Trust, Accountability, and Autonomy in Knowledge Graph-based AI for Self-determination

Luis-Daniel Ibáñez ✉ iD
Department of Electronics and Computer Science, University of Southampton, UK

John Domingue ✉ iD
Knowledge Media Institute, The Open University, UK

Sabrina Kirrane ✉ iD
Institute for Information Systems & New Media, Vienna University of Economics and Business, Austria

Oshani Seneviratne ✉ iD
Department of Computer Science, Rensselaer Polytechnic Institute, USA

Aisling Third ✉ iD
Knowledge Media Institute, The Open University, UK

Maria-Esther Vidal ✉ iD
Leibniz University of Hannover & TIB-Leibniz Information Centre of Science and Technology, Germany

Abstract

Knowledge Graphs (KGs) have emerged as fundamental platforms for powering intelligent decision-making and a wide range of Artificial Intelligence (AI) services across major corporations such as Google, Walmart, and AirBnb. KGs complement Machine Learning (ML) algorithms by providing data context and semantics, thereby enabling further inference and question-answering capabilities. The integration of KGs with neuronal learning (e.g., Large Language Models (LLMs)) is currently a topic of active research, commonly named neuro-symbolic AI. Despite the numerous benefits that can be accomplished with KG-based AI, its growing ubiquity within online services may result in the loss of self-determination for citizens as a fundamental societal issue. The more we rely on these technologies, which are often centralised, the less citizens will be able to determine their own destinies. To counter this threat, AI regulation, such as the European Union (EU) AI Act, is being proposed in certain regions. The regulation sets what technologists need to do, leading to questions concerning: How can the output of AI systems be trusted? What is needed to ensure that the data fuelling and the inner workings of these artefacts are transparent? How can AI be made accountable for its decision-making? This paper conceptualises the foundational topics and research pillars to support KG-based AI for self-determination. Drawing upon this conceptual framework, challenges and opportunities for citizen self-determination are illustrated and analysed in a real-world scenario. As a result, we propose a research agenda aimed at accomplishing the recommended objectives.

1 Introduction

Modern *Artificial Intelligence* (AI) can be traced back to a workshop held at Dartmouth College in the summer of 1956 [67] and is most commonly defined as the use of computers to simulate human intelligence, in particular human reasoning, learning, and problem-solving. Since 1956, AI has lived through times of increased interest and funding, and also 'AI Winters', such as after the 1974 Lighthill report [64], when overall funding was reduced. Over the last few years, however, funding and interest in AI have been high and exploded in November 2022, when ChatGPT, a type of Generative AI, was announced by OpenAI, exposing the power of Large Language Models (LLMs) to the general public. Since its release, ChatGPT has become the fastest-growing app in history, reaching 100M users in just two months, and is now estimated to have 200M users. Generative AI will continue to grow following a significant investment by Microsoft into OpenAI and announcements by Microsoft and Google on how Generative AI will be

embedded in future products [37]. Data-centric AI [115] recognises the immense value of data as crucial resources for training, optimising, and evaluating AI systems. Databricks, a prominent AI company, has defined data-centric AI as the challenge of designing processes for data collection, labelling, and quality monitoring in machine learning (ML) datasets [87] highlighting the need for continuous re-running and re-training, actionable monitoring, and the difficulties of incorporating data inaccessible to human annotators due to privacy concerns as primary research directions. Knowledge Graphs have been used both as a resource and as a structure to support data-centric AI processes. The term *Knowledge Graph* (KG) was first introduced by Google in 2012, and is usually defined as a type of knowledge structure that uses a graph data model to integrate data. KGs are strongly linked to the work of the Semantic Web community, which first began in around 2001 and was introduced in a seminal paper by Tim Berners-Lee [14]. The Semantic Web initiative produced a stack of web standards on which KGs are based. These include the Resource Description Framework (RDF), where data is encoded as subject-predicate-object triples, and the Web Ontology Language (OWL), a set of web-based languages mostly based on description logic. The common theme of these semantic representations is that they facilitate the publishing, use, and re-use of data at the web scale. In particular, they allow disparate heterogeneous data sources to be integrated continuously at scale. Over the past decade, KGs have become a mainstay for a number of key large-scale applications found online. For example, KGs underpin Google Search, which saw 5,900,000 searches in just one minute in April 2022. Similarly, the same minute saw 1,700,000 pieces of content shared on Facebook, 1,000,000 hours streamed, and 347,200 tweets shared on Twitter. All of this content and data are linked to a plethora of AI services that have increasingly been based on KGs, as mentioned above, founded upon machine-readable data and schema representations based on a web stack of standards. AI services cover a wide number of areas, including content recommendation, user input prediction, as well as large-scale search and discovery and form the basis for the business models of companies like Google, Netflix, Spotify, and Facebook. Given the above we define KG-based AI as an AI system (replicating some aspect of human intelligence) based on a KG possibly using the web standards produced by the Semantic Web community.

In addition to privacy concerns, there has been a growing worry about how personal data can be abused and, thus, how AI services impinge on citizen rights. For example, the over-centralisation of data and its misuse led Sir Tim Berners-Lee to call the Web 'anti-human' in an interview in 2018 [19]. Since 2016, hundreds of United States (US) Immigration and Customs Enforcement employees have faced investigations into abuse of confidential law enforcement databases, including stalking and harassment, to passing data to criminals [70]. The subject of much of the proposed legislation today is ensuring that digital platforms, including AI platforms, provide real societal benefit. Within Europe, the proposed European Union (EU) AI Act[1] aims to support safe AI that respects fundamental human rights. The regulation sets what technologists need to do. The concept of data *self-determination*, which is often used in a legal context, implies that individuals are not only aware of who knows what about them but can also influence data processing that concerns them [61]. Given that nowadays, data processing is conducted by opaque AI algorithms behind corporate firewalls, sometimes even without our knowledge, data self-determination is harder than ever before. When it comes to trust in web data and services, Berners-Lee and Fischetti [13] envisaged an "Oh yeah?" button embedded into Web browsers that would provide justifications as to why a page or a service should be trusted. Alas, their vision was never realised in popular web browsers[2]. Instead, we have dedicated websites, e.g.,

---

[1] `https://artificialintelligenceact.eu/`
[2] However, a linked browser prototype, the Tabulator, incorporated this feature in an *Justification UI* (`http:`
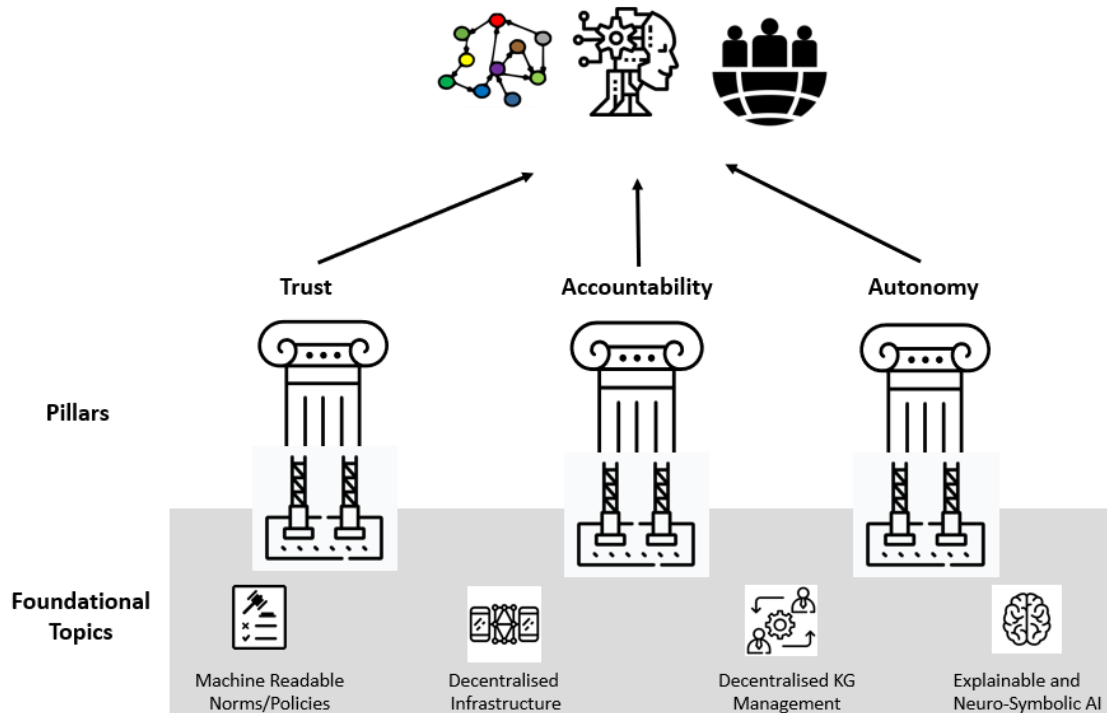
Figure 1 KG-based AI for Self-determination Conceptualisation. KG-based AI for self-determination is supported by the pillars of trust, accountability, and autonomy, built upon the foundational topics of machine-readable norms and policies; decentralised infrastructure; decentralised KG management; and explainable neuro-symbolic AI.

the Ecommerce Europe Trustmark[3] that are used to perform company reputation checks and fact-checking websites, such as Snopes[4], that can be used to check the validity of information posted online. Although some automated fact-checking techniques have been proposed [89], they are used solely for developing trust in information resources and cannot provide any guarantees with respect to AI-based data processing. Moving beyond trust towards accountability, policies have already been used to specify legal data processing requirements that serve as the basis for automated compliance checking, for example, [83]. But what happens when service providers or AI algorithms do not comply? How far can technology go in terms of helping us to determine non-compliance and to make service providers accountable for their actions?

In this paper, we propose a research agenda for ensuring that KG-based AI approaches contribute to user self-determination instead of hindering it. Our vision, which is depicted in Figure 1, is structured around three pillar research topics - trust, accountability, and autonomy - that represent the desired goals for how AI can benefit society and facilitate self-determination. The pillars combine fundamental principles of the proposed EU AI Act and self-determination theory. Both trust and accountability are imperative for safeguarding against adverse impact caused by AI systems, while autonomy is critical for ensuring individuals are able to determine their own destiny. The pillars are supported via four foundational research topics - machine-readable

---

//dig.csail.mit.edu/TAMI/2008/JustificationUI/howto.html#useTab).

[3] https://ecommercetrustmark.eu/

[4] https://www.snopes.com/

norms and policies are needed for humans to declare regulatory frameworks, privacy and usage constraints that can be interpreted by the machines that process their data; explainable neuro-symbolic AI to clearly communicate and prove the decisions AI systems make; and decentralised KG management and decentralised infrastructure to provide alternatives to approaches where a central entity controls a whole process, that are prone to abuse of power. We posit the following research questions:

Q1 What are the key requirements for an AI system to produce trustable results?
Q2 How can AI be made accountable for its decision-making?
Q3 How can citizens maintain autonomy as users or subjects of KG-based AI systems?

In order to facilitate exposition, we ground our discussion in a healthcare scenario inspired by the recently proposed regulation on European Health Data Space[5] that aims to ensure that *"natural persons in the EU have increased control in practise over their electronic health data"* and to facilitate access to health data by various stakeholders in order to *"promote better diagnosis, treatment and well-being of natural persons, and lead to better and well-informed policies"*. The proposed healthcare scenario, which is illustrated in Figure 2, is comprised of the following actors and interactions:

Individuals manage their Personal Knowledge Graphs (PKGs) (aligned with the original Semantic Web vision and modern interpretations [7, 48]). They collect knowledge about their medical conditions, symptoms, treatments, reactions to treatments, etc. Individuals get services from KG-based AI applications that utilise their PKGs, e.g., therapy bots or health assistants.

Experts in healthcare also have PKGs where they collect their knowledge about diseases, results of the treatments they have suggested in the past, links to general medical knowledge graphs, etc. Experts may also be assisted by KG-based AI models.

Knowledge sharing communities are spaces where individuals and healthcare experts may share subsets of their PKGs in the context of specific knowledge, e.g., diseases. We call these *community-based perspectives*. Perspectives from different contributors are aggregated into community KGs (e.g., disease-based). AI applications use these KGs for community benefit, e.g., assessing if a treatment that worked for an individual may work on a different one.

Public and private organisations may negotiate access to data and knowledge from communities to train large KG-based AI models to either improve internal processes or power products sold to communities, experts, or individuals, completing the cycle.

The remainder of the paper is structured as follows: Section 2 introduces the necessary background in terms of KG-based AI. Section 3 highlights the importance of trust, accountability, and autonomy when it comes to ensuring that AI benefits society. Section 4 presents several KG based tools and techniques that can be used to facilitate trust, accountability, and self-determination. In Section 5, we propose a research roadmap that includes several challenges and opportunities for KG-based AI that benefits individuals and society. Finally, we conclude and outline important first steps in Section 6.

2   Knowledge Graph-based AI

---

[5]   https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52022PC0197
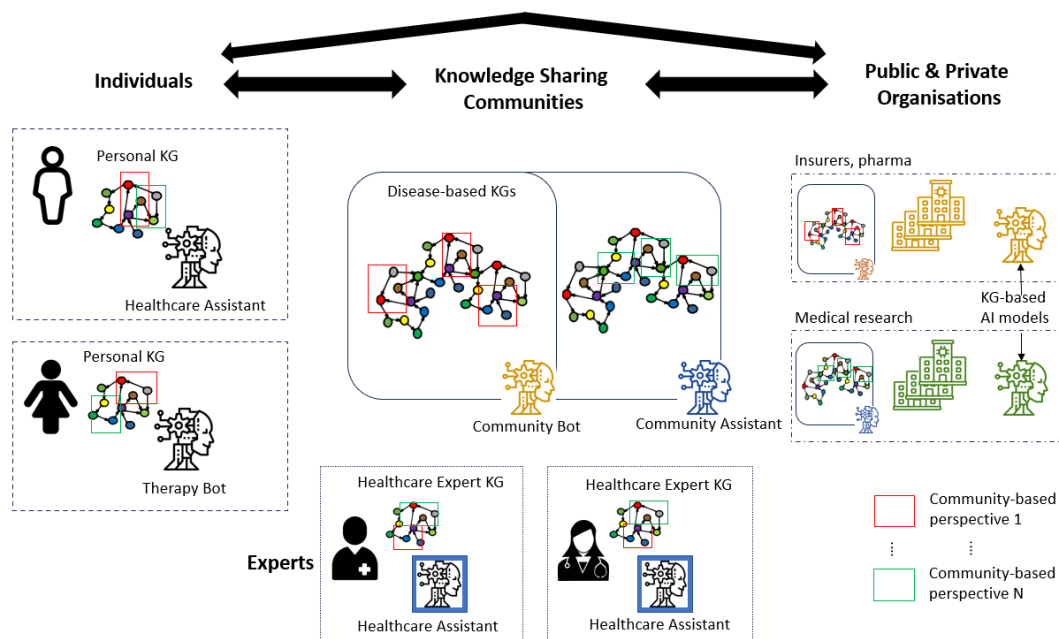
Figure 2 Illustrative Scenario for KG-Based AIs in the healthcare domain. Individuals use AI assistants to make sense of data collected in their PKGs. They may also share perspectives of their PKGs with other individuals and healthcare experts in knowledge-sharing communities that aggregate and curate data to power AI services for the benefit of all members. Public and private organisations can negotiate access to data from communities and individuals to train KG-based AI models, which in turn are used to build services for them.

In his seminal publication, *"Thinking, Fast and Slow"*, Daniel Kahneman [50] presents a comprehensive theory of human intelligence, offering profound insights into the workings of the human mind. This groundbreaking work separates intuition from rationality when approaching problem-solving tasks, defining them as two sets of abilities or *systems*. System 1 operates at an unconscious level, generating responses effortlessly and swiftly. In contrast, System 2 requires conscious attention and concentration, enabling the generation of responses needing complex computations. Kahneman's characterisation of mental cognition aligns with statistical and symbolic learning models that seek to simulate human thinking processes [17]. These systems are known as neuro-symbolic systems [18], and there is a growing interest in emerging hybrid approaches that aim to integrate cognitive capabilities. Specifically, they strive to combine the power of neural networks, such as LLMs, with the interpretability offered by symbolic processing, particularly semantic reasoning over KGs.

2.1 Knowledge Graphs

Google first introduced KGs in 2012 when they enabled 'Knowledge Panels' containing descriptions including pictures for search items. For example, if one types in 'London' to Google Search, the Knowledge Panel displays pictures, the current weather, a map, directions, elevation and related entities (e.g. Paris). The seed for the Google KG was Freebase - a community knowledge base initially launched in 2007 with an add-on RDF service launched at the International Semantic Web Conference in 2008. In 2010, Google bought Metaweb, the company that owned

Freebase and extended the knowledge base into the Google KG[6].

In 2011, Bing, Google and Yahoo! launched Schema.org, a reference website for common data schemas related to web search. The proposal was that website owners would use the published schemas alongside Semantic Web standards such as RDFa and JSON-LD. A number of the schemas, such as Organisation, influence the results returned by Google KG search. Schema.org is an example of a shared vocabulary for semantic representation; the use of such vocabularies or ontologies in KGs, along with the ability to map between equivalent schemas in them, enables integration of heterogeneous data at scale.

Today, KGs are used in a wide range of areas and products outside of search. For example, Netflix, Amazon, and Facebook all use KGs as the foundation for their recommendation engines for television programmes and films, consumer products and posts[7], whereas in the healthcare sector, KGs are used to integrate medical knowledge and support drug discovery.[8]

2.2   Large Language Models

A Large Language Model (LLM) is a specialized machine learning model constructed using a transformer architecture, a category of deep neural networks [121]. LLMs are primarily designed for predicting the next word in a sequence, making them flexible tools for various text processing tasks, such as text generation, summarization, translation, and text completion. Examples of existing LLMs include OpenAI's ChatGPT [95] and Google's PALM [25]. These models have demonstrated high performance in Natural Language Processing (NLP) tasks like code generation, text generation, tool manipulation, and comprehension across diverse domains, often achieving high-quality results in zero-shot and few-shot settings. This success has stimulated advancements in LLM architectures, training techniques, prompt engineering, and question answering [73].

Despite their unquestionable capabilities in emulating human-like conversations, there is an ongoing debate regarding the intelligence exhibited by LLMs, particularly, since their fluency in language does not necessarily imply a cognitive understanding of real-world problems [73]. Additionally, LLMs can only learn knowledge when it appears in the training data and may perform badly when answering questions involving long-tailed facts [32]. Moreover, they may struggle to absorb new knowledge and are not easy to audit [74], suggesting potential risks of discrimination and information hazards.

2.3   Neurosymbolic AI

LLMs– and machine learning models in general– are trained on extensive datasets, resulting in high-quality outcomes whenever applied to specific prediction tasks. However, LLMs– like OpenAI's ChatGPT [95]– lack of causal understanding and may hallucinate in cases which are not statistical in nature (e.g., memories or explanations) [41]. On the other hand, Symbolic AI systems are capable of emulating human-like conscious processes required for causality, logic and counterfactual reasoning, and maintaining long-term memory. As a result, symbolic systems can empower LLMs by modelling human learning and combining knowledge extracted (e.g., from KGs) to formulate prompts that allow for a more fluent communication with users.

Neuro-symbolic AI provides the basis for integrating the discrete approaches implemented by Symbolic AI with high-dimensional vector spaces managed by LLMs. They must decide when and how to combine both systems, e.g., following a principled integration (combining neural and

---

[6] https://en.wikipedia.org/wiki/Schema.org
[7] https://builtin.com/data-science/knowledge-graph
[8] https://www.wisecube.ai/blog/20-real-world-industrial-applications-of-knowledge-graphs/

symbolic while maintaining a clear separation between their roles and representations) or integrated (e.g., a symbolic reasoner integrated into the tuning process of an LLM). Recently, van Bekkum et al. [111] propose 17 fundamental design patterns to model neuro-symbolic systems. These patterns encompass many scenarios where the symbiotic relationship between symbolic reasoning and ML models becomes apparent. Since these combinations may enable symbolic reasoning and enhance contextual knowledge, neuro-symbolic systems may empower explainability and, as a result, also improve transparency by showing how a system works based on the symbolic explanations deduced by the hybrid system.

3 KG-based AI that Benefits Individuals and Society

Considering our vision that KG-based AI can facilitate self-determination, we start by discussing the pertinent role played by trust, accountability, and autonomy when it comes to ensuring that AI benefits society. In each case, we highlight existing challenges and present arguments in favour of KG-based AI system.

3.1 Trust and KG-based AI

One of the primary objectives of the proposed EU AI Act is the *"development of an ecosystem of trust by proposing a legal framework for trustworthy AI"*. The Merriam-Webster dictionary definition of trust includes a *"firm belief in the reliability, truth, or ability of someone or something"* [71]. Questions we address in this paper include understanding how KG-based AI systems can demonstrate reliability, truth, and ability through mechanisms, which add transparency to all elements involved in KG-reasoning. These include: comprehensive provenance tracking of data sources and data elements used for any output; understanding repeatability for all KG-based AI reasoning (e.g., if datasets are altered or disappear altogether, or if other reasoning methods, such as LLMs, are involved); and alleviation mechanisms when KG-based AI system responses are untruthful.

The proliferation of misinformation on the internet has risen significantly in recent years, coinciding with the advancements in generative AI technologies. As AI becomes more sophisticated, it has inadvertently provided tools and techniques for the creation and dissemination of false information, leading to widespread confusion and societal harm [26, 122]. For instance, AI-generated deepfake videos have become a concerning source of misinformation. Deepfakes use AI algorithms to manipulate and superimpose faces onto existing videos, making it difficult to discern real from fabricated content [114]. This technology has been used to create fake videos of public figures saying or doing things they never actually did, leading to potential defamation and manipulation of public opinion. AI-powered chatbots and automated accounts on social media platforms have been employed to spread false information and manipulate public sentiment. These bots can mimic human-like conversations and flood social media platforms with fake news, propaganda, and divisive narratives, influencing public opinion and sowing discord, and have even contributed to misinformation in medical literature [66]. AI-powered recommendation algorithms used by platforms like social media and video-sharing websites can inadvertently contribute to the spread of misinformation. These algorithms aim to maximise user engagement by suggesting content based on user preferences and behaviour. They can create filter bubbles, reinforcing users' existing beliefs and exposing them to a limited range of perspectives, potentially amplifying false information and preventing users from accessing accurate and diverse sources of information [86].

Amidst these challenges, KG technologies have emerged as a potential solution to curb misinformation and enhance trust. Leveraging the power of crowd-supplied and verified knowledge sources, such as Wikidata [112], KGs enable comprehensive fact-checking capabilities. By integ-

rating diverse and reliable information from various trusted sources, these graphs can potentially identify and flag misleading or inaccurate content more effectively. By utilising the collective intelligence of a crowd, KG technologies empower users to contribute to the verification process, enhancing the accuracy and credibility of the information presented. Through collaborative efforts and the utilisation of KG technologies, it is possible to combat the rising tide of misinformation, safeguarding the integrity of online information and fostering a more informed digital society. Coupled with distributed ledgers, it has been proposed that KG-based AI can combat misinformation on the web [97]. There is already a growing body of work in this space, which shows some promise. For example, Mayank et al. [69] and Koloski et al. [58] describe systems that leverage KGs to detect fake news; Kou et al. [60] and and Shang et al. [98] describe how crowd-sourced KGs can be used to mitigate COVID-19 misinformation; Kazenoff et al. [51] use semantic graph analysis to detect cryptocurrency scams propogating in social media.

3.2   Accountability and KG-based AI

According to the proposed EU AI Act, when it comes to high-risk AI, *"accuracy, reliability and transparency is particularly important to avoid adverse impacts, retain public trust and ensure accountability and effective redress"*. Accountability in a KG-based AI context assumes that data scientists, computer scientists, and software engineers will follow best practices and ensure compliance with relevant legislation. In the purely symbolic world, such properties can be achieved via consistency and compliance checking based on formal requirements specified in policy languages such as LegalRuleML [4] and ODRL [47]. When it comes to the sub-symbolic world, these principles are particularly challenging, as ML algorithms are often opaque and could potentially infer confidential information during the training process. In recent years, various Explainable AI (XAI) techniques have been used to build or applied to the output of models such that they can be interpreted and understood by various stakeholders [57]. In the context of KG-based AI this will require the intersection between two strains of explainability: the explanation of why a statement is in the KG that supports the AI, and the explanation of how the model used the statements from the KG to reach a particular decision. KGs can also be used to support the modelling, capturing, and auditing of records useful for accountability throughout the system life cycle [76]

When it comes to AI and accountability technical research should go hand in hand with the interdisciplinary research conducted in communities like FaccT[9]. A recent paper [27] revisited the four barriers of accountability that were developed in the 1990s for accountability of computerised systems in the light of the rise of AI, finding that they are even more important than before. The main barrier is the problem of *many hands* - the large amount of actors involved in the construction of an AI service creates difficulties in the assignment of responsibilities in case of harm. Advancing efficient provenance collection, and verifiability will be the key technical intervention to overcome this barrier. Fields such as data science require strong guarantees for provenance to build context-aware KGs [96]. Similar to explainability, we consider two different approaches zhat need to be combined: the provenance of statements in the KG and the provenance of the pipeline that was followed to construct the ML model.

---

[9] `https://facctconference.org/index.html`

3.3 Autonomy and KG-based AI

Alongside accountability and trust, the third pillar needed to support self-determination is *autonomy*, defined from a self-determination theory[10] perspective as *"the belief that one can choose their own behaviors and actions"*. In the current context, we take this to mean that individuals should be able to make their own decisions about their uses of KG-based AI and about its uses of their data (and have their wishes respected). Assuming that AI systems can be made to be trustable and accountable, how can we best support autonomy in this way? That is to say, if we can know that an AI will behave in a desired and known way, and that its decisions and processes are transparent and traceable, how can we express and enable control over what it does in regard to an individual? A number of approaches have emerged in recent years which facilitate individuals' data sovereignty and how they represent and express their identity online.

The concept of a PKG– introduced in our illustrative scenario– is one means of facilitating autonomy; Solid pods [93, 68] are secure decentralised data stores accessible through standard semantic interfaces for applications that generate and consume linked data. Currently, the default model on the Web is for service providers to host and control access to user data by means of a user account. This denies autonomy to the individuals concerned since all access is mediated via applications and interfaces designed and controlled by service providers. The PKG model is that personal data is independent of any application; PKGs are the primary source of data under the control of individuals, and they mediate service access via standard interfaces. On top of shifting control away from service providers, this approach makes it technically simpler to implement data usage policies, as they can be stored with the data and evaluated at the PKG level.

One prominent way of achieving the second goal is through the notion of Self-Sovereign Identity (SSI) [29]. Traditional digital identity (e.g., as in OpenID Authentication [42]) has been modelled in terms of Identity Providers (IdPs). An individual and an IdP establish a relationship, and the IdP generates a digital identity for them. If the individual wants to authenticate with a third party, the IdP confirms the relationship to them and then asserts that identity to the relying party. Crucially, sovereignty over that identity and decisions about who can see it, the data associated with it, or whether it continues to exist are taken by the IdP. With SSI, an individual generates their *own* digital identity (e.g., a cryptographic key pair), makes their own identity assertions, and therefore has full control over that identity, with correlations between two identities (digital or physical) relying explicitly on attestation by others, and trust relationships with them[11]. The autonomy enabled by SSI makes *selective disclosure* possible, meaning that what identity information gets shared with whom can be made contextually and on a case-by-base basis - much like presenting different aspects of ones personal identity in daily life (e.g., work and home personas).

Considerations of identity pervade any technical considerations for safeguarding self-determination. It seems uncontroversial that there will be scenarios in which an individual's identity is relevant to what they wish to do with a KG-based AI, whether in training, KG contents, or inference, and indeed, even where anonymity is desired, identity must be considered in order to avoid revealing it. Identity is also fundamental to the concept of trust; trust in a person, organisation, system, AI model, KG, etc., is useful only in so far as it is possible to identify relevant entities as needed, and accountability cannot be tracked or apportioned without it. We consider autonomy in terms of the identity, data, and sovereignty afforded to an individual or organisation in terms of what they or others communicate to a KG-based AI ecosystem or elements

---

[10] https://en.wikipedia.org/wiki/Self-determination_theory

[11] As it ultimately does in traditional digital identity, where trust in a small number of well-known IdPs serves as a simplified proxy for more detailed or fine-grained considerations of trust networks.

thereof, what they or others receive from those, and what happens to those (including respect of choices) as data is processed in the ecosystem, with each of these evaluated through the lenses of selective disclosure, relevant identities, and utility.

4   A KG Toolbox for Trust, Accountability, and Autonomy

In order to ground our pillars, we motivate and introduce relevant literature and highlight open research challenges and opportunities concerning our foundational topics: machine-readable norms and policies; decentralised infrastructure; decentralised KG management; and explainable neuro-symbolic AI, each of which plays a pivotal role in facilitating trust, accountability, and autonomy in KG-based AI.

4.1   Machine-readable Norms and Policies

When it comes to KG-based AI, norms and policies could potentially be used to inform data processing based on legal requirements, social norms, privacy preferences, and licensing. Legal documents are designed in natural language for human consumption, thus in order to enable machines and automated agents to evaluate and enforce the agreements embodied in documents, we need to translate them to formats they can read and process efficiently.

Norm and Policy Encoding. Languages to express policies, including but not limited to data access, can be categorised as either general or specific. In the former, the syntax caters to a diverse range of functional requirements (e.g., access control, query answering, service discovery, negotiation), whereas the latter focuses on just one functional requirement. In the early days of the Semantic Web, research into general policy languages that leverage semantic technologies (e.g., KAoS [110], Rei [49], AIR [52], and Protune [16]) was an active area of research. However, despite the huge potential offered by these general purpose languages to date none of them achieved mainstream adoption [56]. More recently, researchers have proposed ontologies that can be used to represent licenses, privacy preferences, and regulatory obligations [54]. When it comes to the legal domain specifically, Semantic Web researchers have proposed cross-domain ontologies that can be used to encode legal text in a machine-readable format using LegalRuleML[12] and adaptations thereof (e.g., [3, 81]). Others focus on facilitating legal document indexing and search using the European Law Identifier (ELI)[13] and the European Case Law Identifier (ECLI)[14] (e.g., [79, 23]), or bridging the gap between the EU and member state legal terminology (e.g., [1, 15]). Besides these cross-domain activities, there have also been various domain-specific initiatives. For instance, the ELI ontology has to be extended to facilitate the encoding of the text of the General Data Protection Regulation (GDPR)[15] (e.g., [85]). While others have focused specifically on modelling privacy policies (e.g., [80, 84]). The Open Digital Rights Language (ODRL)[16], which is a W3C recommendation, has gained a lot of traction in recent years in terms of intellectual property rights management (e.g., [43, 75]). Additionally, the ODRL model and vocabularies have been extended in order to model contracts [40], personal data processing consent [33], and data protection regulatory requirements [28]. There has also been some work on automatically extracting rights

---

[12] https://docs.oasis-open.org/legalruleml/legalruleml-core-spec/v1.0/os/legalruleml-core-spec-v1.0-os.html
[13] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52012XG1026(01)
[14] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52011XG0429(01)
[15] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679&qid=1681238509224
[16] https://www.w3.org/TR/odrl-model/

and conditions from textual documents (e.g., [22, 21]) or extracting important information from legal cases (e.g., [117, 78]). Although many of the proposed approaches are based on existing standards, there is a lot of overhead involved for systems that need to consider different types of policies that are encoded using different languages. General-purpose policy languages are particularly attractive in such scenarios as they lessen the administrative burden. However, considering the potential complexity of such a language, there is a need for policy profiles with well-defined semantics and complexity classes.

Policy Enforcement and Governance. From a policy governance perspective, LegalRuleML researchers have proposed automated compliance approaches based on auditing (e.g., [30, 84]) and business processes (e.g., [82, 10]). While [38] shows how LegalRuleML together with semantic technologies, is used for business process regulatory compliance checking based on a rule-based logic combining defeasible and deontic logic. One of the advantages of description logic-based approaches, when it comes to consistency and compliance checking, is that they can leverage generic reasoners, such as Pellet[17] (e.g., [34]). Although there are presently no ODRL-specific reasoning engines, researchers have demonstrated how ODRL can be translated into rules that can be processed by Answer Set Programming (ASP) [9] solvers such as Clingo [36] (e.g., [43, 28]). Additionally, there have been several custom applications that are designed to support ODRL enforcement or compliance checking, such as a license-based search engine [75]; generalised contract schema and role-based access control enforcement [40]; and access request matching and authorisation [33]. Despite existing efforts, challenges arise when it comes to ensuring that AI and processing algorithms adhere to the policies. This could potentially be achieved either before or during processing using trusted execution environments [11] or after execution by detecting data misuse via automated compliance checking using system logs [55]. The combination of ex-ante and ex-post compliance checking is particularly appealing for supporting risk-based conformance checking such as that envisaged in the proposed EU AI Act. Nevertheless, the practicality, performance and scalability or these proposals remain to be determined. In order to further support self-determination, data owners and processors should be able to engage in on-demand negotiation over policies, assisted by technology that ensures a safe and fair space and helps assessing the compliance of negotiated terms with existing regulation. Negotiation between automated agents has been a topic of interest since the early 2000s but in the context of self-determination we must pay attention to the right balance between artificial representation and human involvement [5, 6].

Grounding based on our Illustrative Scenario. Figure 3 illustrates how machine-readable policies and norms can be used to support self-determination. Considering our illustrated scenario individuals may want to establish policies to precisely define the subset of their PKGs to be shared with communities and what forwarding they allow. For example, *share with the diabetes community my blood in sugar values measured by my connected device and the output of my AI healthcare assistant, or only share and forward anonymised aggregates to medical research institutions, or contact me for negotiation if the pharma company is interested in using my data for clinical studies.* Communities may do the same, *e.g,* requiring specific data to be shared to join the community, but also requiring agreements in order to ensure that participants will abide by social and behavioural norms needed for self-regulation. Public and private organisations may need to adhere not only to privacy preferences and licenses but also to various general regulations, e.g., the GDPR, the proposed AI Act in the EU or the Health Insurance Portability and Accountability Act (HIPAA)[18] in the US, as well as domain-specific regulations (e.g., advanced therapy medicinal

---

[17] https://github.com/stardog-union/pellet
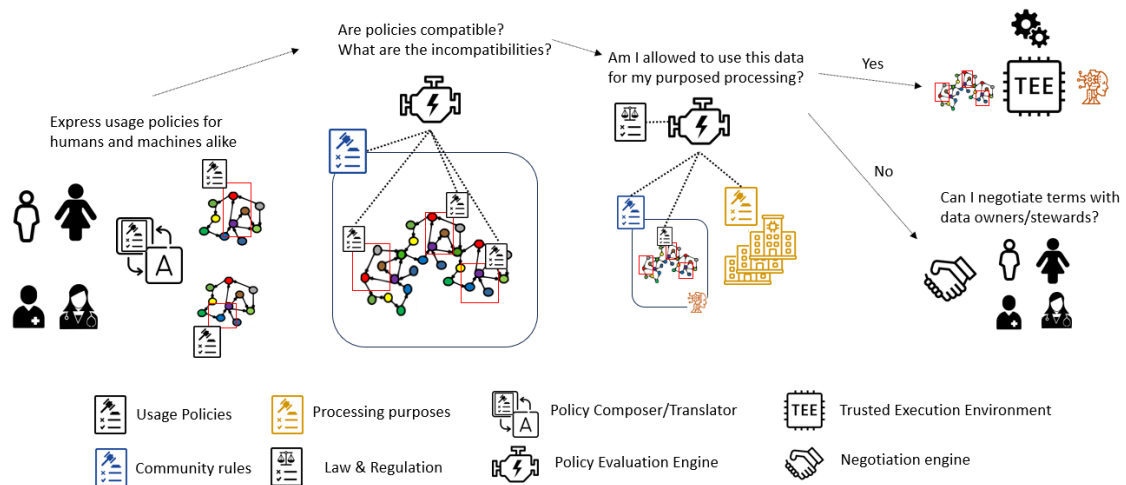[18] https://www.hhs.gov/hipaa/index.html

Figure 3 Machine-readable norms and policies to support self-determination. A Policy Composer/Translator assists individuals in writing data usage policies, communities in defining their rules, and organisations in declaring their data processing purposes in both human- and machine-readable formats. Policy Evaluation Engines assess the acceptability of perspectives in a community by evaluating policies and rules. Engines assess organizations' data usage compliance with regulations. If permitted, processing can occur in a Trusted Execution Environment ensuring compliance. If not allowed, a Negotiation engine may be utilised to seek agreement with data owners/stewards under relevant regulations.

products[19] and rare diseases[20]).

4.2 Decentralised Infrastructure

Over the last 15-20 years, a number of communities have come to accept that centralised computing systems, despite many benefits, can lead to issues such as the over-centralisation of power, the risk of single points of failure, potential abuse of personal data and creation of data silos which can inhibit innovation. A boon from this realisation is that we now have a number of technologies, standards, and approaches to decentralisation which offer benefits in terms of scalability, diversity, and privacy, as well as individually-centred flexibility and control, and is an appealing basis for maintaining and increasing trust, accountability, and autonomy with KG-based AI.

Personal Knowledge Graphs. The concept of a Personal Knowledge Graph (PKG), is that an individual can keep their personal or private data in a space belonging to them, rather than with siloed centralised service providers with limited access and control [8]. A Solid pod[21] is an example of a PKG platform, and the key to the vision of Solid is that there should be standard interfaces and authorisation models to grant or deny access to the contents of a PKG at a granular level. This is argued in particular[22] to enable a highly decentralised architecture for Web applications. Rather than a provider aggregating data from all users into a single location controlled by the provider and application code accessing such data there, instead, an individual permits (or does not permit) Web applications of their choice to access whatever subsets of their data they decide

---

[19] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32007R1394
[20] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32009H0703%2802%29
[21] https://solidproject.org/
[22] https://ruben.verborgh.org/blog/2017/12/20/paradigm-shifts-for-the-decentralized-web/

from their PKG. As well as autonomy, this enables greater accountability since access to the PKG can be filtered via personal machine-readable policies at source, and activities can be tracked directly (e.g., [31]). Although PKGs offer great potential, they also come with challenges in terms of performance and scalability as applications will need to interact with multiple distributed data sources as opposed to a single backend server. These challenges, however, may also simultaneously be opportunities for scalability trade-offs, querying over multiple low-powered data sources rather than a high-powered central one.

Distributed Ledger Technology. Distributed Ledger Technology (DLT) [106] promotes trust and empowerment through the replication of data across contributing nodes, which are geographically distributed across many sites, and the use of consensus algorithms which enable collective fair decision-making with no central control. Blockchains are a type of distributed ledger where an ever-growing list of records or blocks are tied together with cryptographic hashes, often, although not necessarily, associated with a securely exchangeable token system, or 'cryptocurrency'. This technology rose to prominence following the release of Bitcoin [77] in 2008 - a blockchain-based currency that has now been adopted by El Salvador as their legal tender. Ethereum [116], a blockchain platform released in 2015, contains the notion of a 'Smart Contract' [20] (originally coined in the 1990s by Nick Szabo [107]), which is a collection of code that executes in a fully decentralised way. Smart Contracts have been used to implement a range of decentralised applications, including Decentralised Autonomous Organisations (DAOs) [65], which are organisations where decisions are made through blockchain consensus mechanisms. The best-known example of a DAO was 'The DAO' which at one point was worth more than $70M; they have been applied to a number of different activities, including scholarly publishing [44]. Despite the fact that immutability and transparency guarantees offered by DLT are very attractive, when dealing with personal data both the ledgers and the smart contracts themselves will need to be protected against unauthorised access and usage, and designed such that personal data itself is neither stored in, or derivable from, immutable DLT records. Smart contracts may also introduce scalability issues: the default Ethereum model involves every contributing node executing every run of a smart contract, and thus has inherent scale limitations. Relaxing this model may, however, affect trust.

Self Sovereign Identity. In the Web space, Self-Sovereign Identity (SSI) is being developed through a combination of Decentralised Identifiers (DIDs) [103] and Verifiable Credentials (VCs) [104], W3C standards for identity and verifiable attestation claims, respectively. DLT is one of the ways in which DIDs can be grounded, although, by design, the DID standard is open in terms of method. A DID is a URL (`did:<method>:<...>`) which can be resolved in a method-specific manner (e.g., HTTP(S) dereferencing, reading from a smart contract, etc.) to obtain a DID document, a Linked Data set containing information about digital identity in a standard form - for example, how to verify it (e.g., a public key), methods for communicating with the entity controlling it, and so on. DIDs enable SSI; the creation and use of DIDs are open and decentralised, and by using different DIDs with different audiences, individuals can minimise how easily their information can be tracked or correlated across services and can contextually and selectively disclose personal information as desired. This grants individuals significantly greater autonomy than current practices. There is a potential trade-off with trust and accountability of an individual when it comes to information that others need to rely on, which is that effective anonymity of a unique DID can be used to misrepresent oneself (e.g., fake a qualification or entitlement) or pretend to be someone else. VCs are a proposed solution to this. The VC data model is for sharing data alongside information that a recipient can use to verify its integrity or origin, such as a digital signature or DLT record. If a DID is presented to a service that is restricted

to legal adults, for example, the DID owner may also present a VC issued by a government body confirming their adulthood; methods for selective disclosure supported by both DID and VC standards allow this to be done verifiably without requiring disclosure of real-world identity. These technologies are relatively new in comparison with standard digital identity models, and, while intended and designed to address issues in those models, they may also introduce new difficulties or enable different vulnerabilities to, e.g., identity fraud, than current standards.

Federated Learning. In the context of data-driven AI and decentralised infrastructure, there are also techniques for decentralised machine learning. Federated Learning (FL) [119] is the idea that, rather than aggregating training data in one location controlled by a model developer (thereby compromising subject privacy), data holders can run learning algorithms to generate model weights for their own data locally and privately, and then send only the weights to the developer to be incorporated into the larger model. An example might be a smartphone text prediction personalisation algorithm, where a user's own writing is used to generate predictive weights on device, and periodically selections of these can be aggregated to improve general text prediction models. Refinements of FL approaches include sending not the actual learned model weights, but a set of weights with statistically similar properties [113], to further reduce the risk of privacy breaches without affecting model performance. A related approach takes this concept even further, with the idea of embeddings in a larger model, e.g., 'Textual Inversion' [35] to personalise large generative image diffusion models. The intuition here is that if someone wants certain personalised specific types of output from a generative AI, then, if a model is sufficiently large, there is a good chance that the desired concept already exists within it. More recently, the idea of federating for preserving privacy has been applied specifically to deep learning, in particular in the context of Internet of Things. [120] proposes an architecture with a control layer including a distributed ledger, while [118] propose advanced cryptographic mechanisms to reduce the risk of privacy leaks, following more general approaches that apply either differential privacy, homomorphic encryption or secure multi-party computation. Federation also has the positive side effect of potentially speeding up model training when the privacy constraints allow for a helpful distribution of the process [12]. However, when opening the process to multiple parties, there are a number of attack vectors that do not exist in a centralised approach for which we need protection, and pay a communication and computation overhead [45].

Grounding based on our Illustrative Scenario. A decentralised infrastructure supporting self-determination for our illustrative scenario is depicted in Figure 4. Health data is highly sensitive and private, and individuals may want or need to interact with multiple services where it is relevant, including KG-based AI systems. It thus makes sense to create a personal health knowledge graph (PKG) to be a comprehensive and interconnected representation of an individual's health information, including their medical history, lifestyle choices, genetic data, and real-time health monitoring data from IoT devices. Data from various sources, such as wearable devices, mobile applications, electronic health records, and even genomic sequencing, can be linked together to form a holistic view of an individual's health in such a personal health knowledge graph. An early example of a PKG was in [108], where medical, lifestyle, and IoT health monitoring data in a PKG was integrated into a (patient-focused) decision support system built around a public medically-curated KG representing cardiovascular risk factors, giving individuals the autonomy to gain deeper insights into their own health patterns and risks, identify

---

[23] The full picture would have knowledge exchange between multiple parties; to avoid an unreadable cluttered figure, this is left implied by the background network.
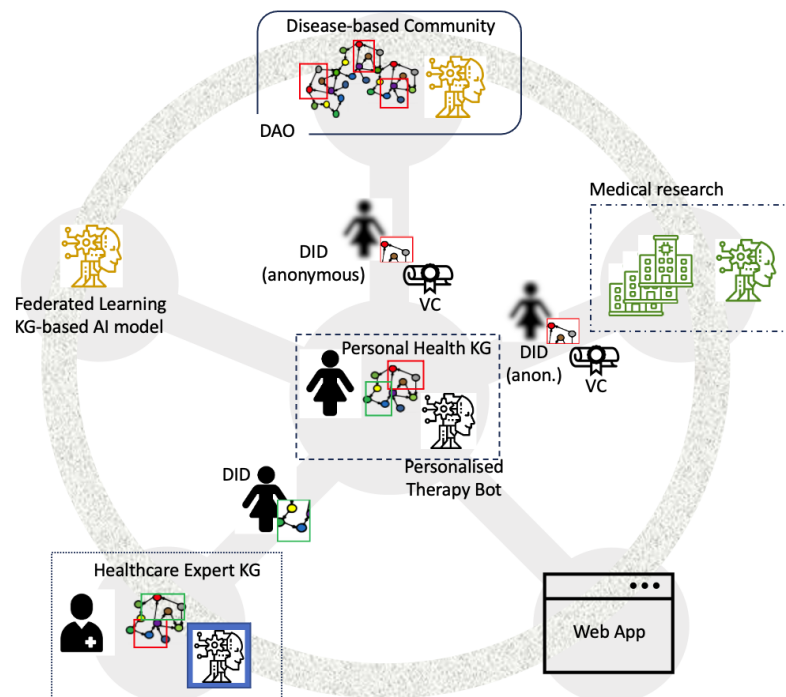
Figure 4 Decentralised Infrastructure supporting self-determination, shown from the perspective of one individual with a PKG[23]. According to individual wishes, portions of the PKG can be shared either directly with a healthcare provider, with web applications for health, or indirectly with peer or research communities. Identity is via DIDs (anonymous in the latter cases), with VCs used for trustable selective disclosure. KG-based AI models can be trained and personalised in federated and private ways on knowledge from diverse sources.

correlations, and make more informed decisions. More recently, BlockIoT [99, 100] aims to integrate health data seamlessly in a decentralised PKG using blockchain and KG technologies, addressing this trust aspect and using PKG-driven smart contracts to trigger the personalised recommendations for lifestyle modifications, medication adjustments, or even timely interventions by the healthcare providers. Furthermore, the PKG can serve as a powerful tool for healthcare beyond the individual. Communities of patients, providers, researchers, etc., or combinations thereof, can share knowledge about various aspects of, e.g., particular conditions, whether that is clinical evidence and best practice, peer advice and support on living with a condition, or data on novel or rare symptoms and side effects, with this knowledge used for support, care, or medical research across populations. De-identified and aggregated data from multiple individual KGs can be collected in community KGs, with trust securely established using DIDs and VCs, and accessed by community, practitioner, researcher, and service provider stakeholders, allowing for decentralised large-scale analysis and identification of broader health trends from multiple perspectives and intersecting factors. This can lead to advancements in disease prevention, treatment protocols, and the development of personalised medicine in a collaborative manner [101]. KG-based AI systems can be both trained and used across this ecosystem, with FL being applied to train larger models (e.g., the organisation models in Figure 2) and personalised embeddings used by individuals to get the best experience from their therapy bots and healthcare assistants while maintaining privacy and autonomy.

4.3   Decentralised KG Management

As the amount of data and knowledge grows exponentially, managing and harnessing this vast information becomes increasingly complex. Traditional centralised approaches to KG management face challenges in terms of scalability, privacy, and control over data, and to address these issues, decentralised KG management emerges as a promising solution. This section explores the key aspects and open challenges in decentralised KG management to enable trust, accountability, and self-determination for individuals in a rapidly evolving AI ecosystem.

Decentralised KG Access and Management. Efficient query processing infrastructures are fundamental for traversing decentralised KGs. There has been notable efforts such as Fedbench [94] in the past. However, these infrastructures should be capable of executing queries against the available KGs while respecting privacy and adhering to norms and policies. With the increasing emphasis on privacy protection with regulations such as GDPR, it is crucial to develop mechanisms that allow users to access and extract knowledge from KGs without compromising sensitive information or violating privacy regulations. Several research directions are worth considering to address the open challenges in decentralised KG management. Firstly, developing the formalisms to describe KG management semantically can provide a common ground for understanding and interoperability across different decentralised KG systems. Such formalisms can enable standardised representations of KGs in the form of ontologies and facilitate seamless integration and collaboration among diverse knowledge sources. Architectures supporting new protocols and standards specific to decentralised KGs are essential for establishing interoperability and seamless communication between knowledge sources and systems. By defining and adopting common protocols and standards, decentralised KGs can collaborate more effectively, share insights, and facilitate cross-domain knowledge discovery.

Note that if we add LLMs to the picture, their current training and execution processes are currently centralised. Decentralised KG management is useful to provide transparency in data used for their training. For approaches involving the interaction between LLM and KGs, the transparency of the LLM itself still depends on the owner.

Provenance and Explanations. Furthermore, explainable methods for data integration and curation, as well as KG validation and distribution, such as the *Explanation Ontology* for user-centric AI, are necessary to ensure the reliability and accuracy of decentralised KGs [24]. By providing transparent and interpretable approaches, users can have better insights into knowledge integration and validation, enhancing trust and accountability of the knowledge contained in the KG and the insights derived. This is especially critical because, in decentralised KGs, data may come from various sources and be represented in different ways. The standardised framework provided in the Explanation Ontology for representing domain-specific explanations of KG entities and relationships helps users and applications understand the meaning and context of the data in the KG. Provenance and traceability also play a vital role in decentralised KG management. Establishing mechanisms to track and validate the origin, history, and lineage of knowledge within KGs is crucial for accountability and the ability to trace back the sources and transformations that contribute to the resulting knowledge. The W3C Provenance Data Management standards [72] provides the basis for encoding provenance attributes in KGs, and subsequent nanopublications specification [39] has gained a lot of traction in the biomedical domains. While these solutions exist, there needs to be a cohesive framework that ties together explanation provenance data management in a decentralized KG context, ensures that users can trace the origins, transformations, and sources of the data, which is crucial for trust, accountability, and data quality assurance. The W3C provenance data management suite of recommendations provides normative interoperable

guidance on recording information about data sources, contributors, and how data is collected or transformed, making integrating heterogeneous data into a coherent KG easier. When data quality issues arise, users can trace back to the source of the problem and take corrective actions, ensuring the KG remains accurate and reliable. The W3C recommendations for decentralized provenance management provide a mechanism for attributing data to its sources or contributors. This attribution is essential for accountability, especially when multiple parties contribute to a KG.

Blockchain Technologies and Tokenomics. In recent years, the integration of blockchain technologies and tokenomics has gained attention in the context of decentralised KG management. Projects such as OriginTrail[24] have contributed to the development of ownable DKGs, which leverage blockchain's inherent properties to enhance trust, provenance, and accountability. By utilising blockchain, KG management systems can ensure the integrity and traceability of data and metadata across various nodes in the network. The OriginTrail protocol aims to create a trustless environment where data providers, consumers, and verifiers can interact and validate the authenticity and reliability of data stored within the knowledge graph. Their protocol issues tokens as incentives for data contributors, validators, and curators within the KG ecosystem. The integration of blockchain technologies and tokenomics in decentralised KG management addresses several critical aspects. Firstly, blockchain's immutability and transparency enable the traceability and provenance of data and metadata, ensuring accountability throughout the KG management pipeline. Secondly, the decentralised nature of blockchain mitigates single points of failure and promotes the distribution of knowledge and decision-making power among participants. This decentralised approach aligns with the principles of self-determination, empowering individuals to have control over their data and knowledge. By rewarding contributors, validators, and curators with tokens, these systems encourage continuous improvement, data quality assurance, and community engagement. Token-based economies can drive the development of sustainable KG management pipelines, enabling the growth and evolution of DKGs over time. However, the tokenomics have to be carefully designed and monitored to avoid the possibility contributors have a motivation (possibly extrinsic) to misbehave. There is also the risk that a sudden churn in blockchain participants impacts performance and availability. There is also the question of the performance of the consensus algorithm of the Blockchain itself.

Grounding based on our Illustrative Scenario. An approach to decentralised knowledge graph management in the context of healthcare, where users retain control over their personal information while benefiting from enhanced privacy measures and seamless collaboration in a community, is illustrated in Figure 5. At the heart of this framework lies the concept of PKGs, such as Solid, which empower individuals to securely store and manage their personal health data. Central to the architecture are specific components aimed at safeguarding user privacy and ensuring data transparency. The process begins with knowledge sanitisation, which anonymises sensitive information and filters the data according to the user's preferences and data policies. These policies encompass not only globally recognised regulations like GDPR and HIPAA but also individual data policies, enabling users to set granular restrictions on how their data is used, such as opting out of genetic data usage for medical research. To ensure interoperability and standardisation, the creation of knowledge graphs leverages community-defined ontologies and vocabularies. These shared frameworks facilitate seamless integration and alignment of personal knowledge graphs within the broader ecosystem, promoting data exchange and collaboration.
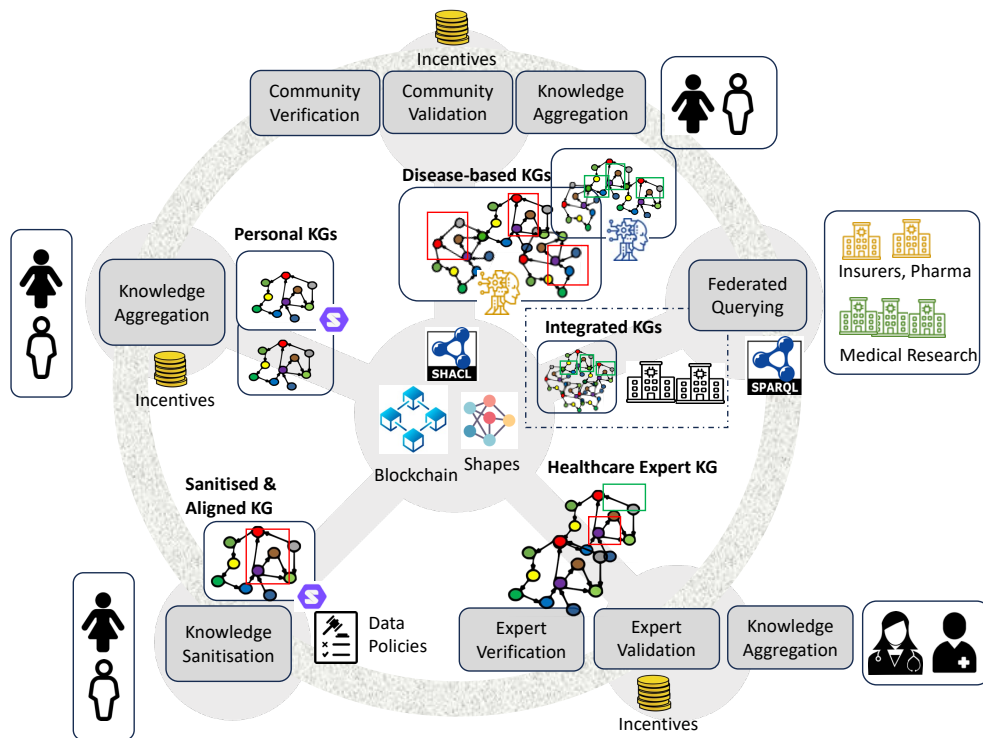
---

[24] https://origintrail.io

Figure 5 Decentralised KG Management Process in Healthcare. Emphasising user empowerment, privacy, and seamless collaboration, users maintain control over their personal health data through personal data stores like Solid, and community and healthcare experts enhance different facets of the KGs in the ecosystem. decentralised KG management involves anonymisation, filtering based on data policies (including GDPR and HIPAA), and alignment with community-defined ontologies. Incentives, driven by blockchain technology, encourage user participation in aggregating KGs and incentivize healthcare professionals for verification, validation, and aggregation activities. SHACL shapes ensure KG validation and federated querying mechanisms enable access to the KGs to stakeholders, e.g., insurers, pharma, and medical research organisations. Integrated KGs are iteratively generated; they comprise a federation of KGs that may be autonomous, distributed, and heterogeneous. A federation query engine enables the traversal of these integrated and connected KGs to provide useful insights to the stakeholders involved.

Users are incentivised to aggregate their knowledge graphs, contributing to the construction of community-based knowledge graphs focused on specific diseases. Through community-based verification, validation, and knowledge aggregation processes, these disease-based knowledge graphs are created, providing valuable insights and fostering collaborative efforts among healthcare professionals, researchers, and the wider community. Blockchain-based incentives drive user participation, rewarding both community users and healthcare experts for their verification, validation, and aggregation activities. The utilisation of an immutable ledger and verifiable credentials ensures the integrity and trustworthiness of the verification process. The validation process, powered by RDF SHACL and Shape descriptions, further enhances data quality and consistency, instilling confidence in the aggregated knowledge. The integrated knowledge graphs, encompassing personal, community-based, and healthcare expert knowledge, can be queried using federated querying mechanisms powered by SPARQL. This allows various institutions, including insurers, pharmaceutical companies, and medical research organisations, to access and leverage the rich insights stored within the knowledge graphs, enabling evidence-based decision-making and advancing medical research and healthcare practices. By combining decentralised knowledge

graph management, user-centric privacy controls, and collaborative data sharing, this innovative framework represents a significant step forward in transforming decentralised KG management, fostering a secure, privacy-enhanced environment that empowers users, facilitates collaboration, and drives advancements in domains such as medical knowledge and patient care.

4.4 Explainable neuro-symbolic AI

Neuro-symbolic systems go beyond generating explanations solely based on the trained model or the individual results derived from applying the model to specific data. They can produce symbolic explanations capturing the essence of an AI model itself. These explanations can be classified as either *instance-level* explanations generated for each specific result of the model, or *model-level* explanations of the structure of a learned model. Previous work on the role of KGs in AI has focused on explainability. [62] frames explainability as a dimension of *trustable* AI and presents challenges, existing approaches, limitations and opportunities for KGs to bring explainable AI to the right level of semantics and interpretability. [109] and [90] conducted independent systematic reviews of existing explainable AI systems to characterise KGs' impact. These results put into perspective the role of KGs in providing symbolic reasoning and learning capabilities with the potential to be precise– as shown by Akrami et al. [2]– in addition to being explainable.

Reasoning and AI. Despite the unquestionable reasoning features of symbolic systems and the studies reporting limitations of LLMs in human-like tasks (e.g., explanations, memories, and reasoning over factual statements) [41], and there is an ongoing debate about LLM's reasoning their causal inference capabilities [53]. Although LLMs excel at certain reasoning tasks, they do poorly in others, raising the question if they genuinely engage in causal reasoning or merely function as unreliable mimics, generating memorized responses (e.g., [46]). Methods to reason can be roughly divided into methods using only the LLM itself (e.g. with prompt-engineering), and methods combining the LLM with an external reasoner and/or external source of knowledge (e.g. a Knowledge Graph) [88]. Our vision posits that external help will always be needed, especially for concrete use cases. There are discussions about the need for knowledge graphs in the era of LLMs. Sun et al. [105] and Dong [32] report on an empirical assessment of ChatGPT [95] with respect to DBpedia, illustrating the need of symbolic systems that *over-fit for the truth* whenever factual statements are collected from KGs. In addition, symbolic approaches can support sanity checking and be easily auditable and traceable. These features position the combination of both approaches in neuro-symbolic AI as a feasible option to provide KG-based AI. The neuro-symbolic AI delivers the basis to integrate the discrete methods implemented by symbolic AI with high-dimensional vector spaces managed by LLMs. They must decide when and how to combine both systems, e.g., following a principled integration (combining neural and symbolic while maintaining a clear separation between their roles and representations) or integrated (e.g., a symbolic reasoner integrated into the tuning process of an LLM).

Trust and AI. Trust in AI systems stems from various factors, including transparency, reproducibility, predictability, and explainability. Neuro-symbolic systems play a vital role in enhancing trustworthiness by enabling communication between modules and facilitating tracing. Modularity enables the specification, verification, and validation of each component and its interactions. As a result, a system's behaviour can be traced and validated. Specifically, within the domain of KG-based AI for self-determination, the seamless integration of KGs and symbolic semantic reasoning offers a comprehensive and unified perspective on curated knowledge. This integration holds immense value in addressing critical tasks such as validating, refuting, and explaining incorrect, biased, or misleading information that may potentially be generated by LLMs. By combining
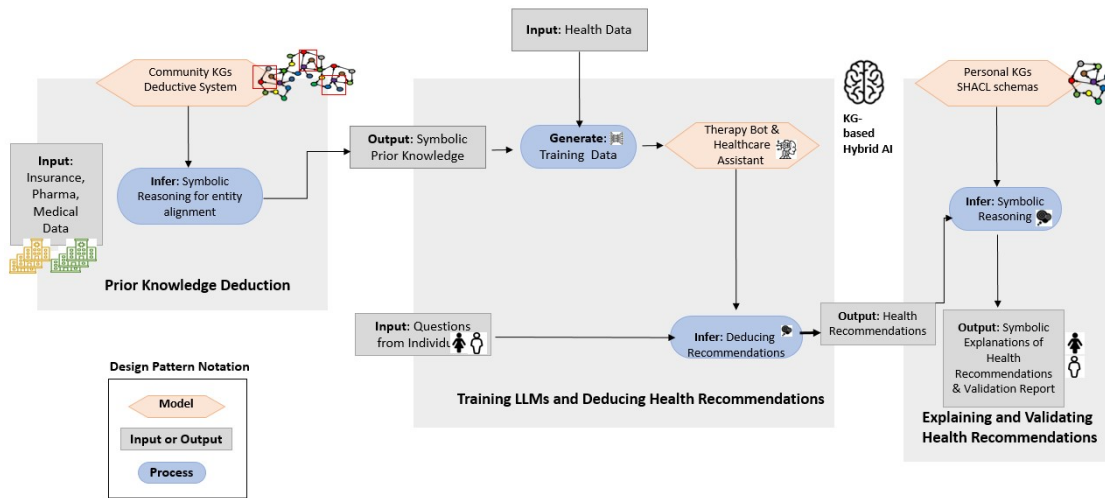
Figure 6 Design Patterns for Hybrid AI. Extension of patterns by van Bekkum et al. [111] for running example in Figure 2. The patterns represent an explainable system with prior knowledge created by the alignments of data from health-related data sources (e.g., insurance, pharma, and medical data).

symbolic reasoning over KGs with LLMs, the propagation of misinformation can be mitigated while simultaneously enhancing the transparency and trustworthiness of AI-generated outputs. Consequently, KG-based AI systems can effectively emulate human behaviour by subjecting mistakes arising from false or incomplete information to a process of validation and enrichment using curated and potentially peer-reviewed sources of knowledge [111].

Quality and AI. A notable application of KGs in neuro-symbolic AI is as a source of informative prior knowledge to increase the quality of machine learning models. An example is the work by Rivas et al. [91], where a deductive database, expressed in Datalog, establishes an axiomatic system of the pharmacokinetic behaviour of a treatment's drugs and enables the deduction of new drug-drug interactions in cancer treatments. This prior knowledge plays a crucial role in elucidating the characteristics of a therapy and justifying its efficacy by considering all the interactions and the dynamic movement of drugs within the body. It encompasses factors such as the absorption, bioavailability, metabolism, and excretion of drugs over time. A KG embedding model improves its prediction of the effectiveness of a treatment, based on the prior knowledge which encodes statements about a treatment's characteristics; these statements are inferred by a deductive system which comprises the symbolic component of the hybrid approach. An approach for explaining link prediction (e.g., [92]) allows the justification of why this added prior knowledge affects the model's decisions, potentially improving trust on the model's results.

Grounding based on our Illustrative Scenario. Grounding on the example presented in Figure 2, when individuals and professionals engage in communities with bots and assistants powered by AI models, it is critical to ensure the transparency of their decision-making process. However, despite the increasing focus on LLMs in healthcare and their continual improvement in terms of precision and accuracy [102], their outcomes can still be susceptible to hidden biases and a lack of traceability [63]. To tackle these challenges, the utilisation of a neuro-symbolic system

can enhance LLMs by incorporating reasoning capabilities. This system operates as a deductive system on a user's Knowledge Graph (KG). These hybrid AI systems can be effectively modelled using patterns proposed by [111]. Figure 6 depicts a pattern describing a hybrid AI system that enhances the explainability of the LLMs described in our running example. At the community level, symbolic reasoning applied to the ontology of shared PKGs can generate prior knowledge, enabling precise and concrete questioning of an LLM and providing additional contextual information. Moreover, a symbolic system facilitates the linking of shared PKGs with corresponding entities in KGs related to insurance, pharmaceuticals, and medical research. By incorporating this prior knowledge, the LLM's answers are improved and validated with the assistance of the symbolic system. The systems operating at the community level and involving heterogeneous sources can be described using the *explainable system with prior knowledge* pattern; data alignments comprising prior knowledge enhance contextual knowledge provided to the therapy bot, facilitating thoughtful health recommendations.

5 Proposed KG-based AI for Self-determination Research Agenda

In this section, we derive a set of requirements concerning KG-based AI for self-determination and map them to the concrete research goals introduced at the start of this vision paper.

5.1 Trust, Accountability, and Autonomy Foundational Goals

In the following, we highlight five open research challenges and opportunities in each of our proposed foundational topics (machine-readable norms and policies; decentralised infrastructure; decentralised KG management; and explainable neuro-symbolic AI). Considering the complex nature of each of these requirements, an assessment of the maturity of existing technologies with respect to the various requirements is beyond the scope of a vision paper.

Machine-readable Norms and Policies.

MRP1: Seamless policy translation. There is a need for humans to express policies in machine-readable format and for machines to express them in natural language or via appropriate visualisations. A major challenge involves checking that machine readable policies faithfully represent their human readable counterpart.

MRP2: Multi-level policy evaluation. Several policy languages exist, however many of them do not have corresponding enforcement mechanisms. Given that usage constraints, community rules, and regulations operate at different, yet interconnected levels, there is a need to devise effective and efficient enforcement and/or compliance checking strategies.

MRP3: Negotiation. Facilitate autonomy via fair and safe negotiation between individuals, communities, and organisations. Here there is a need to study the benefits and tradeoffs between merely assisting humans in taking decisions and developing automated approaches that alleviate individuals from constant affirmations (e.g., the cookie problem).

MRP4: Compliance verification. Provide support for both ex-ante and ex-post compliance checking mechanisms. Despite their potential, it remains to be seen which machine-readable agreements can actually be enforced by trusted execution environments. Additionally, in scenarios where it does not pay data processors to cheat, game theoretic approaches could be used to underpin honours based compliance checking.

MRP5: Data misuse detection. Instil trust and to ensure accountability in KG-based AI, by developing mechanisms that can detect if any party violated policies and norms. In this context, causal reasoning and explanations could potentially be used to both detect misuse and to better understand the root cause.

Decentralised Infrastructure.

DI1: Comprehensive recording. A DLT can provide an immutable ledger but work remains on how best to connect KG-based AI activities, e.g., to a possible federated query engine.

DI2: Personalised tracing. Providing individual and community owners of PKGs with personalised traces of how acquired data was processed and used, will involve dis-aggregating KG-processing and inferencing according to different user data and ensuring that privacy is not violated when individual results are returned.

DI3: 'Decency' check. There is a need for easy-to-use services which allow users and communities to check if an organisation has behaved in a 'decent' way when it processed acquired data. Research here will examine how 'decency' can be defined and validated by comparing PKG declarations of use (e.g., policies) with generated traces of use.

DI4: Interoperability. Develop mechanisms that facilitate comprehensive interoperable identification of human and machine participants in KG-based AI processes. For example, users and communities will wish to know, and be able to validate, claims that a data request comes from a particular organisation, unit and even individual KG processor. This will provide a foundation for accountability at all levels of granularity.

DI5: Self-sovereignty. True self-sovereign KG-based AI needs to be: (i) based upon easy-to-use self-sovereign identities and data management; and (ii) capable of supporting the continuous monitoring of organisational behaviours in a transparent fashion.

Decentralised KG Management.

DKG1: Knowledge Sanitisation. Develop robust techniques for knowledge sanitisation that ensure user privacy by anonymising and filtering sensitive information based on data policies. These policies can be regulations such as GDPR and HIPAA, as well as individual-level data policies enforced at their personal data store, empowering users to specify their sharing preferences and control the aspects of data they disclose.

DKG2: Knowledge Graph Aggregation. Design and implement mechanisms to encourage users to contribute their PKGs towards aggregated knowledge graphs, such as a concerted effort towards developing specific disease KGs. Blockchain-based incentive models that reward users for contributing to constructing such knowledge graphs, fostering collaborative efforts, and enriching the overall quality of shared knowledge are components of this goal.

DKG3: Knowledge Verification. Develop community-based and expert processes to verify the knowledge available in the global KGs. On the community front, it is critical to ensure that a knowledge item that was previously contributed through an individual has not been altered (either through error or with malicious intent), for instance via blockchain primitives, as explained in the previous section.

DKG4: Knowledge Validation. Validation of knowledge is paramount to ensure KG interoperability and the consumption of knowledge in target applications. By employing RDF and SHACL technologies, we ensure that the DKGs across different data stores conform to a specific template, thus, enabling their integration with community-supported KGs.

DKG5: Federated Querying. Explore and implement federated querying mechanisms, specifically utilising SPARQL, to enable efficient querying across integrated KGs. This process includes developing techniques to support various institutions, such as insurers, pharmaceutical companies, and medical research organisations, accessing and extracting insights from the knowledge graphs to enhance decision-making and advance their respective domains.

Explainable Neuro-Symbolic AI.

XNS1: User-dependent Recommendations. Neuro-symbolic systems need to be empowered to transparently present results to the users according to their interests. For example, in our illustrative scenario, an individual may not expect the same level of detail in a health recommendation as a medical doctor or a community representative.

XNS2: Adaptive Hybrid AI. Define models that can adaptively combine predictive models with logical reasoning, encompassing abilities such as generalisation and causal inference. For accountability, the neuro-symbolic system should explain when the combination of logical reasoning with a therapy bot or healthcare assistant will be beneficial. For autonomy, the neuro-symbolic system should include the user in the loop and consider their opinion in this decision. Finally, trust requires verifying and validating these decisions.

XNS3: Contextual-based Hybrid AI. Equip neuro-symbolic systems with contextual knowledge, reasoning capabilities, and causal inference to effectively evaluate the strengths and limitations of machine learning components. This goal empowers the system to identify optimal combinations of statistical and symbolic AI methods, requiring the definition of causal models on top of KGs capable of combining reasoning over KGs with causal inference.

XNS4: Symbolic Reasoning. Employ inference processes, both inductive and deductive, on knowledge graphs to enable ML models, and LLMs in particular, to adjust hyper-parameters and a model's configuration, to new environments (i.e., Personal, community-based, and integrated healthcare KGs) and provide explanations for their decisions. Despite the advances of Automated Machine Learning (AutoML) systems (e.g., AutoML[25] and AutoWeka [59], to best of our knowledge, there are no developments for AutoML over KGs or for neuro-symbolic systems, which will enhance accountability, autonomy, and trust.

XNS5: Learning Transparency. Investigate if existing XAI mechanisms can be tailored for learning transparency, such that it is possible to explain what action was take; how the decision making was performed; and why this was perceived as the outcome offering the greatest expected satisfaction.

5.2 AI for Self-determination

The identified foundational research topic challenges and opportunities can be used to better contextualise concrete goals in relation to trust, accountability, and autonomy from a KG-based AI for self-determination perspective. An overview of this mapping, which is depicted in Table 1, is provided by attempting to answer the overarching questions that guide our vision paper.

(Q1) What are the key requirements for an AI system to produce trustable results? From a trust perspective, it is important that machine-readable policies faithfully represent the human policies (MRP1) in a manner that can be verified automatically (MRP2). Regardless of whether systems are automated or semi-automated, we need to be able to verify that processes behave as expected (MRP4) and any misuse can be detected and rectified (MRP5). Trust could potentially be facilitated via auditing (DI1) and tracing (DI2), as well as certification mechanisms that support decency checks (DI3) and (semi-)automated knowledge verification (DKG3) and validation (DKG4) techniques. While human involvement is paramount to establishing trust in adaptive (XNS2) and contextualised (XNS3) hybrid AI.

---

[25] https://www.automl.org/

Table 1 Mapping of foundational requirements to pillars. A checkmark signifies that the corresponding requirement is necessary for answering a research question related to a pillar.

	Trust	Accountability	Autonomy
Machine-readable norms and policies			
MRP1	✓		✓
MRP1	✓		✓
MRP3			✓
MRP4	✓		✓
MRP5	✓	✓	✓
Decentralised Infrastructur			
DI1	✓		✓
DI2	✓		✓
DI3	✓		✓
DI4		✓	
DI5			✓
Decentralised KG Management.			
DKG1			✓
DKG2		✓	✓
DKG3	✓		
DKG4	✓		
DKG5		✓	
Explainable Neuro-Symbolic AI			
XNS1		✓	
XNS2	✓	✓	
XNS3	✓	✓	
XNS4		✓	
XNS5		✓	

(Q2) How can AI be made accountable for its decision-making? The first step to achieving accountability is to ensure it is possible to detect if any party violated policies and norms (MRP5) and that the recommendations given and decisions taken using both induction and deduction (XNS4) are comprehensible from a users perspective, for instance via user focuses recommendations (XNS1), providing explanations for recommendations and decisions (XNS2), facilitating learning transparency (XNS5), and contextualisation based on causal inference (XNS3). Considering that machines can only work with the knowledge that it has at hand, it is important that systems are able to integrate knowledge from disparate sources (DI4), and are capable of querying (DKG5) and aggregating (DKG2) relevant sources.

(Q3) How can citizens maintain autonomy as users or subjects of KG-based AI systems? Citizens' autonomy in a KG-based AI context is necessary to ensure that humans are able to control not only who has access to their personal data, but also that its usage is in line with existing regulatory requirements. The could be achieved with automated compliance checking (MRP4) and misuse detection (MRP5) built on top of machine-readable policies (MRP1) and evaluation mechanisms (MRP2). Negotiation could potentially enable organisations to gain access to better quality data (MRP3) or to foster collaboration via aggregation (DKG2) and strong privacy guarantees via anonymisation (DKG1). While, self-sovereign identities (DI5), auditing (DI1), tracing (DI2), and decency certification (DI3) have a major role to play when it comes to continuous monitoring.

6   Conclusion

This paper presents a compelling argument for integrating KG-based AI to empower individuals' self-determination and benefit society. This overarching goal is supported by three fundamental pillars: trust, accountability, and autonomy. We advocate the foundations of these pillars require focused research in four areas: machine-readable norms and policies, decentralised infrastructure, decentralised KG management, and explainable neuro-symbolic AI. By drawing on a concrete scenario within the healthcare domain, we demonstrate the relevance of each foundational topic and outline a comprehensive research agenda for each of them.

We aspire for the insights presented in this paper to catalyse the creation of AI services that genuinely support citizens while upholding their rights. Responsible advancement of the foundational topics is crucial to ensure that future KG-based AI solutions are comprehensive and possess the qualities of being traceable, verifiable, and interpretable. It is essential that relevant legislation, such as the EU AI Act, provides clear guidance to steer the development of these forthcoming applications, emphasising the need for accurate, reliable, and transparent AI systems. Within this context, we recognise the Semantic Web community as uniquely positioned to drive transformative change and contribute solutions illuminating opaque AI models' workings. Through this concerted effort, we envision a paradigm shift in KG management and analytics that establishes KG-based AI to empower individuals in their pursuit of self-determination.

References

1 Gianmaria Ajani, Guido Boella, Luigi Di Caro, Livio Robaldo, Llio Humphreys, Sabrina Praduroux, Piercarlo Rossi, and Andrea Violato. The european taxonomy syllabus: A multilingual, multi-level ontology framework to untangle the web of european legal terminology. *Applied Ontology*, 11(4):325–375, 2016. `doi:10.3233/AO-170174`.

2 Farahnaz Akrami, Mohammed Samiul Saeef, Qingheng Zhang, Wei Hu, and Chengkai Li. Realistic re-evaluation of knowledge graph completion methods: An experimental study. In *Proceedings of the 2020 International Conference on Management of Data, SIGMOD Conference 2020, online conference [Portland, OR, USA], June 14-19, 2020*, pages 1995–2010, 2020.

3 Tara Athan, Harold Boley, Guido Governatori, Monica Palmirani, Adrian Paschke, and Adam Wyner. Oasis legalruleml. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Law*, ICAIL '13, page 3–12, 2013. `doi:10.1145/2514601.2514603`.

4 Tara Athan, Guido Governatori, Monica Palmirani, Adrian Paschke, and Adam Wyner. *LegalRuleML: Design Principles and Foundations*, pages 151–188. 2015. `doi:10.1007/978-3-319-21768-0_6`.

5 Reyhan Aydoğan and Catholijn M. Jonker. A survey of decision support mechanisms for negotiation. In *Recent Advances in Agent-Based Negotiation*, Studies in Computational Intelligence, pages 30–51, 2023. `doi:10.1007/978-981-99-0561-4_3`.

6 Tim Baarslag, Michael Kaisers, Enrico H. Gerding, Catholijn M. Jonker, and Jonathan Gratch. *Self-sufficient, Self-directed, and Interdependent Negotiation Systems: A Roadmap Toward Autonomous Negotiation Agents*, pages 387–406. 2022. `doi:10.1007/978-3-030-76666-5_18`.

7 Krisztian Balog and Tom Kenter. Personal Knowledge Graphs: A Research Agenda. In *Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval*, pages 217–220, September 2019. `doi:10.1145/3341981.3344241`.

8 Krisztian Balog and Tom Kenter. Personal knowledge graphs: A research agenda. In *Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval*, pages 217–220, 2019.

9 Chitta Baral. *Knowledge Representation, Reasoning and Declarative Problem Solving.* 2003.

10 Cesare Bartolini, Antonello Calabró, and Eda Marchetti. Enhancing business process modelling with data protection compliance: An ontology-based proposal. In *ICISSP*, pages 421–428, 2019.

11 Davide Basile, Claudio Di Ciccio, Valerio Goretti, and Sabrina Kirrane. Blockchain based resource governance for decentralized web environments. *Frontiers in Blockchain*, 6:1141909, 2023.

12 Tal Ben-Nun and Torsten Hoefler. Demystifying parallel and distributed deep learning: An in-depth concurrency analysis. *ACM Comput. Surv.*, 52(4), aug 2019. `doi:10.1145/3320060`.

13 Tim Berners-Lee and Mark Fischetti. *Weaving the web: the past, present and future of the World Wide Web by its inventor.* Reprinted edition, 2000.

14 Tim Berners-Lee, James Hendler, and Ora Lassila. The Semantic Web. *Scientific American*, 284(5):34–43, 2001. nopublisher: Scientific American, a division of Nature America, Inc.

15 Guido Boella, Luigi Di Caro, and Valentina Leone. Semi-automatic knowledge population in a legal document management system. *Artif. Intell. Law*, 27(2):227–251, 2019. `doi:10.1007/s10506-018-9239-8`.

16 PA Bonatti, JL De Coi, D Olmedilla, and L Sauro. Protune: A rule-based provisional trust negotiation framework. *IEEE Trans. on Knowl. and Data Eng.(TKDE)*, 22(11):1507–1520, 2010.

17 Grady Booch, Francesco Fabiano, Lior Horesh, Kiran Kate, Jonathan Lenchner, Nick Linck, Andrea Loreggia, Keerthiram Murugesan, Nicholas Mattei, Francesca Rossi, and Biplav Srivastava. Thinking fast and slow in AI. In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021*, pages 15042–15046, 2021.

18 Anna Breit, Laura Waltersdorfer, Fajar J. Ekaputra, Marta Sabou, Andreas Ekelhart, Andreea Iana, Heiko Paulheim, Jan Portisch, Artem Revenko, Annette ten Teije, and Frank van Harmelen. Combining machine learning and semantic web: A systematic mapping study. *ACM Comput. Surv.*, mar 2023. Just Accepted. `doi:10.1145/3586163`.

19 Katrina Brooker. "I Was Devastated": Tim Berners-Lee, the Man Who Created the World Wide Web, Has Some Regrets. *Vanity Fair*, (August 2018), 2018.

20 Vitalik Buterin et al. A next-generation smart contract and decentralized application platform. *white paper*, 3(37):2–1, 2014.

21 Elena Cabrio, Alessio Palmero Aprosio, and Serena Villata. These are your rights - A natural language processing approach to automated RDF licenses generation. In *The Semantic Web: Trends and Challenges - 11th International Conference, ESWC 2014, Anissaras, Crete, Greece, May 25-29, 2014. Proceedings*, volume 8465 of *Lecture Notes in Computer Science*, pages 255–269, 2014. `doi:10.1007/978-3-319-07443-6\_18`.

22 Cristian Cardellino, Serena Villata, Laura Alonso Alemany, and Elena Cabrio. Information extraction with active learning: A case study in legal text. In *Computational Linguistics and Intelligent Text Processing: 16th International Conference, CICLing 2015, Cairo, Egypt, April 14-20, 2015, Proceedings, Part II 16*, pages 483–494. Springer, 2015.

23 Ilias Chalkidis, Charalampos Nikolaou, Panagiotis Soursos, and Manolis Koubarakis. Modeling and querying greek legislation using semantic web technologies. In *The Semantic Web - 14th International Conference, ESWC 2017, Portorož, Slovenia, May 28 - June 1, 2017, Proceedings, Part I*, pages 591–606, 2017. `doi:10.1007/978-3-319-58068-5\_36`.

24 Shruthi Chari, Oshani Seneviratne, Daniel M Gruen, Morgan A Foreman, Amar K Das, and Deborah L McGuinness. Explanation ontology: a model of explanations for user-centered ai. In *The Semantic Web–ISWC 2020: 19th International Semantic Web Conference, Athens, Greece, November 2–6, 2020, Proceedings, Part II*, pages 228–243. Springer, 2020.

25 Aakanksha Chowdhery, Sharan Narang, Jacob Devlin, Maarten Bosma, Gaurav Mishra, Adam Roberts, Paul Barham, Hyung Won Chung, Charles Sutton, Sebastian Gehrmann, et al. Palm: Scaling language modeling with pathways. *arXiv preprint arXiv:2204.02311*, 2022.

26 Giovanni Luca Ciampaglia, Alexios Mantzarlis, Gregory Maus, and Filippo Menczer. Research challenges of digital misinformation: Toward a trustworthy web. *AI Magazine*, 39(1):65–74, 2018.

27 A. Feder Cooper, Emanuel Moss, Benjamin Laufer, and Helen Nissenbaum. Accountability in an Algorithmic Society: Relationality, Responsibility, and Robustness in Machine Learning. In *2022 ACM Conference on Fairness, Accountability, and Transparency*, pages 864–876, June 2022. `doi:10.1145/3531146.3533150`.

28 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.

29 Uwe Der, Stefan Jähnichen, and Jan Sürmeli. Self-sovereign identity − opportunities and challenges for the digital revolution, 2017. `arXiv:1712.01767`.

30 Johannes Dimyadi, Guido Governatori, and Robert Amor. Evaluating legaldocml and legalruleml as a standard for sharing normative information in the aec/fm domain. In *Proceedings of the Joint Conference on Computing in Construction (JC3)*, volume 1, pages 637–644. Heriot-Watt University, Edinburgh, UK. Heraklion, Greece, 2017.

31 John Domingue, Aisling Third, Maria-Esther Vidal, Philipp Rohde, Juan Cano, Andrea Cimmino, and Ruben Verborgh. Trusting decentralised knowledge graphs and web data at the web conference. In *Companion Proceedings of the ACM Web Conference 2023*, pages 1422–1423, 2023.

32 Xin Luna Dong. Generations of knowledge graphs: The crazy ideas and the business impact. *CoRR*, abs/2308.14217, 2023. `doi:10.48550/arXiv.2308.14217`.

33 Beatriz Esteves, Harshvardhan J Pandit, and Víctor Rodríguez-Doncel. Odrl profile for expressing consent through granular access control policies in solid. In *2021 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, pages 298–306. IEEE, 2021.

34 Enrico Francesconi. A description logic framework for advanced accessing and reasoning over normative provisions. *Artificial intelligence and Law*, 22(3):291–311, 2014.

35 Rinon Gal, Yuval Alaluf, Yuval Atzmon, Or Patashnik, Amit H Bermano, Gal Chechik, and Daniel Cohen-Or. An image is worth one word: Personalizing text-to-image generation using textual inversion. In *Proceedings of the Twelfth International Conference on Learning Representations*, 2023.

36 Martin Gebser, Roland Kaminski, Benjamin Kaufmann, and Torsten Schaub. Clingo = ASP + control: Preliminary report. *CoRR*, abs/1405.3694, 2014.

37 Cindy Gordon. ChatGPT Is The Fastest Growing App In The History Of Web Applications, February 2023.

38 Guido Governatori, Mustafa Hashmi, Ho-Pun Lam, Serena Villata, and Monica Palmirani. Semantic business process regulatory compliance checking using LegalRuleML. In *European Knowledge Acquisition Workshop*, 2016.

39 Paul Groth, Andrew Gibson, and Jan Velterop. The anatomy of a nanopublication. *Information services & use*, 30(1-2):51–56, 2010.

40 Susanne Guth, Gustaf Neumann, and Mark Strembeck. Experiences with the enforcement of access rights extracted from odrl-based digital contracts. In *Proceedings of the 3rd ACM workshop on Digital rights management*, pages 90–102, 2003.

41 Kristian J. Hammond and David B. Leake. Large language models need symbolic AI. In Artur S. d'Avila Garcez, Tarek R. Besold, Marco Gori, and Ernesto Jiménez-Ruiz, editors, *Proceedings of the 17th International Workshop on Neural-Symbolic Learning and Reasoning, La Certosa di Pontignano, Siena, Italy, July 3-5, 2023*, volume 3432 of *CEUR Workshop Proceedings*, pages 204–209. CEUR-WS.org, 2023. URL: `https://ceur-ws.org/Vol-3432/paper17.pdf`.

42 Dick Hardt. The OAuth 2.0 Authorization Framework. RFC 6749, October 2012. `doi:10.17487/RFC6749`.

43 Giray Havur, Simon Steyskal, Oleksandra Panasiuk, Anna Fensel, Victor Mireles, Tassilo Pellegrini, Thomas Thurner, Axel Polleres, and Sabrina Kirrane. Automatic license compatibility checking. In *SEMANTiCS Posters&Demos*, 2019.

44 Michał Robert Hoffman, Luis-Daniel Ibáñez, and Elena Simperl. Scholarly publishing on the blockchain–from smart papers to smart informetrics. *Data Science*, 2(1-2):291–310, 2019.

45 Hongsheng Hu, Zoran Salcic, Lichao Sun, Gillian Dobbie, Philip S. Yu, and Xuyun Zhang. Membership inference attacks on machine learning: A survey. *ACM Comput. Surv.*, 54(11s), sep 2022. `doi:10.1145/3523273`.

46 Jie Huang and Kevin Chen-Chuan Chang. Towards reasoning in large language models: A survey. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 1049–1065, Toronto, Canada, July 2023. Association for Computational Linguistics. URL: `https://aclanthology.org/2023.findings-acl.67`, `doi:10.18653/v1/2023.findings-acl.67`.

47 Renato Ianella and Serena Villata. ODRL Information Model, 2018.

48 Eleni Ilkou. Personal Knowledge Graphs: Use Cases in e-learning Platforms. In *Companion Proceedings of the Web Conference 2022*, pages 344–348, April 2022. `doi:10.1145/3487553.3524196`.

49 Lalana Kagal et al. Rei: A policy language for the me-centric project. 2002.

50 Daniel Kahneman. *Thinking, fast and slow.* 2011.

51 Daniel Kazenoff, Oshani Seneviratne, and Deborah L McGuinness. Semantic graph analysis to combat cryptocurrency misinformation on the web. In *ASLD@ ISWC*, pages 168–176, 2020.

52 Ankesh Khandelwal, Jie Bao, Lalana Kagal, Ian Jacobi, Li Ding, and James Hendler. Analyzing the air language: a semantic web (production) rule language. In *International Conference on Web Reasoning and Rule Systems*, pages 58–72. Springer, 2010.

53 Emre Kiciman, Robert Ness, Amit Sharma, and Chenhao Tan. Causal reasoning and large language models: Opening a new frontier for causality. *CoRR*, abs/2305.00050, 2023. `doi: 10.48550/arXiv.2305.00050`.

54 Sabrina Kirrane. Intelligent software web agents: A gap analysis. *Journal of Web Semantics*, 71:100659, November 2021. `doi:10.1016/j.websem.2021.100659`.

55 Sabrina Kirrane, Javier D. Fernández, Piero Bonatti, Uros Milosevic, Axel Polleres, and Rigo Wenning. The special-k personal data processing transparency and compliance platform, 2021. `arXiv:2001.09461`.

56 Sabrina Kirrane, Alessandra Mileo, and Stefan Decker. Access control and the Resource Description Framework: A survey. *Semantic Web*, 8(2):311–352, December 2016. `doi: 10.3233/SW-160236`.

57 Pang Wei Koh and Percy Liang. Understanding black-box predictions via influence functions. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 1885–1894, 06–11 Aug 2017.

58 Boshko Koloski, Timen Stepišnik Perdih, Marko Robnik-Šikonja, Senja Pollak, and Blaž Škrlj. Knowledge graph informed fake news classification via heterogeneous representation ensembles. *Neurocomputing*, 496:208–226, 2022.

59 Lars Kotthoff, Chris Thornton, Holger H. Hoos, Frank Hutter, and Kevin Leyton-Brown. Auto-weka 2.0: Automatic model selection and hyperparameter optimization in WEKA. *J. Mach. Learn. Res.*, 18:25:1–25:5, 2017.

60 Ziyi Kou, Lanyu Shang, Yang Zhang, and Dong Wang. Hc-covid: A hierarchical crowdsource knowledge graph approach to explainable covid-19 misinformation detection. *Proceedings of the ACM on Human-Computer Interaction*, 6(GROUP):1–25, 2022.

61 Tahu Kukutai and Donna Cormack. "Pushing the space" Data sovereignty and self-determination in Aotearoa NZ. In *Indigenous Data Sovereignty and Policy*, Routledge Studies in Indigenous Peoples and Policy. 2021.

62 Freddy Lecue. On the role of knowledge graphs in explainable AI. *Semantic Web*, 11(1):41–51, January 2020. `doi:10.3233/SW-190374`.

63 Hanzhou Li, John T Moon, Saptarshi Purkayastha, Leo Anthony Celi, Hari Trivedi, and Judy W Gichoya. Ethics of large language models in medicine and medical research. *The Lancet, Digital Health*, 5, 2023. `doi:https://doi.org/10.1016/S2589-7500(23)00083-3`.

64 James Lighthill. Artifical Intelligence: A General Survey. Technical report, UK Science Research Council.

65 Lu Liu, Sicong Zhou, Huawei Huang, and Zibin Zheng. From technology to society: An overview of blockchain-based dao. *IEEE Open Journal of the Computer Society*, 2:204–215, 2021. `doi:10.1109/OJCS.2021.3072661`.

66 James H Lubowitz. Chatgpt, an artificial intelligence chatbot, is impacting medical literature. *Arthroscopy*, 39(5):1121–1122, 2023.

67 George F. Luger. Modern AI and How We Got Here. In *Knowing our World: An Artificial Intelligence Perspective*, pages 49–74. 2021. `doi:10.1007/978-3-030-71873-2_3`.

68 Essam Mansour, Andrei Vlad Sambra, Sandro Hawke, Maged Zereba, Sarven Capadisli, Abdurrahman Ghanem, Ashraf Aboulnaga, and Tim Berners-Lee. A demonstration of the solid platform for social web applications. In *Proceedings of the 25th international conference companion on world wide web*, pages 223–226, 2016.

69 Mohit Mayank, Shakshi Sharma, and Rajesh Sharma. Deap-faked: Knowledge graph based approach for fake news detection. In *2022 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 47–51. IEEE, 2022.

70 Dhruv Mehrotra. ICE Records Reveal How Agents Abuse Access to Secret Data, April 2023.

71 Merriam-Webster.com Dictionary. Trust. https://www.merriam-webster.com/dictionary/trust. Accessed 14 Jul. 2023.

72 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.

73 Melanie Mitchell and David C. Krakauer. The debate over understanding in ai's large language models. *Proceedings of the National Academy of Sciences*, 120(13):e2215907120, 2023. `doi:10.1073/pnas.2215907120`.

74 Jakob Mökander, Jonas Schuett, Hannah Rose Kirk, and Luciano Floridi. Auditing large language models: a three-layered approach. *AI Ethics*, 2023. `doi:10.1007/s43681-023-00289-2`.

75 Benjamin Moreau, Patricia Serrano-Alvarado, Matthieu Perrin, and Emmanuel Desmontils. A license-based search engine. In *The Semantic Web: ESWC 2019 Satellite Events: ESWC 2019 Satellite Events, Portorož, Slovenia, June 2–6, 2019, Revised Selected Papers 16*, pages 130–135. Springer, 2019.

76 Iman Naja, Milan Markovic, Peter Edwards, and Caitlin Cottrill. A Semantic Framework to Support AI System Accountability and Audit. In *The Semantic Web*, volume 12731, pages 160–176. 2021. Series Title: Lecture Notes in Computer Science. `doi:10.1007/978-3-030-77385-4_10`.

77 Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system, 2008.

78 María Navas-Loro and Cristiana Santos. Events in the legal domain: first impressions. In *TERECOM@JURIX*, pages 45–57, 2018.

79 Arttu Oksanen, Minna Tamper, Jouni Tuominen, Eetu Mäkelä, Aki Hietanen, and Eero Hyvönen. Semantic finlex: Transforming, publishing, and using finnish legislation and case law as linked open data on the web. *Knowledge of the Law in the Big Data Age*, 317:212–228, 2019.

80 Alessandro Oltramari, Dhivya Piraviperumal, Florian Schaub, Shomir Wilson, Sushain Cherivirala, Thomas B. Norton, N. Cameron Russell, Peter Story, Joel R. Reidenberg, and Norman M. Sadeh. Privonto: A semantic framework for the analysis of privacy policies. *Semantic Web*, 9(2):185–203, 2018. `doi:10.3233/SW-170283`.

81 Monica Palmirani, Guido Governatori, Antonino Rotolo, Said Tabet, Harold Boley, and Adrian Paschke. LegalRuleML: XML-based rules and norms. In *International Workshop on Rules and Rule Markup Languages for the Semantic Web*, pages 298–312. Springer, 2011.

82 Monica Palmirani, Michele Martoni, Arianna Rossi, Cesare Bartolini, and Livio Robaldo. Legal ontology for modelling gdpr concepts and norms. In *Legal Knowledge and Information Systems*, pages 91–100. 2018.

83 Monica Palmirani, Michele Martoni, Arianna Rossi, Cesare Bartolini, and Livio Robaldo. PrOnto: Privacy Ontology for Legal Reasoning. In *Electronic Government and the Information Systems Perspective*, volume 11032, pages 139–152. 2018. Series Title: Lecture Notes in Computer Science. `doi:10.1007/978-3-319-98349-3_11`.

84 Monica Palmirani, Michele Martoni, Arianna Rossi, Cesare Bartolini, and Livio Robaldo. Pronto: Privacy ontology for legal reasoning. In *Electronic Government and the Information Systems Perspective - 7th International Conference, EGOVIS 2018, Regensburg, Germany,*

September 3-5, 2018, Proceedings, volume 11032 of *Lecture Notes in Computer Science*, pages 139–152, 2018. `doi:10.1007/978-3-319-98349-3\_11`.

85 Harshvardhan J. Pandit, Kaniz Fatema, Declan O'Sullivan, and Dave Lewis. Gdprtext - GDPR as a linked data resource. In *The Semantic Web - 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3-7, 2018, Proceedings*, pages 481–495, 2018. `doi:10.1007/978-3-319-93417-4\_31`.

86 Eli Pariser. *The filter bubble: How the new personalized web is changing what we read and how we think*. 2011.

87 Neoklis Polyzotis and Matei Zaharia. What can Data-Centric AI Learn from Data and ML Engineering? 2021. nopublisher: arXiv Version Number: 1. `doi:10.48550/ARXIV.2112.06439`.

88 Shuofei Qiao, Yixin Ou, Ningyu Zhang, Xiang Chen, Yunzhi Yao, Shumin Deng, Chuanqi Tan, Fei Huang, and Huajun Chen. Reasoning with language model prompting: A survey. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 5368–5393, Toronto, Canada, July 2023. Association for Computational Linguistics. URL: `https://aclanthology.org/2023.acl-long.294`, `doi:10.18653/v1/2023.acl-long.294`.

89 Umair Qudus, Michael Röder, Muhammad Saleem, and Axel-Cyrille Ngonga Ngomo. HybridFC: A Hybrid Fact-Checking Approach for Knowledge Graphs. In *The Semantic Web – ISWC 2022*, pages 462–480, 2022.

90 Enayat Rajabi and Kobra Etminani. Knowledge-graph-based explainable AI: A systematic review. *Journal of Information Science*, page 016555152211128, September 2022. `doi:10.1177/01655515221112844`.

91 Ariam Rivas, Diego Collarana, Maria Torrente, and Maria-Esther Vidal. A neuro-symbolic system over knowledge graphs for link prediction. *Semantic Web*, 2023.

92 Andrea Rossi, Donatella Firmani, Paolo Merialdo, and Tommaso Teofili. Explaining link prediction systems based on knowledge graph embeddings. In *SIGMOD '22: International Conference on Management of Data, Philadelphia, PA, USA, June 12 - 17, 2022*, pages 2062–2075, 2022. `doi:10.1145/3514221.3517887`.

93 Andrei Vlad Sambra, Essam Mansour, Sandro Hawke, Maged Zereba, Nicola Greco, Abdurrahman Ghanem, Dmitri Zagidulin, Ashraf Aboulnaga, and Tim Berners-Lee. Solid: a platform for decentralized social applications based on linked data. *MIT CSAIL & Qatar Computing Research Institute, Tech. Rep.*, 2016.

94 Michael Schmidt, Olaf Görlitz, Peter Haase, Günter Ladwig, Andreas Schwarte, and Thanh Tran. Fedbench: A benchmark suite for federated semantic data query processing. In *The Semantic Web–ISWC 2011: 10th International Semantic Web Conference, Bonn, Germany, October 23-27, 2011, Proceedings, Part I 10*, pages 585–600. Springer, 2011.

95 John Schulman, Barret Zoph, Christina Kim, Jacob Hilton, Jacob Menick, Jiayi Weng, Juan Felipe Ceron Uribe, Liam Fedus, Luke Metz, Michael Pokorny, et al. Chatgpt: Optimizing language models for dialogue. *OpenAI blog*, 2022.

96 Oshani Seneviratne. Data provenance and accountability on the web. *Provenance in Data Science: From Data Models to Context-Aware Knowledge Graphs*, pages 11–24, 2020.

97 Oshani Seneviratne. Blockchain for social good: Combating misinformation on the web with ai and blockchain. In *Proceedings of the 14th ACM Web Science Conference 2022*, pages 435–442, 2022.

98 Lanyu Shang, Ziyi Kou, Yang Zhang, Jin Chen, and Dong Wang. A privacy-aware distributed knowledge graph approach to qois-driven covid-19 misinformation detection. In

2022 IEEE/ACM 30th International Symposium on Quality of Service (IWQoS), pages 1–10. IEEE, 2022.

99   Manan Shukla, Jianjing Lin, and Oshani Seneviratne. Blockiot: blockchain-based health data integration using iot devices. In *AMIA Annual Symposium Proceedings*, volume 2021, page 1119. American Medical Informatics Association, 2021.

100 Manan Shukla, Jianjing Lin, and Oshani Seneviratne. Blockiot-RETEL: Blockchain and iot based read-execute-transact-erase-loop environment for integrating personal health data. In *2021 IEEE International Conference on Blockchain (Blockchain)*, pages 237–243. IEEE, 2021.

101 Manan Shukla, Jianjing Lin, and Oshani Seneviratne. Collaboratively learning optimal patient outcomes using smart contracts in limited data settings. In *2022 IEEE/ACM Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*, pages 133–137. IEEE, 2022.

102 Karan Singhal, Tao Tu, Juraj Gottweis, Rory Sayres, Ellery Wulczyn, Le Hou, Kevin Clark, Stephen Pfohl, Heather Cole-Lewis, Darlene Neal, Mike Schaekermann, Amy Wang, Mohamed Amin, Sami Lachgar, Philip Andrew Mansfield, Sushant Prakash, Bradley Green, Ewa Dominowska, Blaise Agüera y Arcas, Nenad Tomasev, Yun Liu, Renee Wong, Christopher Semturs, S. Sara Mahdavi, Joelle K. Barral, Dale R. Webster, Gregory S. Corrado, Yossi Matias, Shekoofeh Azizi, Alan Karthikesalingam, and Vivek Natarajan. Towards expert-level medical question answering with large language models. *CoRR*, abs/2305.09617, 2023. `doi:10.48550/arXiv.2305.09617`.

103 Manu Sporny, Amy Guy, Markus Sabadello, Drummond Reed, Orie Steele, and Christopher Allen. Decentralized Identifiers (DIDs), 2022.

104 Manu Sporny, David Longley, and David Chadwick. Verifiable Credentials Data Model, 2022.

105 Kai Sun, Yifan Ethan Xu, Hanwen Zha, Yue Liu, and Xin Luna Dong. Head-to-tail: How knowledgeable are large language models (llm)? A.K.A. will llms replace knowledge graphs? *CoRR*, abs/2308.10168, 2023. `doi:10.48550/arXiv.2308.10168`.

106 Ali Sunyaev and Ali Sunyaev. Distributed ledger technology. *Internet computing: Principles of distributed systems and emerging internet-based technologies*, pages 265–299, 2020.

107 Nick Szabo. Formalizing and securing relationships on public networks. *First monday*, 1997.

108 A. Third, G. Gkotsis, E. Kaldoudi, G. Drosatos, N. Portokallidis, S. Roumeliotis, K. Pafili, and J. Domingue. Integrating medical scientific knowledge with the semantically quantified self. In *The Semantic Web–ISWC 2016: 15th International Semantic Web Conference, Kobe, Japan, October 17–21, 2016, Proceedings, Part I 15*, pages 566–580. Springer, 2016.

109 Ilaria Tiddi and Stefan Schlobach. Knowledge graphs as tools for explainable machine learning: A survey. *Artificial Intelligence*, 302:103627, January 2022. `doi:10.1016/j.artint.2021.103627`.

110 Andrzej Uszok, Jeffrey Bradshaw, Renia Jeffers, Niranjan Suri, Patrick Hayes, Maggie Breedy, Larry Bunch, Matt Johnson, Shriniwas Kulkarni, and James Lott. Kaos policy and domain services: Toward a description-logic approach to policy representation, deconfliction, and enforcement. In *Proceedings POLICY 2003. IEEE 4th International Workshop on Policies for Distributed Systems and Networks*, pages 93–96. IEEE, 2003.

111 Michael van Bekkum, Maaike de Boer, Frank van Harmelen, André Meyer-Vitali, and Annette ten Teije. Modular design patterns for hybrid learning and reasoning systems. *Appl. Intell.*, 51(9):6528–6546, 2021. `doi:10.1007/s10489-021-02394-3`.

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

[113] Kang Wei, Jun Li, Ming Ding, Chuan Ma, Howard H Yang, Farhad Farokhi, Shi Jin, Tony QS Quek, and H Vincent Poor. Federated learning with differential privacy: Algorithms and performance analysis. *IEEE Transactions on Information Forensics and Security*, 15:3454–3469, 2020.

[114] Mika Westerlund. The emergence of deepfake technology: A review. *Technology innovation management review*, 9(11), 2019.

[115] Steven Euijong Whang, Yuji Roh, Hwanjun Song, and Jae-Gil Lee. Data collection and quality challenges in deep learning: A data-centric ai perspective. *The VLDB Journal*, 32(4):791–813, 2023.

[116] Gavin Wood. Ethereum: A secure decentralised generalised transaction ledger. *Ethereum Project Yellow Paper*, 2014.

[117] Adam Z Wyner and Wim Peters. Lexical semantics and expert legal knowledge towards the identification of legal case factors. In *JURIX*, volume 10, pages 127–136, 2010.

[118] Guowen Xu, Hongwei Li, Yun Zhang, Shengmin Xu, Jianting Ning, and Robert H. Deng. Privacy-preserving federated deep learning with irregular users. *IEEE Transactions on Dependable and Secure Computing*, 19(2):1364–1381, 2022. `doi:10.1109/TDSC.2020.3005909`.

[119] Qiang Yang, Yang Liu, Tianjian Chen, and Yongxin Tong. Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(2):1–19, 2019.

[120] Bo Yin, Hao Yin, Yulei Wu, and Zexun Jiang. Fdc: A secure federated deep learning mechanism for data collaborations in the internet of things. *IEEE Internet of Things Journal*, 7(7):6348–6359, 2020. `doi:10.1109/JIOT.2020.2966778`.

[121] Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, Yifan Du, Chen Yang, Yushuo Chen, Zhipeng Chen, Jinhao Jiang, Ruiyang Ren, Yifan Li, Xinyu Tang, Zikang Liu, Peiyu Liu, Jian-Yun Nie, and Ji-Rong Wen. A survey of large language models. *CoRR*, abs/2303.18223, 2023. `doi:10.48550/arXiv.2303.18223`.

[122] Jiawei Zhou, Yixuan Zhang, Qianni Luo, Andrea G Parker, and Munmun De Choudhury. Synthetic lies: Understanding ai-generated misinformation and evaluating algorithmic and human solutions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–20, 2023.