management*, we propose **WikiPrompts**, a collaborative knowledge graph of prompts and prompt workflows typically used in generative AI applications. WikiPrompts combines the ideas of structured collaborative wikis like Wikidata [66], and prompt management, interaction, sharing, and documentation. WikiPrompts facilitates a systematic description and documentation of prompts through *properties* and *values*, similarly to how Wikidata documents items as entities of the real world.

WikiPrompts has the following features:

- It is a collaborative knowledge graph specifically about prompts and prompt workflows for generative AI systems, editable by anyone;
- It proposes a data model, or ontology, that specifies vocabularies and metadata terms to document prompts and their workflows
- It can be systematically queried by people and especially applications, incentivising prompt reuse and bringing prompts closer to FAIR [68]
- Its type system allows the documentation of various types of prompts: simple prompts, which can be reused directly; prompt templates, with variables that can be instantiated at runtime; chain-of-thought (CoT) prompts; etc
- It allows *prompt composition*, *prompt chaining* and *prompt workflows* by leveraging internal linking. Various workflow models can be implemented but our current work uses the W3C PROV provenance model as basis [48]
- It builds in the ability to automate the execution of prompts through bots that regularly update the output produced by each prompt in a variety of generative AI models, automatically feeding back content into the knowledge graph, and facilitating the study of prompt-output relations; this strengthens the value of the platform in documenting model inputs and outputs [36]
- Since it is based on a wiki, provenance is fully supported via the edit history of each contributed prompt (i.e. who edited what, when, how)

We have built a preliminary version of WikiPrompts with Wikibase[23], the opens-source software that powers Wikidata, and is available online[24] (see Fig. 4).

Many challenges remain open for WikiPrompts. First, WikiPrompts is in its early stages and faces the steep start of any community-building project: this is thus a call to arms to join and contribute. Second, its data model is tangential to various competing efforts for modelling prompt documentation and needs to find a spot between technical accuracy and user friendliness. Third, more work is needed in thinking about use cases that could link WikiPrompts to other KGs, e.g. in how new prompts could be automatically derived from Wikidata items [5], or automatically generated in general rather than having them written by people. Finally, the infrastructure needed for WikiPrompts, both for its knowledge graph operations (e.g. importing, querying) and for prompt execution (i.e. output documentation) needs long-term planning in terms of scalability, as it currently relies on accessible but limited resources (e.g. Wikibase Cloud[25], Hugging Face Serverless Inference API[26]). An important observed limitation of using KGs for prompt and workflow management is the inherent textual nature of modern AI inputs and outputs: some might be hard to represent symbolically and query, pointing at potential scalability issues.

4. Discussion and Vision

We reflect here on the common themes across the various layers of KG.GOV and the use cases.

[21] See e.g. https://github.com/f/awesome-chatgpt-prompts, https://github.com/ai-boost/awesome-prompts and https://github.com/promptslab/Awesome-Prompt-Engineering.

[22] https://huggingface.co/datasets/Gustavosta/Stable-Diffusion-Prompts

[23] https://wikiba.se/

[24] https://wikiprompts.wikibase.cloud/

[25] https://www.wikibase.cloud/

[26] https://huggingface.co/docs/api-inference/en/index

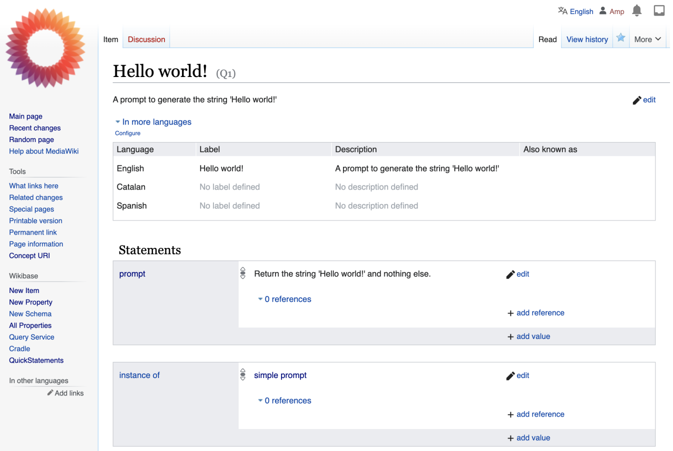


Figure 4: Snapshot of a WikiPrompts item, documenting a prompt with properties and values.

Although **modelling data** is well exemplified with Croissant, as providing a vocabulary for ML datasets is primarily a data modelling activity, we notice that data modelling is also a challenge in WikiPrompts (i.e. how to describe a prompt and complex prompt workflows appropriately?) and in Multimodal Transformations (MT; i.e. how to explain the process that generated this content?). We observe here requirements of data modelling at different levels: from the high-level, abstract metadata to describe datasets and their publishing, authors, versioning, licensing, etc. (Croissant); to the more specific, low-level of behaviour control and monitoring (WikiPrompts) and fine-grained provenance of how data came to be (MT). These levels are interrelated and reference each other, but importantly our framework proposes to use KGs for their governance and representation, thus facilitating interoperability, exchange, integration, and to the extent possible, automation. Queries about data licensing for ML (Croissant) will most probably need to be combined with behaviour (WikiPrompts) and provenance traces (MT) to e.g. assess compliance; allowing for traceability from the technical to the social and legal dimensions.

This social aspect is another common, orthogonal theme to all dimensions and use cases, making us reflect on **the role of citizens and volunteer communities** in the governance of AI through KGs. Although in WikiPrompts the role of such a community is an essential part of its functioning and content provision (i.e. similar to Wikidata or Wikipedia), the involvement of the civic society in Croissant and MT are also fundamental for the appropriate gathering of requirements. In Croissant, new use cases and needs will arise when developers are confronted with the consumption of ML datasets and the need of explicitly linking them with the models they train them with and downstream decisions. In MT, their intervention is key in at least two ways: (a) as main actors in the evaluation of methods that propose AI-generated content, as many automated metrics continue to fail to assess issues of semantic content, copyright infringement, relatedness, etc. [2]; and (b) as a volunteer force that monitors and seeks for ethical and moral issues in models deployed in society [54] in a similar way to how communities check open-source license infringements in commercial software [42]. An open question is still how to address issues of scalability and infrastructure for the public governance of AI. Wikimedia has been very successful in providing equal access to knowledge for all (e.g. Wikidata's SPARQL endpoint[27]) through a model of donations and grants (for personnel, hardware, etc.); but whether that model can work for publicly governed AI infrastructure remains to be seen.

This leads to the final common theme: **evaluation**. The assessment of how good do KG data models, KG-enhanced AI behaviour platforms, and KG-powered AI-generated content work for a stable, long-term AI governance poses an enormous challenge. One place to look here are holistic evaluation metrics and dashboards such as HELM [41] (Holistic Evaluation of Language Models). Although WikiPrompts proposes to address some common issues, an extension, repurposing, or abstraction of HELM for AI infrastructures that use KGs as their governance backbone will be needed to answer questions such as: what combinations of datasets, prompts and models score low in metrics related to ethics and bias for AI-generated multimodal KG content? Some work on this direction, including benchmarks and evaluation frameworks for generative AI, is already on the way [25, 44]. What architectures allow fully traceable provenance and explainable workflows, from the final generated content to the original training datasets [29]? Such a holistic evaluation framework will surely need more vocabularies for metadata descriptions (Croissant) and prompt-based workflows (WikiPrompts)–besides metrics, models, datasets, etc.– in order to provide assessments that can be used to increase trust in AI. This could be combined with a different use of KGs for AI governance in ongoing efforts in neurosymbolic approaches to generative AI [31] and common-sense KGs [34], in which KGs play a central role as controllers and safeguard policy providers in end-to-end architectures. With multimodality at play, scene or music symbolic representations [11, 45] could play this role in novel ways of thinking about Multimodal Retrieval Augmented Generation (MMRAG) [69]. Crucially for evaluation, we should look into assessing how much of that improved safety and safeguarding is due to the use of KGs [13].

Here, we have presented KG.gov, a framework for thinking about KGs as the backbone of data governance in AI. With this work, we intend to show that KGs can have a central role in the provision of data models, behaviour descriptions, and as enablers of data representations that can be accessed by all. However, this also implies new challenges and research questions for KGs, which stem from our observed limitations and common themes: (1) data

[27]https://query.wikidata.org/

and system aspects will continue to raise requirements for representations and metadata vocabularies, which will require extensions and combinations; (2) scalability of KGs for governance poses issues around supporting the representation and querying of multimodal data, abundant text, and complex workflows; (3) uptake by communities and human acceptance to work with interfaces to AI systems; and (4) evaluation of KG data models, AI systems impacting KGs, and KGs impacting AI systems, which remains as possibly the biggest challenge in both AI governance and AI in general.

Playing different roles, either as providers of vocabulary terms to describe ML datasets (Data modelling), workflows to publish and consume AI prompts on the Web (Describing behaviour), or as hosts of multimedia, accessible content that help us think about inherently multimodal knowledge representations (Alternative representations), KGs present themselves as a useful tool to automate AI governance. In this paper we have described various examples of this. If appropriately used, Croissant license metadata can be used to semi-automatically trace the provenance of data used in training, relate it to model properties, document model outputs, and assess compliance to regulations. By using AI as a tool for generating and translating representations, we can think about new tasks such as multimodal KG completion from a safety perspective, where governing KG have an active role in safeguarding and limiting the generated media, and flag and annotate them if they violate the conditions they specify. The management of knowledge bases of AI workflows and prompts can help in automating knowledge management processes, from acquisition to formalisation and use. The future of AI governance seems to stem from this at least partial automation facilitated by KG data and workflows. Besides this automation, however, the role of humans in a KG-powered AI governance will continue to be essential as high-quality data generators, initiators of questioning, assemblers of communities and proponents of better ways of assessing the outcomes and impacts of AI-generated content; but primarily as the main actors studying the metadata, prompts, workflows and representations that AI entails.

Acknowledgements

This work was partially supported by the HE project MuseIT, cofunded by the European Union under Grant Agreement No. 101061441. The views and opinions expressed in this work are those of the authors and do not necessarily reflect the views of the European Union or the European Research Executive Agency. Part of this work was supported by King's College London High Performance Computing environment: King's College London. (2024). King's Computational Research, Engineering and Technology Environment (CREATE). Retrieved November 14, 2024, from `https://doi.org/10.18742/rnvf-m076`

References

[1] David Abián, Albert Meroño-Peñuela, and Elena Simperl. An analysis of content gaps versus user needs in the wikidata knowledge graph. In *International Semantic Web Conference*, pages 354–374. Springer, 2022.

[2] Raia Abu Ahmad, Martin Critelli, Sefika Efeoglu, Eleonora Mancini, Célian Ringwald, Xingyue Zhang, and Albert Merono Penuela. Draw me like my triples: Leveraging generative ai for wikidata image completion. In *The 4th Wikidata Workshop*, 2023.

[3] Mubashara Akhtar, Omar Benjelloun, Costanza Conforti, Pieter Gijsbers, Joan Giner-Miguelez, Nitisha Jain, Michael Kuchnik, Quentin Lhoest, Pierre Marcenac, Manil Maskey, et al. Croissant: A metadata format for ml-ready datasets. In *Proceedings of the Eighth Workshop on Data Management for End-to-End Machine Learning*, pages 1–6, 2024.

[4] Rohan Anil, Sebastian Borgeaud, Yonghui Wu, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew M. Dai, Anja Hauth, Katie Millican, David Silver, Slav Petrov, Melvin Johnson, Ioannis Antonoglou, Julian Schrittwieser, Amelia Glaese, Jilin Chen, Emily Pitler, Timothy P. Lillicrap, Angeliki Lazaridou, Orhan Firat, James Molloy, Michael Isard, Paul Ronald Barham, Tom Hennigan, Benjamin Lee, Fabio Viola, Malcolm Reynolds, Yuanzhong Xu, Ryan Doherty, Eli Collins, Clemens Meyer, Eliza Rutherford, Erica Moreira, Kareem Ayoub, Megha Goel, George Tucker, Enrique Piqueras, Maxim Krikun, Iain Barr, Nikolay Savinov, Ivo Danihelka, Becca Roelofs, Anaïs White, Anders Andreassen, Tamara von Glehn, Lakshman Yagati, Mehran Kazemi, Lucas Gonzalez, Misha Khalman, Jakub Sygnowski, and et al. Gemini: A family of highly capable multimodal models. *CoRR*, abs/2312.11805, 2023.

[5] Ryan Brate, Minh-Hoang Dang, Fabian Hoppe, Yuan He, Albert Meroño-Peñuela, and Vijay Sadashivaiah. Improving language model predictions via prompts enriched with knowledge graphs. In *DL4KG@ ISWC2022*, 2022.

[6] Arno Breitfuss, Karen Errou, Anelia Kurteva, and Anna Fensel. Representing emotions with knowledge graphs for movie recommendations. *Future Generation Computer Systems*, 125:715–725, 2021.

[7] Hao Chen and Jun Hou. Intelligent data governance: building an enterprise data management system using KG and LLM. In *Proceedings of the 2024 International Conference on Cloud Computing and Big Data*, pages 266–271, 2024.

[8] Zhuo Chen, Yichi Zhang, Yin Fang, Yuxia Geng, Lingbing Guo, Xiang Chen, Qian Li, Wen Zhang, Jiaoyan Chen, Yushan Zhu, Jiaqi Li, Xiaoze Liu, Jeff Z. Pan, Ningyu Zhang, and Huajun Chen. Knowledge graphs meet multi-modal learning: A comprehensive survey. *CoRR*, abs/2402.05391, 2024.

[9] Tek Raj Chhetri, Anelia Kurteva, Rance J DeLong, Rainer Hilscher, Kai Korte, and Anna Fensel. Data protection by design tool for automated gdpr compliance verification based on semantically modeled informed consent. *Sensors*, 22(7):2763, 2022.

[10] I. Glenn Cohen, Theodoros Evgeniou, and Martin Husovec. Navigating the new risks and regulatory challenges of GenAI, 2023. Available at https://hbr.org/2023/11/navigating-the-new-risks-and-regulatory-challenges-of-genai, Accessed on 14 Nov 2024.

[11] Jacopo de Berardinis, Albert Meroño-Peñuela, Andrea Poltronieri, and Valentina Presutti. Choco: a chord corpus and a data transformation workflow for musical harmony knowledge graphs. In *Manuscript under review*, 2023.

[12] Marina De Vos, Sabrina Kirrane, Julian Padget, and Ken Satoh. Odrl policy modelling and compliance checking. In *Rules and Reasoning: Third International Joint Conference, RuleML+RR 2019, Bolzano, Italy, September 16–19, 2019, Proceedings 3*, pages 36–51. Springer, 2019.

[13] Chiara Di Bonaventura[1], Arianna Muti, and Marco Antonio Stranisci. O-dang at hodi and haspeede3: A knowledge-enhanced approach to homotransphobia and hate speech detection. In *EVALITA Proceedings of the Eighth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian Final Workshop: Parma, Italy, September 7-8th, 2023*, volume 113, page 26. Accademia University Press, 2024.

[14] Mohamad Diab, Julian Herrera, Bob Chernow, and Coco Mao.

Stable diffusion prompt book, 2022.

[15] Patrick Esser, Sumith Kulal, Andreas Blattmann, Rahim Entezari, Jonas Müller, Harry Saini, Yam Levi, Dominik Lorenz, Axel Sauer, Frederic Boesel, Dustin Podell, Tim Dockhorn, Zion English, and Robin Rombach. Scaling rectified flow transformers for high-resolution image synthesis. In *Forty-first International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21-27, 2024*. OpenReview.net, 2024.

[16] EU Parliament and Concil. Artificial Intelligence Act. *Official Journal of the European Union*, April 2021.

[17] EU Parliament and Concil. Directive 95/46/ec (General Data Protection Regulation). *Official Journal of the European Union, L119*, May 2016.

[18] Kaniz Fatema, Ensar Hadziselimovic, Christophe Debruyne, and Dave Lewis. Compliance through informed consent: Semantic based consent permission and data management model. *PrivOn@ ISWC*, 2017.

[19] Dieter Fensel. Ontology-based knowledge management. *Computer*, 35(11):56–59, 2002.

[20] Dieter Fensel, Federico Michele Facca, Elena Simperl, and Ioan Toma. *Semantic web services*, volume 357. Springer, 2011.

[21] Dieter Fensel and Dieter Fensel. *Ontologies*. Springer, 2001.

[22] National Institute for Standards and Technology. Trustworthy and responsible ai.

[23] Mikhail Galkin, Sören Auer, Maria-Esther Vidal, and Simon Scerri. Enterprise knowledge graphs: A semantic approach for knowledge management in the next generation of enterprise information systems. In *International Conference on Enterprise Information Systems*, volume 2, pages 88–98. SCITEPRESS, 2017.

[24] Leo Gao, Stella Biderman, Sid Black, Laurence Golding, Travis Hoppe, Charles Foster, Jason Phang, Horace He, Anish Thite, Noa Nabeshima, et al. The pile: An 800gb dataset of diverse text for language modeling. *arXiv preprint arXiv:2101.00027*, 2020.

[25] Dhruba Ghosh, Hanna Hajishirzi, and Ludwig Schmidt. Geneval: An object-focused framework for evaluating text-to-image alignment, 2023.

[26] Rohit Girdhar, Alaaeldin El-Nouby, Zhuang Liu, Mannat Singh, Kalyan Vasudev Alwala, Armand Joulin, and Ishan Misra. Imagebind one embedding space to bind them all. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2023, Vancouver, BC, Canada, June 17-24, 2023*, pages 15180–15190. IEEE, 2023.

[27] Delaram Golpayegani, Isabelle Hupont, Cecilia Panigutti, Harshvardhan J. Pandit, Sven Schade, Declan O'Sullivan, and Dave Lewis. Ai cards: Towards an applied framework for machine-readable ai and risk documentation inspired by the eu ai act, 2024.

[28] Delaram Golpayegani, Harshvardhan J Pandit, and Dave Lewis. Airo: An ontology for representing ai risks based on the proposed eu ai act and iso risk management standards. In *Towards a Knowledge-Aware AI*, pages 51–65. IOS Press, 2022.

[29] Jack Hardinges, Elena Simperl, and Nigel Shadbolt. We must fix the lack of transparency around the data used to train foundation models. *Harvard Data Science Review*, 2023.

[30] Leah Henrickson and Albert Meroño-Peñuela. Prompting meaning: a hermeneutic approach to optimising prompt engineering with chatgpt. *AI & SOCIETY*, pages 1–16, 2023.

[31] Pascal Hitzler and Md Kamruzzaman Sarker. Neuro-symbolic artificial intelligence: The state of the art. 2022.

[32] Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia d'Amato, Gerard De Melo, Claudio Gutierrez, Sabrina Kirrane, José Emilio Labra Gayo, Roberto Navigli, Sebastian Neumaier, et al. Knowledge graphs. *ACM Computing Surveys (Csur)*, 54(4):1–37, 2021.

[33] Soheil Human, Harshvardhan J Pandit, Victor Morel, Cristiana Santos, Martin Degeling, Arianna Rossi, Wilhelmina Botes, Vitor Jesus, and Irene Kamara. Data protection and consenting communication mechanisms: Current open proposals and challenges. In *2022 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, pages 231–239. IEEE, 2022.

[34] Filip Ilievski, Pedro Szekely, and Bin Zhang. CSKG: The Commonsense Knowledge Graph. In *The Semantic Web: 18th International Conference, ESWC 2021, Virtual Event, June 6–10, 2021, Proceedings 18*, pages 680–696. Springer, 2021.

[35] Arcchit Jain, Tal Friedman, Ondrej Kuzelka, Guy Van den Broeck, and Luc De Raedt. Scalable rule learning in probabilistic knowledge bases. In *The 1st Conference On Automated Knowledge Base Construction (AKBC)*, may 2019.

[36] Marijn Janssen, Paul Brous, Elsa Estevez, Luis S. Barbosa, and Tomasz Janowski. Data governance: Organizing data for trustworthy artificial intelligence. *Government Information Quarterly*, 37(3):101493, 2020.

[37] Nithish Kannen, Arif Ahmad, Marco Andreetto, Vinodkumar Prabhakaran, Utsav Prabhu, Adji Bousso Dieng, Pushpak Bhattacharyya, and Shachi Dave. Beyond aesthetics: Cultural competence in text-to-image models, 2024.

[38] Sehoon Kim, Suhong Moon, Ryan Tabrizi, Nicholas Lee, Michael W. Mahoney, Kurt Keutzer, and Amir Gholami. An LLM compiler for parallel function calling. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp, editors, *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 24370–24391. PMLR, 21–27 Jul 2024.

[39] Sabrina Kirrane, Javier D Fernández, Piero Bonatti, Uros Milosevic, Axel Polleres, and Rigo Wenning. The special-k personal data processing transparency and compliance platform. *arXiv preprint arXiv:2001.09461*, 2020.

[40] Anelia Kurteva. Making sense of consent with knowledge graphs, 2022. PhD Thesis. Semantic Technology Institute (STI) Innsbruck, University of Innsbruck, Austria.

[41] Percy Liang, Rishi Bommasani, Tony Lee, Dimitris Tsipras, Dilara Soylu, Michihiro Yasunaga, Yian Zhang, Deepak Narayanan, Yuhuai Wu, Ananya Kumar, Benjamin Newman, Binhang Yuan, Bobby Yan, Ce Zhang, Christian Cosgrove, Christopher D. Manning, Christopher Ré, Diana Acosta-Navas, Drew A. Hudson, Eric Zelikman, Esin Durmus, Faisal Ladhak, Frieda Rong, Hongyu Ren, Huaxiu Yao, Jue Wang, Keshav Santhanam, Laurel Orr, Lucia Zheng, Mert Yuksekgonul, Mirac Suzgun, Nathan Kim, Neel Guha, Niladri Chatterji, Omar Khattab, Peter Henderson, Qian Huang, Ryan Chi, Sang Michael Xie, Shibani Santurkar, Surya Ganguli, Tatsunori Hashimoto, Thomas Icard, Tianyi Zhang, Vishrav Chaudhary, William Wang, Xuechen Li, Yifan Mai, Yuhui Zhang, and Yuta Koreeda. Holistic evaluation of language models, 2023.

[42] GNU General Public License. Gnu general public license. *Retrieved December*, 25:2014, 1989.

[43] Chunhyeok Lim, Shiyong Lu, Artem Chebotko, and Farshad Fotouhi. Prospective and retrospective provenance collection in scientific workflow environments. In *2010 IEEE International Conference on Services Computing*, pages 449–456. IEEE, 2010.

[44] Zhiqiu Lin, Deepak Pathak, Baiqi Li, Jiayao Li, Xide Xia, Graham Neubig, Pengchuan Zhang, and Deva Ramanan. Evaluating text-to-visual generation with image-to-text generation, 2024.

[45] Albert Meroño-Peñuela, Rinke Hoekstra, Aldo Gangemi, Peter Bloem, Reinier de Valk, Bas Stringer, Berit Janssen, Victor de Boer, Alo Allik, Stefan Schlobach, and Kevin Page. The midi linked data cloud. In *The Semantic Web – ISWC 2017*, Cham, 2017. Springer International Publishing.

[46] Microsoft. The microsoft responsible ai standard v2, June 2022.

[47] Victor Mireles, Stefan Gindl, Sotiris Karampatakis, and Michael Boch. D3. 7 data governance, trusts knowledge graph i. *Nature*, 30:06, 2021.

[48] Paolo Missier, Khalid Belhajjame, and James Cheney. The w3c prov family of specifications for modelling provenance metadata. In *Proceedings of the 16th International Conference on Extending Database Technology*, pages 773–776, 2013.

[49] Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, In-

ioluwa Deborah Raji, and Timnit Gebru. Model cards for model reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, FAT* '19, page 220–229. ACM, January 2019.

[50] Kevin Paeth, Daniel Atherton, Nikiforos Pittaras, Heather Frase, and Sean McGregor. Lessons for editors of ai incidents from the ai incident database, 2024.

[51] Harshvardhan J Pandit, Axel Polleres, Bert Bos, Rob Brennan, Bud Bruegger, Fajar J Ekaputra, Javier D Fernández, Roghaiyeh Gachpaz Hamed, Elmar Kiesling, Mark Lizar, et al. Creating a vocabulary for data privacy: The first-year report of data privacy vocabularies and controls community group (dpvcg). In *On the Move to Meaningful Internet Systems: OTM 2019 Conferences: Confederated International Conferences: CoopIS, ODBASE, C&TC 2019, Rhodes, Greece, October 21–25, 2019, Proceedings*, pages 714–730. Springer, 2019.

[52] Jay Pujara, Hui Miao, Lise Getoor, and William Cohen. Using semantics statistics to turn data into knowledge. *AI Magazine*, 36(1):65–74, 2015.

[53] Mahima Pushkarna, Andrew Zaldivar, and Oddur Kjartansson. Data cards: Purposeful and transparent dataset documentation for responsible ai. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, pages 1776–1826, 2022.

[54] Jessica Quaye, Alicia Parrish, Oana Inel, Charvi Rastogi, Hannah Rose Kirk, Minsuk Kahng, Erin Van Liemt, Max Bartolo, Jess Tsang, Justin White, Nathan Clement, Rafael Mosquera, Juan Ciro, Vijay Janapa Reddi, and Lora Aroyo. Adversarial nibbler: An open red-teaming method for identifying diverse harms in text-to-image generation. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '24, page 388–406, New York, NY, USA, 2024. Association for Computing Machinery.

[55] Miriam Redi, Martin Gerlach, Isaac Johnson, Jonathan Morgan, and Leila Zia. A taxonomy of knowledge gaps for wikimedia projects (second draft), 2021.

[56] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022*, pages 10674–10685. IEEE, 2022.

[57] Pranab Sahoo, Ayush Kumar Singh, Sriparna Saha, Vinija Jain, Samrat Mondal, and Aman Chadha. A systematic survey of prompt engineering in large language models: Techniques and applications. *arXiv preprint arXiv:2402.07927*, 2024.

[58] Maarten Sap, Ronan LeBras, Emily Allaway, Chandra Bhagavatula, Nicholas Lourie, Hannah Rashkin, Brendan Roof, Noah A. Smith, and Yejin Choi. Atomic: An atlas of machine commonsense for if-then reasoning, 2019.

[59] Yonadav Shavit, Sandhini Agarwal, Miles Brundage, Steven Adler, Cullen O'Keefe, Rosie Campbell, Teddy Lee, Pamela Mishkin, Tyna Eloundou, Alan Hickey, et al. Practices for governing agentic ai systems. *Research Paper, OpenAI, December*, 2023.

[60] Sopra Steria. Sopra Steria Next predicts AI market will double to $1,270 billion by 2028. `https://www.soprasteria.co.uk/insights/press-releases/details/sopra-steria-next-predicts-ai-market-will-double-to-1\-270-billion-by-2028`, 2024.

[61] Julia Stoyanovich and Bill Howe. Nutritional labels for data and models. *A Quarterly bulletin of the Computer Society of the IEEE Technical Committee on Data Engineering*, 42(3), 2019.

[62] Amar Tauqeer, Anelia Kurteva, Tek Raj Chhetri, Albin Ahmeti, and Anna Fensel. Automated gdpr contract compliance verification using knowledge graphs. *Information*, 13(10):447, 2022.

[63] Ilaria Tiddi and Stefan Schlobach. Knowledge graphs as tools for explainable machine learning: A survey. *Artificial Intelligence*, 302:103627, 2022.

[64] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in Neural Information Processing Systems*, 2017.

[65] Bertie Vidgen, Adarsh Agrawal, Ahmed M. Ahmed, Victor Akinwande, Namir Al-Nuaimi, Najla Alfaraj, Elie Alhajjar, Lora Aroyo, Trupti Bavalatti, Max Bartolo, Borhane Blili-Hamelin, Kurt Bollacker, Rishi Bomassani, Marisa Ferrara Boston, Siméon Campos, Kal Chakra, Canyu Chen, Cody Coleman, Zacharie Delpierre Coudert, Leon Derczynski, Debojyoti Dutta, Ian Eisenberg, James Ezick, Heather Frase, Brian Fuller, Ram Gandikota, Agasthya Gangavarapu, Ananya Gangavarapu, James Gealy, Rajat Ghosh, James Goel, Usman Gohar, Sujata Goswami, Scott A. Hale, Wiebke Hutiri, Joseph Marvin Imperial, Surgan Jandial, Nick Judd, Felix Juefei-Xu, Foutse Khomh, Bhavya Kailkhura, Hannah Rose Kirk, Kevin Klyman, Chris Knotz, Michael Kuchnik, Shachi H. Kumar, Srijan Kumar, Chris Lengerich, Bo Li, Zeyi Liao, Eileen Peters Long, Victor Lu, Sarah Luger, Yifan Mai, Priyanka Mary Mammen, Kelvin Manyeki, Sean McGregor, Virendra Mehta, Shafee Mohammed, Emanuel Moss, Lama Nachman, Dinesh Jinenhally Naganna, Amin Nikanjam, Besmira Nushi, Luis Oala, Iftach Orr, Alicia Parrish, Cigdem Patlak, William Pietri, Forough Poursabzi-Sangdeh, Eleonora Presani, Fabrizio Puletti, Paul Röttger, Saurav Sahay, Tim Santos, Nino Scherrer, Alice Schoenauer Sebag, Patrick Schramowski, Abolfazl Shahbazi, Vin Sharma, Xudong Shen, Vamsi Sistla, Leonard Tang, Davide Testuggine, Vithursan Thangarasa, Elizabeth Anne Watkins, Rebecca Weiss, Chris Welty, Tyler Wilbers, Adina Williams, Carole-Jean Wu, Poonam Yadav, Xianjun Yang, Yi Zeng, Wenhui Zhang, Fedor Zhdanov, Jiacheng Zhu, Percy Liang, Peter Mattson, and Joaquin Vanschoren. Introducing v0.5 of the ai safety benchmark from mlcommons, 2024.

[66] Denny Vrandečić and Markus Krötzsch. Wikidata: a free collaborative knowledgebase. *Communications of the ACM*, 57(10):78–85, 2014.

[67] Jan Philip Wahle, Terry Ruas, Saif M Mohammad, Norman Meuschke, and Bela Gipp. Ai usage cards: Responsibly reporting ai-generated content. In *2023 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, pages 282–284. IEEE, 2023.

[68] Mark D Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E Bourne, et al. The fair guiding principles for scientific data management and stewardship. *Scientific data*, 3(1):1–9, 2016.

[69] Peng Xia, Kangyu Zhu, Haoran Li, Tianze Wang, Weijia Shi, Sheng Wang, Linjun Zhang, James Zou, and Huaxiu Yao. Mmed-rag: Versatile multimodal rag system for medical vision language models, 2024.

[70] Zhengyuan Yang, Linjie Li, Kevin Lin, Jianfeng Wang, Chung-Ching Lin, Zicheng Liu, and Lijuan Wang. The dawn of lmms: Preliminary explorations with gpt-4v(ision). *CoRR*, abs/2309.17421, 2023.

[71] Bohui Zhang, Ioannis Reklos, Nitisha Jain, Albert Meroño Peñuela, and Elena Simperl. Using Large Language Models for Knowledge Engineering (LLMKE): A Case Study on Wikidata. *KBC-LM'23: Knowledge Base Construction from Pre-trained Language Models workshop at ISWC 2023*, 2023.