An Ontology for Ethical AI Principles

Andrew Harrison [a,*], Dayana Spagnuelo [b] and Ilaria Tiddi [b]
[a] *Informatics Institute, University of Amsterdam, Amsterdam, Netherlands*
E-mail: andrew.harrison@student.uva.nl
[b] *Department of Computer Science , Vrije Universiteit Amsterdam, Amsterdam, Netherlands*
E-mails: d.spagnuelo@vu.nl, i.tiddi@vu.nl

Abstract. The initial trickle of organisations releasing Artificial Intelligence (AI) principles documents has turned into a flood, termed the proliferation of principles, with current counts exceeding over 300 of such documents. This has led researchers to apply traditional systematic review techniques to the growing corpus of knowledge. Aims vary from meta-analytic accounts of country of origin, gender of authors, and type of organisations, to mapping principles across documents, to attempts to consolidate the vast number of principles down to a set of core authoritative principles, to authors selecting principle documents to support a research hypothesis. The commonality underlying all these efforts is traditional research techniques, which are arguably inefficient, and create static artefacts with low reusability. The Semantic Web offers a different way, an avenue to examine this proliferating body of knowledge, creating dynamic knowledge graphs, richly and more objectively connecting principles as concepts, providing enhanced semantic querying, and incorporating the existing resources from the Linked Open Data cloud. In order to achieve this, an ontology for AI principles is first required. This work presents the first ontology for Ethical AI principles (AIPO), leveraging ontology vocabularies including Dublin Core, SKOS, FOAF and DCAT2 among others, and shows its applicability through a use-case based on the OECD's AI principles set. We further discuss the benefits of AIPO, including the facilitation of systematic studies and its impact over the AI principle sets landscape.

Keywords: Artificial Intelligence, Principles, Ethics, Policy, Ontology, Semantic Web, Linked Data

1. Introduction

The intelligence explosion predicted after the first ultra-intelligent machine is invented [1] is yet to arrive, but an Artificial Intelligence (AI) principles explosion certainly has, with the largest study revealing that 88% of principles were released after 2016 [2], giving the appearance that any organisation connected to technology policy is either producing their own or endorsing another's set of AI principles [3]. These principle sets serve as non-legislative policy instruments also known as soft-law [2], and the organisations producing them cover the whole gamut from civil society, through private sector, national government, intergovernmental and supranational organisations. A tally from April 2020 listed over 160 AI ethics-specific guidelines[1] while a broader count of more general AI policy initiatives revealed over 300[2]. This "principle proliferation" [4] rather than consolidation and uniformity has spurred meta-analyses of the principle sets, which are an attempt to utilise the growing body of knowledge by re-engineering it into a human consumable volume and format.

While the scale and purpose of these meta-studies varies, the methods that are employed are limited and can involve a significant amount of manpower. Most of these methods involve manual search and selection of principle sets through keyword and citation linking, followed by human analysis of the documents and manual coding of the content to derive discursive or analytical results, that are finally presented by researchers through an article. There is poor reusability between the meta-studies, with cross-citations often serving only to validate the difference and thus relevance of the current meta-study. Further, the PDF arte-

*Corresponding author. E-mail: andrew.harrison@student.uva.nl.
[1] https://inventory.algorithmwatch.org
[2] https://oecd.ai/countries-and-initiatives

fact outdates rather quickly, exacerbated by delays between submission and publishing, with articles taking pains to mention research cutoff dates, especially during this period of AI principle proliferation. While the article form itself is a paper-native means to freeze and mount research, there is a shift towards web-native systems to enhance the current research outputs [5]. The application of Semantic Web [6] and Linked Data [7] technology and methods allows browsability of graphs and rich querying services. Indeed, there are efforts towards moving academic research to full semantic publishing [8–10]).

The burgeoning field of AI ethics and principles serves as a bulwark to the multitude of malfeasant usages for AI [11], along with addressing safety concerns even when it is not used for ill [12], and it aims to protect from the short-term issues around bias and fairness [13] through to the longer-term debates over existential threats to humanity [14]. As mentioned, the principle sets need to be re-engineered to make them human-consumable in volume and format, with traditional meta-study methods an attempt to do this, but the creation of an ontology and use of Semantic Web tools and technology it is argued, a better way. Additionally, semantic technologies and Linked Data can make the principle sets machine readable, serving both to assist human comprehension through use of data mining, and becoming directly implementable into AI entities themselves. The consumability and use of these AI principles is important to ensure public accessibility and accountability, shared understanding between actors to prevent AI arms races, assistance in the design and deployment of AI systems, and finally to help drive and shape the presumably forthcoming hard-law and regulation of AI.

In this work, we then investigate whether the existing AI ethical principle sets can be structured using the existing semantic technologies, in order to facilitate systematic studies. To answer this question, we require two research questions, namely:

1. Can we formulate an ontology to transform the knowledge of AI ethical principles into a linked data consumable format?
2. What use can be made of such an ontology to facilitate the systematic study of AI principle sets, and further the impact of such studies?

Based on these questions, we focus on creating an ontology for ethical AI principles according to Semantic Web standards, where possible using existing ontology vocabularies to assist the integration with other Linked Data sources.

We first analyse 8 systematic studies of AI principles to identify their common goals, results and limitations (Section 2). We then implement the first ontology for ethical AI principles using the established ontology engineering steps, with a particular focus on promoting its interoperability through vocabulary reuse, our methodology is presented in Section 3. We show the applicability of the ontology through a grounded example based on the OECD's AI principles set (Section 4). We finally discuss the benefits of AIPO, namely how it supports researchers in performing systematic studies and what impact it has on the AI principle sets landscape (Section 5). AIPO will serve as the starting point for two future streams of work which we detail in Section 6. First, it will allow the creation of a standalone knowledge graph populated with AI principles sets, which will provide additional inferred knowledge and make available an endpoint for user-defined queries of this body of knowledge, to help drive further research and insights. Second, it will promote the integration of this body of knowledge into the Web of (Linked) Data, with the inherent advantages and use-cases from the volume and variety of other data sources.

2. Related Work

In this section, we first examine eight AI principle systematic studies, with the goal of highlighting the current efforts and their results in more detail. The section is purposefully exhaustive in description to highlight the difficulty of traditional methods, irreproducibility when methodologies are not adequately described, and the variance in their applications and findings. We then follow by descriptions of the three AI principle repositories hosted online along with their functionality, and finally provide examples of the use of ontologies in differing fields.

2.1. Survey of Meta-studies

The purposes of the systematic studies reviewed are non-exclusively categorised as:

i *meta-analysis*, i.e. aiming at deriving quantitatively meaningful information, such as female authors or number of principles;

ii *coverage maps*, showing which "principle sets" the specific principles belong to or are covered by in various applications;
iii *principle consolidations*, which attempt to derive an authoritative set of principles from the various principle sets; and
iv *hypothesis validation*, which seeks to provide evidence from the analysis of principle sets to support the research hypothesis proposed by the authors.

Of the 8 papers reviewed, 5 conduct meta-analysis, 6 use coverage maps, 4 attempt principle consolidation, and 4 seek to validate hypotheses. The number of AI principle sets reviewed in each study ranges from 7 to 84, with a median of 28.5, average of 31.9 and standard deviation of 24.1, with the selection criteria when mentioned, differing widely (see overview in Table 1).

The first study [4], argues that the volume of AI ethical principles is becoming overwhelming and confusing, and asks whether there is either similarity, meaning redundant repetition, or divergence, meaning ambiguity and disorientation. Results from the comparative analysis answer affirmatively to the former (redundancy of principles). They choose 6 principles sets comprising 47 principles in total, selected based on them being (i) recent, defined as less than 3 years old, (ii) relevant, i.e. impacting the whole of society, and (iii) reputable, i.e. employed by multi-stakeholder organisations with national or higher scope. These 47 principles are consolidated down to 5 core principles: beneficence, non-maleficence, autonomy, justice and explicability. The core principles are then supported via quotations selected from the principle sets, and principle mapping is done against 9 documents (3 additional added). The authors argue that their systematic analysis and principle consolidation serves as a basis for the development of laws and technical standards.

The second study [15], a self-described semi-systematic analysis, aims to analyse the ethics of AI through an evaluation of 22 AI principle sets sourced from database searches and the Algorithm Watch's AI Ethics Guidelines Global Inventory mentioned in the next subsection. The authors aim to show that the principle sets are having no impact on decision-making, have a lack of enforcement around their normative claims, and discourage agreement on a binding legal framework. Arguments are made that the principles are aimed at calming the public critical voices, while still maintaining the criticised practices within the organisations. The authors also want to show what is omitted from the principles, namely existential threats, machine consciousness, robot ethics, social cohesion, political abuse of AI systems, lack of diversity, Kantian brutalisation arguments, ecological costs, and industry-funded "buyout" of research institutions. The principle sets analysed were in English and had to be less than 5 years old, should not be applicable to only a national context with an exception made for Chinese, US, and European Commission works as these bodies are considered "AI superpowers". Also, corporate policies were not considered unless they had garnered a lot of media coverage or were multi-stakeholder such as IEEE or Partnership on AI. Other criteria and exception sub-clauses were considered, with an eventual summation that it includes principle sets functioning as a comprehensive mapping of the normative issues of AI ethics. The analysis revealed that accountability, privacy, and fairness appear in 80% of the principle sets, that industry sets averaged 9.1 principles compared to 10.8 in science sets, and that male authors are over-represented particularly in the technical solution-focused sets.

The third [2] is the most exhaustive of the studies and uses the PRISMA template [16] for their systematic review. The study seeks to understand the groups producing AI principles, whether principles are converging, and if diverging, whether they are reconcilable. It aims to inform governmental and intergovernmental organisations, scientists, research institutions, and funding agencies involved in ethical AI. Citing the absence of a unified database, the authors use a sequential search structure, with first a manual search of four Linkhub webpages, followed by keyword search on Google, and finally manual and complete citation chaining screening of the retrieved principle sets. Inclusion criteria were: being in English, German, Greek, French, and Italian languages, being issued by private and public sector institutions, including AI or similar concepts explicitly in the title, and expressing normative ethical stances. This resulted in 84 AI principle sets being used in the study, with two researchers conducting two manual coding cycles and one code mapping cycle, with additional assistance from specialist ethics researchers. They identified 11 clusters of ethical principles: transparency, justice and fairness, non-maleficence, responsibility, privacy, beneficence, freedom and autonomy, trust, sustainability, dignity, and solidarity. No cluster is common to all principle sets, but the first 5 of the clusters are present in over half the principle sets. Private companies produce the most principle sets, more economically developed countries are over-represented, and most principle sets are targeted towards multiple stakeholder groups. Substantive divergence is identified and grouped into 4 areas, namely how principles are interpreted, why they are important, the issue, domain or actors they refer to, and how to implement them. They view these as a gap at the junction of principle formulation and practical implementation, with implications for research ethics, technology governance, and public policy.

The fourth study [3] is more exhaustive discursively, with Harvard's Berkman Klein Center for Internet and Society releasing a 72 page report on AI principles aiming to map consensus. The report is targeted at scholars and policy makers, and through side-by-side comparison and analysis, seeks to reveal trends in what they propose is "a fractured global conversation". Authors aim to provide a high-level snapshot of current thinking in AI governance, fostering avenues for future research including those drafting new sets of principles, and include visualisations, timelines, and bibliographies as indexes to engage with primary sources. Sampling is employed with the purpose of variety across geography, content, date, and stakeholders with focus on the highly visible and most influential documents. Document search was conducted, primarily not through academic databases, but rather through search engines, cross-citation linking, and personal network recommendations. Legislative and regulatory documents were excluded, along with documents focused on a specific technology; however, sector-specific documents were included. Documents needed to represent the views of organisations as a whole and be authored by senior staff, and languages were limited to Chinese, English, French, German and Spanish. Identified documents were reviewed in team meetings, and if they met conditions, were assigned to an individual for hand-coding, who used the principle's title from the document or, if not available, paraphrased the principle's content. Thus, 36 documents were selected, and after merging the similar principles coded by individuals, 47 principles remained, which were clustered under 8 themes. A key aim of the research was a richly informative spider chart-esque data visualisation, to the extent that it locked down their earlier theme decisions from subsequent changes. The 8 themes established were privacy, accountability, safety and security, transparency and explainability, fairness and non-discrimination, human control of technology, professional responsibility, and promotion of human values, with between 3 to 10 principles listed within each of these themes. As the more recent sets they analysed tended to incorporate all 8 themes, it suggested some convergence around a normative core. The authors also analysed human rights, with 64% of documents containing a reference, and 5 documents using human rights as a framework for their AI principles. Finally, noting that the field is emergent and their sampling being subjective, they expect perspectives to evolve beyond their captured snapshot.

The fifth article [17] uses the analysis of AI principle sets to examine whether policy makers conceptualisation (derived from document analysis) of the term Artificial Intelligence is different to the conceptualisation in research environments (which they established through surveys). Again, the authors started with Algorithm Watch's AI Ethics Guidelines Global Inventory with documents published between 2017-2019, and restricted it to English, resulting in 40 Principle sets used, with three authors manually coding. They found that only 28% of researchers used "human" in their definition, against 62% of the AI Principle sets used by policymakers. Thus, while researchers and policy makers are equally focused on issues such as inequality and discrimination, policy makers appear much more focused on the existential threat of human-like AI, which the authors argue may result in policy makers focusing on longer-term theoretical AI, rather than the pressing ethical issues with existing deployed AI.

Instead of using ethics methods, the sixth article [18] uses risk assessment and risk management techniques from business to create the principles needed for responsibly developing AI. The authors provide 50 self-developed principles grouped under 10 themes, produced from their previous research along with analysis of 30 AI principle sets selected for their diversity, though restricted to English. They then test the coverage of the AI Principle sets against their self-developed principles, to evidence that their consolidated set is more comprehensive, while also being an immediately applicable checklist for businesses to use.

The seventh article [19] argues that AI principles have a high degree of overlap and a number of shortcomings. Interpreted differently by different groups, the principles are highly generalisable and hence hard to put into practice, while also coming into conflict with each other. Examining 7 principles sets, the authors identify 4 key tensions, and argue that focusing on such tensions should drive research questions, bridge the principle to practice gap, acknowledge value differences, highlight where new solutions are required, and identify knowledge gaps and ambiguities.

Table 1
Summary of the systematic studies reviewed.

(a) Number of studied sets, goals, and purpose/output from each study. Goals: Meta-analysis (MA), Coverage Maps (CM), Principle Consolidation (PC), and Hypothesis Validation (HV).

Study	# Sets	MA	CM	PC	HV	Purpose/Output
[4]	9	-	✓	✓	-	Defining 5 principles for laws and technical standards
[15]	22	✓	✓	-	✓	Critical studies analysis on focus, impact and missing areas
[2]	84	✓	✓	✓	-	Defining 11 clusters of agreement and 4 areas of divergence
[3]	36	✓	✓	✓	-	Defining 47 principles under 8 themes for scholars/policy
[17]	40	✓	-	-	✓	Revealing differences between researchers and policy makers
[18]	30	-	✓	✓	✓	Defining 50 principles under 10 themes for businesses
[19]	7	-	-	-	✓	Identifying 4 key tensions between principles to focus on
[20]	27	✓	✓	-	-	Creating a platform for linking and analysing AI principles

(b) Assessing the number of AI principle sets reviewed within each study and the proportion of the studies covering each goal.

	Principle sets per study metrics				
	Min	**Max**	**Med.**	**Avg.**	**S.D.**
Number of Sets	7	84	28.5	31.9	24.1

	Goals per study metrics			
	MA	**CM**	**PC**	**HV**
Fraction of Studies	5/8	6/8	4/8	4/8

The last article [20] takes an explicit multiplicity approach, and rather than adopting one set of AI principles, aims to link them together into a framework where principles can interact with and complement each other. The authors first manually extract 10 general topic terms identified from 27 different AI principle sets, with topics represented by the chosen terms: humanity, collaboration, share, fairness, transparency, privacy, security, safety, accountability, and Artificial General Intelligence/Artificial Super Intelligence. They then calculated topic coverage, which is the number of topics that are contained (i.e. the term is present) in a specific AI principles set. Google's Word2vec [21] trained on a news corpus is used to return the ranked list of words with the closest cosine similarity to each of the topic terms. The result is a wider list of keywords for each topic, where the first semantically different word in the returned list serves as a cutoff point for inclusion. With the new expanded set of keywords on the topic, the coverage of the AI principle sets was again examined, and shown to have expanded, with the level of coverage based ordering of the principle sets also changing, implying that their new coverage mapping is more robust and accurate. Topic frequency by actor groupings showed that governments include security but avoid accountability, and that the private sector includes collaboration but avoids topics such as privacy and security.

2.2. *AI Principles repositories*

In [20], the authors also produce the Linked Artificial Intelligence Principles (LAIP) online platform[3], which actually uses Semantic Web standards (`owl:sameAs`) to represent the semantic linkages between similar AI principles words, and shows original topics, as well as keyword coverage, for 74 of the AI principles sets. Principle set summary cards are shown, with title, publisher, type of publisher, country or region, abbreviation, and date of publication. Keyword-based search functionality is also available, and returns paragraph-level results from respective documents where the keyword is found. Finally, click through functionality is also available from the coverage maps, both to original document sets, and to sentence level listings of the topics or keywords *in situ* within all documents. This richer exploratory functionality is based solely off semantically linking topic words and keywords, but is enough to evidence the value of the Semantic Web approach.

Two other major efforts to catalogue the AI principle sets have been presented, namely the AI Ethical Guidelines Global Inventory[4] launched in April 2019 as a crowd-sourced repository hosted by the German non-profit organisation Algorithm Watch, and the OECD's AI Policy Observatory[5] which launched in

[3]http://www.linking-ai-principles.org
[4]https://inventory.algorithmwatch.org
[5]https://oecd.ai/ai-principles

February 2020. Both of these resources provide search and retrieval functionalities built on relational database concepts. The Ethical Guidelines Global Inventory offers a search field, four drop-downs, and sorting on pre-defined tags leading to hyperlinks to PDFs. The AI Policy Observatory has two distinct classes: the first one contains national AI policies and additional requested text (received by the OECD from member states), and a rich interactive visual interface with analytics dashboards and click through navigation, while the second class includes stakeholder initiatives (e.g. complement of nation states) and contains only tags for stakeholder type, stakeholder name, publishing date, and hyperlinks to the PDFs. Beyond some minimal filtering, the information about the principle sets and ability to compare between them is buried within the individual PDFs.

2.3. Ontologies for Artificial Intelligence

Finally, we discuss work that evidences the successful role that ontologies can also play in the context of AI principles. An early survey on ontologies provided as its first use-case an ontology as a conceptual framework to tie together AI Planning, Decision Theory, and Distributed Systems Theory, to allow these fields to share the research results currently inhibited by differing terms and perspectives on the same underlying ideas [22]. There is a clear parallel with the AI principles proliferation, and the need to explicitly match the same underlying thoughts to the terms used in the documents. Ontologies have been used across multiple domains, from broader work on automating the hypothesis testing within large data repositories of scientific data [23] and modelling the scholarly process in digital humanities [24], to narrower domains such as biomedical investigations [25] and social science systematic literature reviews [26]. Especially pertinent to the AI principles ontology developed herein, are the ontologies supporting systematic reviews in software engineering [27], and in continuous integration of data and hypothesis testing [28]. In short, the continuous re-evaluation of hypotheses can be achieved through setting up data analysis workflows that trigger when new data becomes available. Thus, every time a new AI principles set is added to the knowledge graph, new meta-study results are available via stored SPARQL queries.

The only similar ontological work the authors are aware of is the IEEE P7007 – Ontological Standard for Ethically Driven Robotics and Automation Systems[6], ongoing since 2017 but currently with no publicly available output, and is focused on creating a set of ontologies at different levels of abstraction for ethical design methodologies for robots and automation systems.

Having described the various meta-study and cataloguing efforts for the AI principles, along with the usage of ontologies in other domains, we now lead into the methodology used.

3. Methodology

Given the scope of our work, we use a multi-phase approach, as shown in Figure 1. Using the principle of a life-cycle from software engineering, ontology engineering can be used and broken into the phases of requirements analysis, ontology creation, and ontology assurance [29]. Our methodology is based on those phases.

The first phase incorporates gathering and reading primary AI principle documents and secondary AI principle systematic studies along with identifying AI principle repositories. The second phase incorporates design of the AI principles ontology according to ontology engineering and ontology design methods. A note on terminology is required, as determining AI ethical principles versus more generic AI policy statements and governance initiatives is a difficult task [3]. With a view to extensibility for the ontology to AI principles generally, the aim of the design is thus for extensive coverage of AI ethical principle sets that should also prove sufficient for other AI related documents. Finally, the third phase refers to the assurance and validation of our ontology with the implementation of a proof of concept knowledge graph.

Phase 1 - Requirement Analysis. The primary source of AI principle documents is not academic articles, but rather documents produced by a range of different actors, then published on websites directly or as PDFs available for download from websites. Hence, we did not use academic databases as the predominant source to retrieve the documents, but rather search engines and news articles, along with principle sets previously collected by the authors during prior research, as well as the principle sets that were referred to in the secondary systematic studies. A sampling of these princi-

[6]https://site.ieee.org/sagroups-7007/

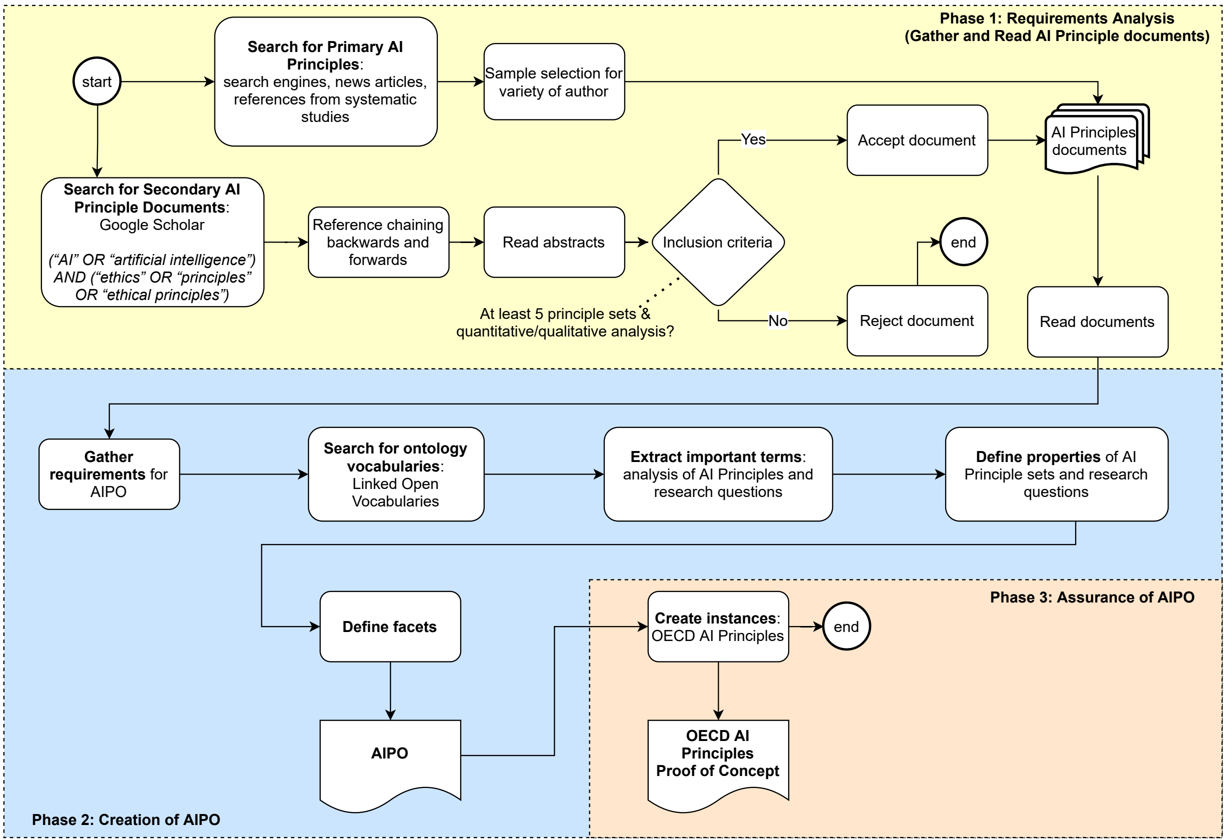

Fig. 1. Our methodology and its phases.

ple sets were read, selected for variety, ranging from corporate (e.g. Google), corporate groups (e.g. Partnership on AI), and professional bodies (e.g. IEEE), to Catholic religious groups (e.g. the Pontifical Academy for Life), supranational organisations such as the European Commission, multinational organisations such as the OECD, and multi-stakeholder initiatives such as the Montreal Declaration for Responsible Development of AI.

The secondary AI principle document search focused on academic articles, with Google Scholar and search terms ("AI" OR "artificial intelligence") AND ("ethics" OR "principles" OR "ethical principles") used. The pool of articles was further increased by reference chaining forwards using Google Scholar, and backwards using citations. Abstracts were read and the articles selected based on them, incorporating at least 5 principle sets and performing quantitative or qualitative analysis on the principles to obtain the research objectives. Note that principle set comparisons are also present in the primary source documents themselves such as the Montreal Declaration, and governmental body briefs such as the European Parliament studies, but these were excluded. Two of the AI principle repositories identified were referred to in systematic studies directly, while the third was identified while sourcing a primary document.

Phase 2 - Creation of AIPO. The following step is to build the AI principles Ontology (AIPO). In designing an ontology, there is not a singular best modelling approach for a domain, the design process is inherently iterative, and the ontology's concepts should be close to the physical or logical objects and relationships in the domain [30]. Further, the use case for the ontology and level of granularity needed should drive design decisions. A seven step ontology design process commonly used is:

Step 1. Determine the domain and scope of the ontology.
Step 2. Consider reusing existing ontologies.
Step 3. Enumerate important terms in the ontology.
Step 4. Define the classes and the class hierarchy.
Step 5. Define the properties of classes-slots.

Step 6. Define the facets of the slots.
Step 7. Create instances.

The first six steps are depicted in Figure 1 under Phase 2 (blue area), while the seventh corresponds to the validation part of our methodology, and can be seen in the same figure under Phase 3 (orange area).

In lieu of accessible human experts, domain knowledge had to be developed by the authors in order to create the AI principles ontology. The first phase of reading primary and secondary sources and analysing the principle repositories was thus necessary to develop this knowledge and gain insights into the requirements for the ontology – consequently defining how best to address the research questions. The Linked Open Vocabularies resource was used to source a number of ontology vocabularies for review that were considered for inclusion in the ontology. Ontology mapping and integration was done manually by the authors without the use of automated tools.

Phase 3 - Assurance of AIPO. Finally, in phase three to ensure the applicability of the ontology, we instantiate a small knowledge graph. For this we chose the OECD's AI principles set as it allowed us to instantiate every component in our ontology. We focus specifically on complex AI principle that relates to other principles in the same set as it allows us to test the validity of our ontology to its full extent.

4. The AI Principles Ontology

The domain and scope of the ontology is already clearly set by the research question, establishing the ontology is meant to allow for the systematic review of AI principles, as described by the four purposes of such studies, and obtain similar results to the meta-studies described above. Additionally, we aim at keeping the ontology as lean and simple as possible. The Linked Open Vocabularies resource[7] was used to identify a number of vocabularies, including SKOS, VOID, DCAT2, Dublin Core, PROV-O, FOAF, Schema.org, ORG, ModSci, and Geonames among others. These ontologies were read and assessed in detail, using ontology mapping and subsequent ontology merging to ensure reuse rather than recreation. As a result, only 3 unique properties were created within the new ontology, namely `:hasPrincipleSet`, `:hasWordCount` and `:hasPageLength`. The full list of reused namespaces is included in Table 2.

Table 2
List of the merged ontologies and their namespaces.

Ontology	Namespace
rdf	⟨http://www.w3.org/1999/02/22-rdf-syntax-ns#⟩
rdfs	⟨http://www.w3.org/2000/01/rdf-schema#⟩
xsd	⟨http://www.w3.org/2001/XMLSchema#⟩
owl	⟨http://www.w3.org/2002/07/owl#⟩
skos	⟨http://www.w3.org/2004/02/skos/core#⟩
dcat	⟨http://www.w3.org/ns/dcat#⟩
foaf	⟨http://xmlns.com/foaf/0.1/⟩
dctype	⟨http://purl.org/dc/dcmitype/⟩
dct	⟨http://purl.org/dc/terms/⟩
modsci	⟨https://w3id.org/skgo/modsci⟩
org	⟨http://www.w3.org/ns/org#⟩

The resulting AIPO (AI Principles Ontology) ontology is visually presented in Figure 2. The important terms identified relate to documents and their metadata, principles and their meanings, agents and their roles. These main concepts drove the creation of classes and their hierarchy, followed by the properties and attributes. Additional classes exist to ensure the merged ontologies are used correctly. Note that the ontology is publicly available and online[8]; all source images can be found under the Assets folder.

At a high level, a `dcat:Resource` is the document itself, with a `:hasPrincipleSet` property linking to `skos:ConceptScheme` to capture the set of principles as a whole idea. The individual principles in `skos:Concept` are linked via `skos:inScheme`. A number of SKOS match properties are used to allow richness of semantic linking, with strings capturing key contexts and examples to give meaning to the principle beyond its title. `skos:relate` was also added, as inter-principle set links between principles was found after testing the ontology with the OECD principles described below. Dublin Core Terms (`dct`) capture the multitude of metadata regarding the document, required for the meta-analytic components of systematic studies. The `dct:subject` property uses `modsci:ArtificialIntelligence` to ensure that a resource is correctly identified when added to the Web of Linked Open Data. No adequate ethics designation was found, but as the ontology is designed for AI principles in general, we decided to leave it out. When implementing the OECD principles, we also identified the need for a `dct:source` property, as they were

[7]https://lov.linkeddata.es/dataset/lov

[8]https://github.com/AndrewHarrison/AIPO

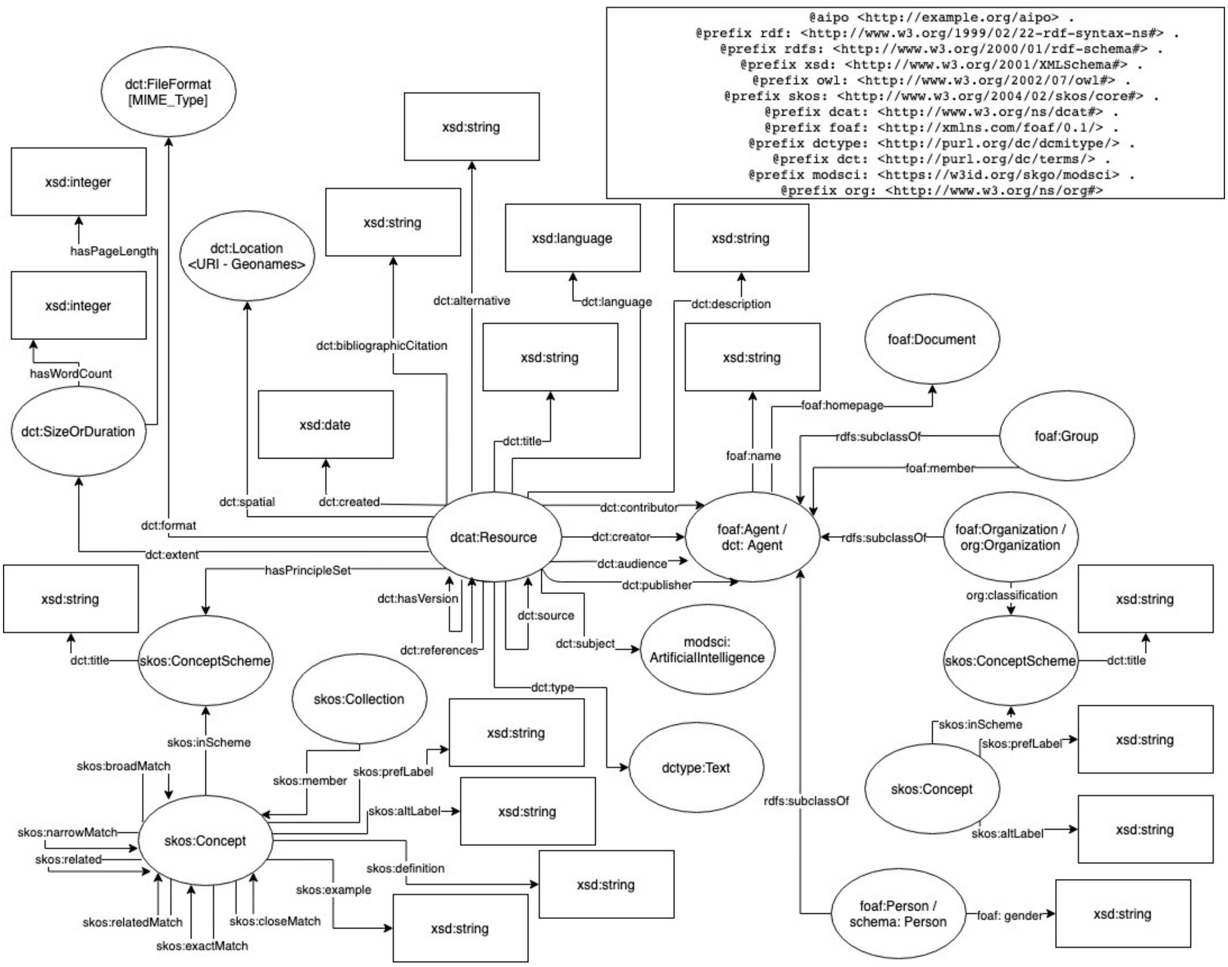

Fig. 2. AIPO - the AI Principles Ontology.

used to develop the G20 AI principles, and this linkage should be recognised. The FOAF ontology provides richness in describing Agents, and due to usage of its properties was the predominant ontology merged in. However, AIPO also captures other ontologies where the entity is the same, in order to better assist integration into Linked Data: in other words, `foaf:Agent`/`dct:Agent`, `foaf:Person`/`schema:Person` and `foaf:Organization`/`org:Organization` are also included. The Org ontology describes relations and hierarchy within an organisation, but no suitable ontology was found for describing the governmental, private sector, civil society distinction needed. The `org:classification` property was therefore used with `skos:ConceptScheme` capturing such categorisation. A full subject-predicate-object tabular view of the ontology is presented in Table 3.

A Proof of Concept with partial instantiation of the OECD principles was done as validation of the ontology, and led to the ontology enhancements described above. This is presented in Figure 3. The example shows strengths of the ontology, such as capturing that accountability (the only principle fully shown) is related to the other four principles (only one is partially shown) via them needing to be implemented as a necessary condition for accountability. Alternate labels for the term "accountability" are also captured, along with a definition and an example that ensures that the context and meaning of the concept "accountability" used by OECD is correctly matching the term "accountability" used in other principle sets. The differing languages the document is available in are also shown, with all of the principles also linkable on the Semantic Web to their translation in other languages (for instance, on ConceptNet [31]). The audience for the doc-

Table 3
The AIPO ontology: tabular view.

Subject	Predicate	Object
skos:Concept	skos:relatedMatch	skos:Concept
skos:Concept	skos:exactMatch	skos:Concept
skos:Concept	skos:closeMatch	skos:Concept
skos:Concept	skos:narrowMatch	skos:Concept
skos:Concept	skos:broadMatch	skos:Concept
skos:Concept	skos:related	skos:Concept
skos:Concept	skos:example	xsd:string
skos:Concept	skos:definition	xsd:string
skos:Concept	skos:altLable	xsd:string
skos:Concept	skos:prefLabel	xsd:string
skos:Collection	skos:member	skos:Concept
skos:Concept	skos:inScheme	skos:ConceptScheme
skos:ConceptScheme	dct:title	xsd:string
dcat:Resource	:hasPrincipleSet	skos:ConceptScheme
dcat:Resource	dct:hasVersion	dcat:Resource
dcat:Resource	dct:references	dcat:Resource
dcat:Resource	dct:type	dctype:Text
dcat:Resource	dct:source	dcat:Resource
dcat:Resource	dct:subject	modsci:ArtificialIntelligence
dcat:Resource	dct:extent	dct:SizeorDuration
dct:SizeorDuration	:hasWordCount	xsd:integer
dct:SizeorDuration	:hasPageLength	xsd:integer
dcat:Resource	dct:format	dct:FileFormat [MIME_Type]
dcat:Resource	dct:spatial	dct:Location ⟨URI-Geonames⟩
dcat:Resource	dct:created	xsd:date
dcat:Resource	dct:bibliographicCitation	xsd:string
dcat:Resource	dct:alternative	xsd:string
dcat:Resource	dct:title	xsd:string
dcat:Resource	dct:language	xsd:language
dcat:Resource	dct:description	xsd:string
dcat:Resource	dct:contributor	foaf:Agent
dcat:Resource	dct:creator	foaf:Agent
dcat:Resource	dct:audience	foaf:Agent
dcat:Resource	dct:publisher	foaf:Agent
foaf:Agent	owl:sameAs	dct:Agent
foaf:Agent	foaf:name	xsd:string
foaf:Agent	foaf:homepage	foaf:Document
foaf:Group	foaf:member	foaf:Agent
foaf:Group	rdfs:subclassOf	foaf:Agent
foaf:Organization	rdfs:subclassOf	foaf:Agent
foaf:Organization	owl:sameAs	org:Organization
foaf:Organization	org:classification	skos:ConceptScheme
foaf:Person	rdfs:subclassOf	foaf:Agent
foaf:Person	foaf:gender	xsd:string
foaf:Person	owl:sameAs	schema:Person

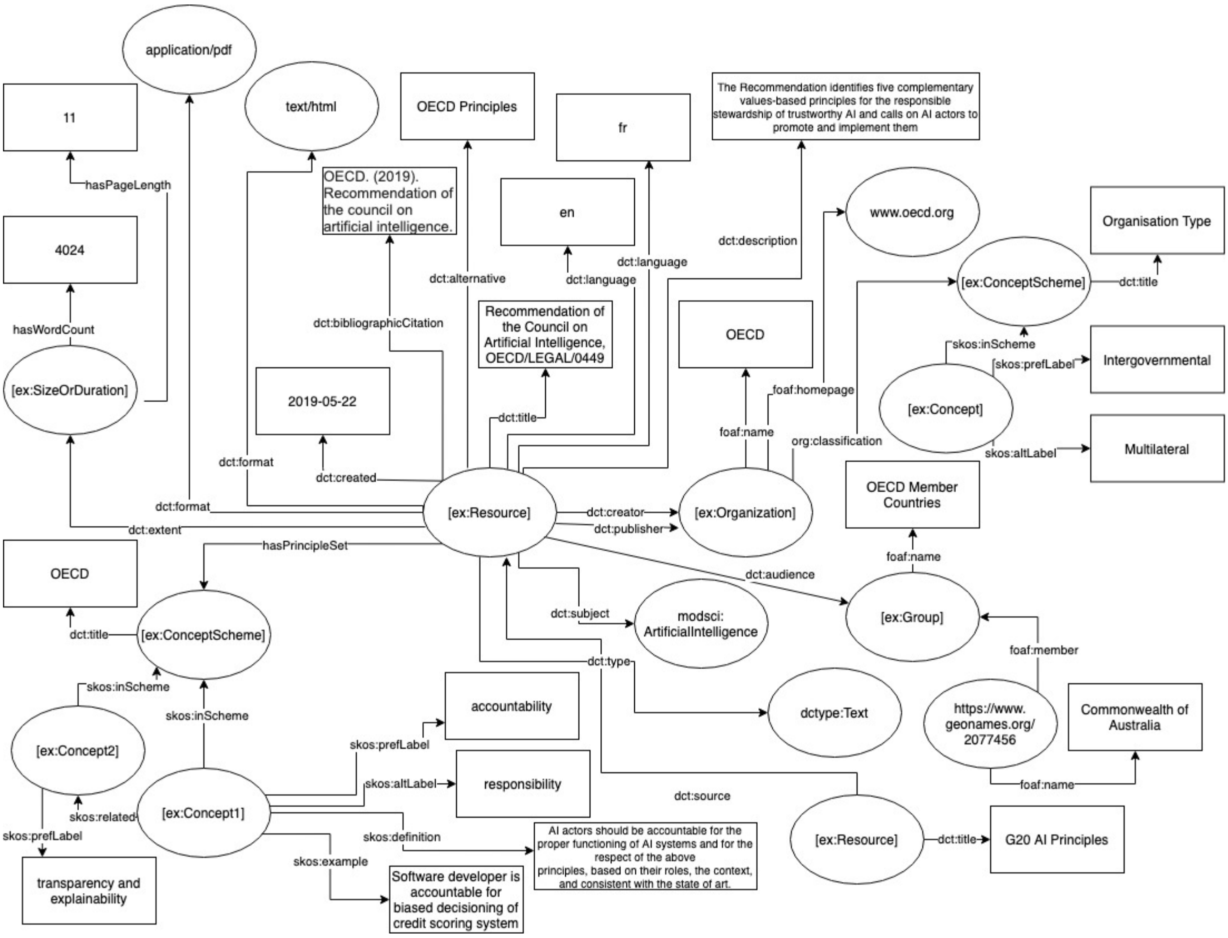

Fig. 3. Proof of Concept of AIPO using OECD principles.

ument, being the 37 OECD member countries can also be captured, with the example given of Australia. Finally, details on the creator and publisher (being the same here) are captured, with an example of the organisation concept scheme showing that it is a intergovernmental organisation, also termed a multilateral organisation.

5. Benefits of AIPO

The ontology and the instantiation of the OECD principles shown in the previous section answer affirmatively our secondary research question, i.e. whether an ontology can be developed to transform AI ethical principles into a Linked Data consumable format. Given that, we turn to the remaining research question, concerning how such an ontology can facilitate the study of AI principle sets and further their impact. Assessing the facilitator value of the ontology includes two steps: first, investigating whether it does obviate steps in the current meta-study process for AI principles; second, whether it does produce the same results available from current meta studies. In order to answer whether the ontology furthers the impact of AI principle sets, we discuss the value added by the use of semantic technologies over relational databases.

5.1. AIPO facilitates meta-studies

Starting with facilitation, the creation of AIPO is a necessary step toward the creation of a knowledge graph containing details from all of the primary principle sets. This will remove the need for using generic search engines, with their high false-positive rate and lack of semantic search that subsequently requires manual review and selection of principle sets from the returned results. Instead, finding AI principle sets can done through the knowledge graph, with search functionality via SPARQL to identify the desired documents. The step of extensive human analysis and manual coding of the different principle sets is also significantly reduced, with principles from the same set and relations to similar/differing principles in other sets already represented in the knowledge graph through SKOS's `skos:ConceptScheme`, `skos:Concept`, and `skos:Collection` classes. The meta-information about the principle set documents, such as created date and location, and the documents' relations to each others (e.g. OECD as a source for G20 AI principles) are provided through various Dublin Core Terms. Finally, FOAF and SKOS provide information on the Agents related to the document, whether they are the audience, creator, contributor etc.

Further to this, the encoding of the information in the knowledge graph allows rich semantic querying of the actual semantic content extracted from within the principle sets, inherently available via SPARQL queries performed on the knowledge graph. Thus to conduct a desired analysis, manual re-coding and re-analysis for every new researcher for every new purpose is not required. The removal of this costly and laborious process reduces a barrier to entry for novel research on AI principles, and also makes publicly available the coding schema used, serving both reproducibility and assurance purposes. Additionally, it is worth noting that the knowledge graph is not a static artefact locked into a quickly dating article, but rather a dynamic living data source, that is already integrated and interrogatable in conjunction with a host of other data sources contained on the Semantic Web.

Efficiency should not be at the loss of effectiveness. The existing systematic studies were grouped into four purposes and presented a series of results. Assessing whether these purposes and results can be derived from querying the knowledge graph is a litmus test for the external validity of the ontology. Firstly, meta-analytic results can be obtained through SPARQL queries that count the triples returned, such as gender of authors, type of organisation, or number of principles. Secondly, the coverage maps can also be derived from the knowledge graph through SPARQL queries, via properties such as `skos:inScheme` and `:hasPrincipleSet`. Further, arbitrarily defined coverage maps using `skos:Collection` via the property `skos:member` can also be created and queried. Thirdly, principle consolidation is built-in by design, with `skos:Concept` representing unique principle entities, and the various SKOS match properties used to comprehensively detail such consolidation. Finally, hypothesis testing while not automated explicitly in the design, can at least be assisted by stored SPARQL queries that can be re-run by researchers to test and re-validate their hypotheses when new data is added to the knowledge graph, or if a Linked Data source is found that provides new bootstrapped reasoning.

Comparing explicitly to the only paper using semantic technologies in this context [20], our use of `skos:exactMatch` removes data merging issues that are introduced by `owl:sameAs`, such as unique information from two different `skos:example` properties being lost when principles are matched using `owl:sameAs`. Further, polysemy and context, with a word such as "school" having multiple meanings (an institution, a physical building, or the cohort of people for instance), is a stated issue in their work. And as the terms and principles used in AI documents, let us take "fairness" as an example, have multi-interpretability depending upon their context, polysemy may be one of the greatest challenges for regulating AI systems [32]. Even when there is some commonality in understanding, drawn out semantic debates and misunderstandings can ensue without contextual definition of each term. Through the use of `skos:altlabel`, `skos:definition`, and `skos:example` in addition to SKOS match properties to explicitly contextualise words, this polysemy issue can instead be addressed. Given the obviation of research steps and equal if not improved effectiveness, the facilitation aspect of the ontology has thus also been answered affirmatively.

5.2. *Impact of AIPO*

Let us now turn toward impact, particularly when comparing AIPO to the existing AI principle repositories.

In deciding to use a semantic representation over a classical database structure, there are two major advantages: 1) the knowledge semantically represented is easily integratable and exchangeable with other sources of knowledge, and 2) further to the explicit knowledge asserted in the semantic specification, implicit knowledge can be deduced [29]. Any of the data sources available on the Linked Data Cloud can be integrated via value chains, with avenues of research more constrained by investigators' imaginations rather than lack of data. Further, organisations such as the BBC are using the Semantic Web to drive their data stories [33], and data stories are growing in importance in journalistic storytelling and in educational media [34]. As the informed participation of stakeholders is required for the development of responsible AI, meaning education and communication to ensure its potential impacts are known to stakeholders, along with making them aware that they can participate in shaping its development within society [32]. Then, the value of an AI principles ontology derived knowledge graph, feeding data into such journalistic and educational initiatives, can directly impact social awareness of AI principles.

Further, researchers using Algorithm Watch's AI Ethics Guidelines Global Inventory when conducting research, as detailed in Section 2.1, evidence that they would also be users of the knowledge graph as a repository too. In addition, the restriction to the English language found in many of the studies, though common to a lot of academic research, is a particular shortcoming for AI Principles as they are ostensibly meant for global application. Though some studies were conducted by multi-lingual researchers thus allowing for additional major languages, principle sets were essentially convenience sampled for the researchers' known languages. But the availability and wide usage of language tags on the Semantic Web, which are not just at document level, but with individual principles able to be linked to the same or similar term across languages, allows principle sets in different languages to be easily interlinked. More concretely, this means that also stakeholders not from Western liberal democracies are both more able to listen to, and to add to, the conversation on AI principles. This is of particular importance, given these stakeholders' susceptibility to AI's misuse by overbearing authorities.

A final valuable impact from the use of Semantic Web technologies pertains to their machine readability. For both digital and physically instantiated entities, the ethically imposed constraints on the systems are in the purview of their designers, who are charged with choosing the functionality which is allowed and the data that is used. Thus programmers and engineers are expected to adhere to ethical guidelines in the development of AI, see as example the ACM guidelines, or IEEE guidelines (both critiqued as short and theoretical [35]). The ACM's code of ethics, introduced in 1972 and not updated since 1992, was finally updated in 2018, presumably in light of news scandals such as Volkswagen's "diesel gate" which cost the company over $30B [36]. Volkswagen's programmers raised objections internally, but there was no whistle-blowing (as companies actively disincentive whistle-blowing [35]), and it thus stands to question the enforceable value of these ethical guidelines. Experimental testing using ethical vignettes from real-world situations, with the experimental group being exposed to the ACM's guidelines have bore out for their lack of effect, while also revealing a positive impact on ethical decisions when ethical issues have hit news cycles [36] (an extra value add to the media's use of knowledge graphs feeding data stories described above). But with machine-readable ethical AI principles, an agreed schema or even legally enforced standard could be conceivably set in the future (see recent EU white-paper [37]). This could be read directly into AI entities, with the presence and implementation of the principles auditable by regulatory authorities, and adherence even standing to lower legal liability and damages apportioned [3] to developers or the owners of AI systems. This ontology work can be viewed as a step towards such schema.

Categorising the current situation, there is a growth in AI principles, growth in systematic analyses of these principles, both second-order (meaning and functioning of ethics) and first-order (ought behaviours and ethical norms) [38], a multitude of applied uses that are also contested, a susceptibility to ethics shopping, and finally a key role for AI principles to play in making safe AI arms races. Given this flurry of activity and its importance, and the scaling and bias impracticalities of manual researcher coding, there is both space and opportunity for the introduction of semantic technology in this context. In summary, the research objective of creating an ontology for AI ethical principles is met, with a number of existing ontology vocabularies used to assist ontology mapping and integration with other linked data sets. We have shown that through using semantic technologies we can structure AI ethical principle sets to facilitate systematic studies, and that this can further the impact of AI principle sets.

6. Conclusion

Major scandals of recent years such as echo chambers, propaganda bots, and fake-news would not have been possible without AI, and in many cases AI ethics is failing, with institutionally driven principle sets serving mainly as a marketing strategy [15]. Systematic reviews can help reveal such shortcomings, for example, the UN Global Pulse's chief data scientist published a letter to the editor in Nature Machine Learning [39], questioning why solidarity was rarely present in AI ethical principles (using data from [2]), while being present in 30% of the world's constitutions. Thus, meta-analysis drives insights, insights drive thoughts and opinions, and these drive discussions between stakeholders that have real impact. For example, academic researchers have been calling for a halt to facial recognition technology until it is regulated [40], citing amongst other things a 35% error rate for dark-skinned women versus 1% for white males. Additional voices have come from the employee activism initiatives (*e.g.*, [15, 35]), and the societal impact is now starting to show. During the same week in early June 2020, IBM announced they would stop selling facial recognition technology, followed by Amazon announcing a moratorium on use of their facial recognition technology for 1 year, and then Microsoft stating it will await US Federal laws on safe deployment before providing facial recognition technology[9].

By structuring AI principle sets using semantic technologies to assist systematic studies, we believe that our AI Principles Ontology can help drive impact. Immediate future work involves expanding the ontology according to best practices [41]. Additionally, a reason for the variance in AI ethics and principles is that the term AI can refer to the computational technology, a field of scientific research, or an autonomous entity itself [32]. This variance is not currently addressed in the AIPO, with the generic subject "Artificial Intelligence" not differentiating between the technology, research fields, and actual entities. Another two components for future analysis are identified, the first regards the phase of AI development, as it can also impact how an AI principle is interpreted. For example, "transparency" and "auditability" mean different things to different stakeholders, during system design and development, as opposed to when the system is deployed and in use within an organisation [18]. The second is the domain of usage and level of impact on targeted groups and individuals, such as with facial recognition, recidivism prediction, and autonomous weapons. Further refinement of the ontology may be required to capture these intricacies. Lastly, while the use of the SKOS ontology gives context and connects similar usages of the same ethical terms, it does not resolve the issue that these moral terms can conflict with each other, nor resolve situations of moral overload/hard choices where agents must make a choice that will violate their values [42]. However, socio-technical attempts to address this normative uncertainty through democratic dissent during AI development (see [43] for a hard choices framework) can make use of the information captured in the knowledge graph, with its accessibility to semantically meaningful SPARQL queries and use of data stories.

In closing, at web scale, local theories can be combined, and knowledge does not have to be consistent, as the goal is not an all-encompassing ontology with universal agreement [44]. When commonality of a concept has not led to commonality of terms, the Semantic Web, through identifying and describing the relationships between similar concepts that are developed independently by different groups, provides the wider common language for communication and collaboration [6]. But the Semantic Web had a slow start, and it is of little practical use if it is not instantiated with suitable amounts of data, for else it remains only in the purview of enthusiastic academics producing sophisticated ontologies [44]. Its power is realised when users create applications that gather diverse information from across the Web, process it, and exchange it with other applications [6]. Thus incentivising researchers and AI principle producers to use the ontology and add to the knowledge graph will be a critical challenge, requiring bootstrapping of an initial viable knowledge graph, and awareness building in the community. The next step is thus to use the ontology created herein to populate a knowledge graph with a core volume of AI principle sets, and connect this to the Linked Data Cloud.

[9] https://www.theverge.com/21288053/microsoft-facial-recognition-police-law-enforcement-pledge-regulation

References

[1] I.J. Good, *Speculations on perceptrons and other automata*, International Business Machines Corporation, 1959.

[2] A. Jobin, M. Ienca and E. Vayena, The global landscape of AI ethics guidelines, *Nature Machine Intelligence* **1**(9) (2019), 389–399.

[3] J. Fjeld, N. Achten, H. Hilligoss, A. Nagy and M. Srikumar, Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI, *Berkman Klein Center Research Publication* (2020).

[4] L. Floridi and J. Cowls, A unified framework of five principles for AI in society, *Harvard Data Science Review* **1**(1) (2019).

[5] J. Priem, Beyond the paper, *Nature* **495**(7442) (2013), 437–440.

[6] T. Berners-Lee, J. Hendler and O. Lassila, The semantic web, *Scientific american* **284**(5) (2001), 34–43.

[7] C. Bizer, T. Heath and T. Berners-Lee, Linked data: The story so far, in: *Semantic services, interoperability and web applications: emerging concepts*, IGI Global, 2011, pp. 205–227.

[8] T. Kuhn and M. Dumontier, Genuine semantic publishing, *Data Science* **1**(1–2) (2017), 139–154.

[9] P. Groth, A. Gibson and J. Velterop, The anatomy of a nanopublication, *Information Services & Use* **30**(1–2) (2010), 51–56.

[10] S. Bechhofer, D. De Roure, M. Gamble, C. Goble and I. Buchan, Research objects: Towards exchange and reuse of digital knowledge, *Nature Precedings* (2010), 1–1.

[11] M. Brundage, S. Avin, J. Clark, H. Toner, P. Eckersley, B. Garfinkel, A. Dafoe, P. Scharre, T. Zeitzoff, B. Filar et al., The malicious use of artificial intelligence: Forecasting, prevention, and mitigation, *arXiv preprint arXiv:1802.07228* (2018).

[12] D. Amodei, C. Olah, J. Steinhardt, P. Christiano, J. Schulman and D. Mané, Concrete problems in AI safety, *arXiv preprint arXiv:1606.06565* (2016).

[13] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman and A. Galstyan, A survey on bias and fairness in machine learning, *arXiv preprint arXiv:1908.09635* (2019).

[14] N. Bostrom and E. Yudkowsky, The ethics of artificial intelligence, *The Cambridge handbook of artificial intelligence* **1** (2014), 316–334.

[15] T. Hagendorff, The ethics of AI ethics: An evaluation of guidelines, *Minds and Machines* (2020), 1–22.

[16] D. Moher, A. Liberati, J. Tetzlaff, D.G. Altman, P. Group et al., Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement, *PLoS med* **6**(7) (2009), e1000097.

[17] P. Krafft, M. Young, M. Katell, K. Huang and G. Bugingo, Defining AI in Policy versus Practice, in: *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 2020, pp. 72–78.

[18] R. Clarke, Principles and business processes for responsible AI, *Computer Law & Security Review* **35**(4) (2019), 410–422.

[19] J. Whittlestone, R. Nyrup, A. Alexandrova and S. Cave, The role and limits of principles in AI ethics: towards a focus on tensions, in: *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 2019, pp. 195–200.

[20] Y. Zeng, E. Lu and C. Huangfu, Linking artificial intelligence principles, *arXiv preprint arXiv:1812.04814* (2018).

[21] T. Mikolov, I. Sutskever, K. Chen, G.S. Corrado and J. Dean, Distributed representations of words and phrases and their compositionality, *Advances in neural information processing systems* **26** (2013), 3111–3119.

[22] M. Uschold, M. Gruninger et al., Ontologies: Principles, methods and applications, *TECHNICAL REPORT-UNIVERSITY OF EDINBURGH ARTIFICIAL INTELLIGENCE APPLICATIONS INSTITUTE AIAI TR* (1996).

[23] Y. Gil, D. Garijo, V. Ratnakar, R. Mayani, R. Adusumilli, H. Boyce and P. Mallick, Automated hypothesis testing with large scientific data repositories, in: *Proceedings of the Fourth Annual Conference on Advances in Cognitive Systems (ACS)*, Vol. 2, 2016, p. 4.

[24] V. Pertsas and P. Constantopoulos, Scholarly Ontology: modelling scholarly practices, *International Journal on Digital Libraries* **18**(3) (2017), 173–190.

[25] A. Bandrowski, R. Brinkman, M. Brochhausen, M.H. Brush, B. Bug, M.C. Chibucos, K. Clancy, M. Courtot, D. Derom, M. Dumontier et al., The ontology for biomedical investigations, *PloS one* **11**(4) (2016), e0154556.

[26] I. Tiddi, D. Balliet and A. ten Teije, Fostering Scientific Meta-analyses with Knowledge Graphs: A Case-Study, in: *European Semantic Web Conference*, Springer, 2020, pp. 287–303.

[27] J.C. de Almeida Biolchini, P.G. Mian, A.C.C. Natali, T.U. Conte and G.H. Travassos, Scientific research ontology to support systematic review in software engineering, *Advanced Engineering Informatics* **21**(2) (2007), 133–151.

[28] Y. Gil, D. Garijo, V. Ratnakar, R. Mayani, R. Adusumilli, H. Boyce, A. Srivastava and P. Mallick, Towards Continuous Scientific Data Analysis and Hypothesis Evolution., in: *AAAI*, 2017, pp. 4406–4414.

[29] P. Hitzler, M. Krotzsch and S. Rudolph, *Foundations of semantic web technologies*, CRC press, 2009.

[30] N.F. Noy, D.L. McGuinness et al., Ontology development 101: A guide to creating your first ontology, Stanford knowledge systems laboratory technical report KSL-01-05 and . . . , 2001.

[31] R. Speer, J. Chin and C. Havasi, Conceptnet 5.5: An open multilingual graph of general knowledge, in: *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 31, 2017.

[32] A. Theodorou and V. Dignum, Towards ethical and socio-legal governance in AI, *Nature Machine Intelligence* **2**(1) (2020), 10–12.

[33] G. Kobilarov, T. Scott, Y. Raimond, S. Oliver, C. Sizemore, M. Smethurst, C. Bizer and R. Lee, Media meets semantic web–how the bbc uses dbpedia and linked data to make connections, in: *European semantic web conference*, Springer, 2009, pp. 723–737.

[34] E. Segel and J. Heer, Narrative visualization: Telling stories with data, *IEEE transactions on visualization and computer graphics* **16**(6) (2010), 1139–1148.

[35] B. Mittelstadt, Principles alone cannot guarantee ethical AI, *Nature Machine Intelligence* **1**(11) (2019), 501–507.

[36] A. McNamara, J. Smith and E. Murphy-Hill, Does ACM's code of ethics change ethical decision making in software development?, in: *Proceedings of the 2018 26th ACM joint meeting on european software engineering conference and symposium on the foundations of software engineering*, 2018, pp. 729–733.

[37] E. Commission et al., White Paper on Artificial Intelligence—A European Approach to Excellence and Trust, *COM (2020)* **65** (2020).

[38] D. Greene, A.L. Hoffmann and L. Stark, Better, nicer, clearer, fairer: A critical assessment of the movement for ethical artificial intelligence and machine learning, in: *Proceedings of the 52nd Hawaii international conference on system sciences*, 2019.

[39] M. Luengo-Oroz, Solidarity should be a core ethical principle of AI, *Nature Machine Intelligence* **1**(11) (2019), 494–494.

[40] K. Crawford, Halt the use of facial-recognition technology until it is regulated., *Nature* **572**(7771) (2019), 565–566.

[41] D. Garijo and M. Poveda-Villalón, Best Practices for Implementing FAIR Vocabularies and Ontologies on the Web, *arXiv preprint arXiv:2003.13084* (2020).

[42] J. Van den Hoven, G.-J. Lokhorst and I. Van de Poel, Engineering and the problem of moral overload, *Science and engineering ethics* **18**(1) (2012), 143–155.

[43] R. Dobbe, T.K. Gilbert and Y. Mintz, Hard Choices in Artificial Intelligence: Addressing Normative Uncertainty through Sociotechnical Commitments, *arXiv preprint arXiv:1911.09005* (2019).
[44] J. Euzenat et al., A map without a legend, *Semantic Web* **11**(1) (2020), 63–68.