Check for updates

SI: Re-Imagining a More Trustworthy Social Media Future

SM+S
social media + society

# A Rights-Based Approach to Trustworthy AI in Social Media

Social Media + Society
July-September 2020: 1–13
© The Author(s) 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2056305120954672
journals.sagepub.com/home/sms
SAGE

**Dave Lewis**[1] **and Joss Moorkens**[2]

**Abstract**

Social media platforms increasingly use powerful artificial intelligence (AI) that are fed by the vast flows of digital content that may be used to analyze user behavior, mental state, and physical context. New forms of AI-generated content and AI-driven virtual agents present new forms of risks in social media use, the harm of which will be difficult to predict. Delivering trustworthy social media will therefore be increasingly predicated on effectively governing the trustworthiness of its AI components. In this article, we examine different approaches to the governance AI and the Big Data processing that drives it being explored. We identify a potential over-reliance on individual rights at the expense of consideration of collective rights. In response, we propose a collective approach to AI data governance grounded in a legal proposal for universal, non-exclusive data ownership right. We use the Institutional Analysis and Development (IAD) framework to explore the relative costs and benefits on stakeholders in two use cases, one focused on digital content consumers the other focused on digital content knowledge workers. Following an analysis that looks at self-regulation and industry-state co-regulation, we propose governance through shared data ownership. In this way, future social media platforms may be able to maintain trust in their use of AI by committing to no datafication without representation.

**Keywords**

trustworthy AI, data ownership, social media, collective bargaining

## Introduction

Social media platforms increasingly use powerful artificial intelligence (AI) that are fed by the vast flows of speech, video, and sensed data that platforms capture to analyze user behavior, mental state, and physical context. New forms of AI-generated content and AI-driven virtual agents that are rapidly being integrated into social media platforms. These may be accompanied by new forms of risks in social media use, the harm of which will be difficult to predict, for example, from manipulation of audio and video content such as deep-fakes, highly persuasive bargaining/sales agents, accurate lie detection or context-sensitive micro-targeting of content. Delivering trustworthy social media will therefore be increasingly predicated on effectively governing the trustworthiness of its AI components.

Some prospective requirements for governing trustworthy AI, such as non-discrimination, respect for privacy, robustness, and safety are already subject to existing regulation or are well aligned with commercial company requirements. These may therefore represent the low hanging fruit for new governance approaches designed to deliver trustworthiness. However, other requirements including accountability, data governance, design-for-all, human oversight, respect for human autonomy, and transparency may conflict with the business motives driving AI use in social media. These will present major objectives in designing effective AI governance regimes and may require new innovative forms of governance. However, many strong trends in the development of AI may impede the successful implementation of new governance approaches. Confounding trends include the accelerating pace of AI innovation, especially for organization with access to vast data stores; the diminishing barriers to entry in terms of AI skills and computing power; the transnational nature of AI-driven service provision; the opaque nature of modern machine learning algorithms; and the immense technical and legal expertise that the large digital platforms providing deploying AI-driven social media can bring to bear on any regulatory conflicts. Regulatory approaches to meeting trustworthy AI requirements may also be impeded politically and socially by its perceived chilling

[1]Trinity College Dublin, Ireland
[2]Dublin City University, Ireland

**Corresponding Author:**
Dave Lewis, ADAPT Centre, Trinity College Dublin, Dublin 2, Ireland.
Email: dave.lewis@adaptcentre.ie

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).

effect on AI innovation, especially as AI capabilities to be seen as a strategic national economic and security asset. AI governance regimes that require centralized state regulation and corresponding AI regulatory expertise will therefore require strong political drive to establish and to scale at a pace that can keep up with AI's accelerating impact on society.

In this article, we examine different approaches to the governance AI and the Big Data processing that drives it being explored. We identify a potential over-reliance on individual rights at the expense of consideration of collective rights. In response, we propose a collective approach to AI data governance grounded in a legal proposal for universal, non-exclusive data ownership right. We use the Institutional Analysis and Development (IAD) framework to explore the relative costs and benefit to actors in two use cases, one focused on digital content consumers the other focused on digital content knowledge workers, and propose further routes for analysis.

## Governance for Trustworthy AI

AI trustworthiness has risen to prominence in public policy discussion from 2017, popularized by the widespread signup of prominent AI practitioners to the Principles developed by the Asilomar conference on Beneficial AI (Future of Life Institute, 2017). Policy work has built to an extent from existing work on accountable algorithms, which was concerned with fairness and bias in AI applications and the role of transparency and accountability to counteract the opaque nature of modern machine learning algorithms such as deep neural networks (Kroll et al., 2016; Mittelstadt et al., 2016).

In the context of social media, an important work is that of Latzer et al. (2016), which analyses the economics and governance options for what they refer to as algorithmic selection technologies, that is, those that mediate and form the digital reality experienced by individuals and by society. This class of technologies have become dominated by AI techniques and underpin the core functions of social media applications, namely search, aggregation, surveillance, predictions, filtering, recommenders, scorers, content generation, and resource allocation. Risks of algorithmic selection are identified as:

- Manipulation of individuals or groups
- Diminishing variety that creates biased views and distortion of reality
- Constraints on communication and freedom of expression
- Threats to privacy and data protection rights
- Social discrimination
- Violation of intellectual property rights
- Impact on the human brain and cognitive capacity
- Algorithmic power over human behavior and development

They outline the prospects for different forms of governance of AI that may be followed to reinforce benefits and mitigate the perceived ethical and societal risks of employing AI technology, namely market-based, self-organization, self-regulation, and co-regulation. Market-based approaches rely on suppliers and customers, especially in consumer markets, applying pressure on organizations to respect their value systems, or risk losing their business. However, this relies on consumers and market partners having a clear understanding of potentially harmful business behavior, whereas may players in the data value chains feeding AI applications operate in obscure data brokerage roles that do not interface directly with consumers. Also, as large digital platforms benefiting from network effects often occupy near monopoly positions there is a high barrier for exit for consumers.

Self-organization involves an organization setting itself principles or standards to follow in developing and deploying AI solutions. These may be administered through internal ethics boards that must be staff with appropriately qualified personnel and sufficiently empowered by senior management to make recommendations that may come into conflict with existing business objectives. Several large digital platform companies such as Microsoft and Google already have such boards in place. However, they suffer from limits to the transparency in their decision-making, as they may need to deal with commercially sensitive information or plans.

Self-regulation involves the development of codes of conduct, industry standards, quality seals and certification bodies, ombudsman schemes, and ethic committees across an industry sector. Self-regulation works better in mature sectors characterized by players with a shared outlook on maintaining public and market trust. The Institute of Electrical and Electronic Engineers (IEEE) has commenced as number of initiatives aimed at improving professional practice and trustworthiness in AI technology (Adamson et al., 2019). Significant among these IEEE initiatives has been the Global Initiative on Ethics of Autonomous and Intelligent Systems, which has conducted a wide-ranging iterative expert consultation on Ethically Aligned Design (EAD), the First Edition of which was launched in February 2019 (The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, 2019). This sets out general principles for ethically aligned design, highlights a range of issues, and makes recommendations for organization governance, professional practice, government policy, and further research. Many of these issues are now being addressed by the P7000 series of IEEE standards, to provide specific standards for organizations, ranging from process models for ethical design, guidelines on transparency, privacy, bias, and well-being, and specific domains for ethical AI including child data, nudging, trustworthy news, and facial analysis. Industry ethical standards rarely involve independent enforcement mechanisms and have been criticized as an anti-competitive means for larger incumbents to raise barriers to new market entrants (Calo, 2017).

These limitations of self-regulation therefore give rise to proposal for co-regulation, where industry behavior is impacted by statutory regulation and other government actions such as tax incentives, state funding for research and mitigation measures, and government procurement policy. Such state action has to balance the realization of the potential economic and societal benefits of AI with the anticipated risks. Several governments have started to produce policy documents about the trade-offs necessary between realizing the benefits of AI. A 2018 review of European national and transnational regulatory proposals identified policy activity (Access Now, 2018). It conducted a gap analysis against criteria related to transparency, accountability, privacy, freedom of conscience and expression, equality and non-discrimination, due process in law, data protection and user control, collective rights for free press and election, economic impact and the future of work and AI in weapons. It identifies the European Union's (EU) move to use an ethical grounding to differentiate Europe's approach to AI, contrasting with the laissez-faire approach of the United States and centralized government-driven approach of China, building on the existing commitment to European-wide regulation already implemented via the General Data Protection Regulation (GDPR) (EU, 2016), but identifies several gaps AI policy across Europe. Of relevance to AI in social media, gaps are identified in terms of the disparity between standards to which public and private organizations are held, especially given the dominant role of the private sector in developing AI technology. This public–private disparity also identified as an issue in relation to big data and AI ethics in a US study (Metcalf et al., 2016). The comparative analysis of EU policy work also highlights relevant lack of policy outside of Germany related to freedom of conscience and expression, both areas heavily impacted by search bubbles and personalized selection of social media news feeds. The profiling of individual is also a highly relevant area, given the inherent power of AI in pattern matching and the vast amounts of personal behavioral data accessible via social media. Though European data protection authorities can interpret such inferred data as personal data, potential for bias exists outside the remit of GDPR when AI application profile places or other non-personal characteristics rather than people directly. Gaps are also identified in the purpose limitations of data protection in the face of AI's power in identifying previously unanticipated patterns, such as those inherent in large social media data set. Furthermore, the collective rights to free press and elections are not address in any of the analyzed national policies, despite these being area now being heavily impacted by social media. Attention to the impact of AI on the future of work is identified across the analyzed policies, but primarily in the form of pragmatic of guidance on retraining, and without reference to labor rights. This analysis of European AI policy proposals identifies some important general issues in relation to AI ethics:

- A general "wait and see" attitude is prevalent, driven in part by a desire to not to stifle AI innovation and investment. It is also in part driven by uncertainty arising from the rapid progress of AI and therefore the potential for technical solution to emerge (i.e., relying on market-driven measure), which may quickly render legislation redundant.
- A lack of consideration of collective issues in identifying and mitigating the risks of AI, rather than relying solely on individual rights and data protection frameworks.
- The relationship between ethics and legislation, where ethical guidelines may prove useful in reducing harm while appropriate regulation is developed or where the adoption of guidelines simply avoids addressing the difficult issues of regulation, or as pointed out in Calo (2017), can be used by large entities to project a solution without submitting to external enforcement.

These general trends seem to persist in the most recent detailed governmental work on trustworthy AI. The EU has recently established a set of ethical guidelines for achieving trustworthy AI that aims to foster research, reflection, and discussion, but which is positioned as a complement to possible future EU regulation of AI technology (European Commission, 2019). This presents ethical AI principles derived from the EU Charter of Fundamental rights, but consistent with the Access Now review, it focused on the fundamental rights assigned to individuals in relation to chapters on dignity, freedom, and justice but says little in relation to collective rights under the chapter of solidarity including the rights of workers for information, communication, and collective bargaining (Official Journal of the European Communities, 2018). Similarly, recent recommendation on AI from the Organisation for Economic Co-operation and Development (OECD) lays out a similar set of principles that should be followed in the development and use of AI systems (OECD/LEGAL/0449, 2019). However, though it recommends that government should review and adapt policies related to innovation and competition of trustworthy AI, it does not make any concrete proposals on the form of regulation. There have been some proposals for possible co-regulatory structures such as new national regulatory bodies responsible for the governance of big data and its use for algorithms (Tutt, 2016) as well as internal ethics boards that may help organizations implement best practice (Calo, 2013; Polonetsky et al., 2015). However, there seem to be few concrete plans in place to establish the public regulatory bodies and standards needed to implement co-regulation. This is in part due to uncertainty about the direction and impact of AI technology, but also in part due to the lack of relevant technical expertise available to public policy bodies in comparison with the large digital platforms and other major private adopters of AI (Calo, 2017).

In summary, while the risks of applying AI to social media data may be profound, market-based, self-organization, self-regulation, and co-regulation solutions to AI governance all face challenges in balancing the rapid development of complex AI technology in service of maximizing shareholder value and realization of the societal benefit of AI, with the need to track and mitigate the societal risks the stream of AI innovation present. In the next section, we therefore examine an alternative strand of AI governance proposals that takes a more collective approach by those involved in the provision of data that drives AI, rather than requiring external inspection of algorithmic behavior.

## Collective Governance Approaches for AI Trustworthiness

Another, earlier, thread of research that has had a visible influence on analysis of trustworthy AI is work on the ethics of Big Data, especially in relation to privacy and data protection concerns (Metcalf et al., 2016; Mittelstadt & Floridi, 2016; Richards & King, 2014; Zwitter, 2014). A common theme here is the criticism of the notice and consent mode of personal data self-management that is prevalent as a requirement of most data protection regulation. Privacy researchers already understood that individuals were poorly equipped to fully understand and anticipate the impact of personal data usage to which they asked to consent (Solove & Washington, 2012). This had led to the assessment that the predominant current model of making use of personal data based on initial consent may not be ethically justified and therefore requires some *collective* rather than individual engagement with organizations doing so (Sax, 2016). Big Data analytics, which was later enhanced by rapidly improving AI capabilities in pattern matching, increasingly delivers societal level impacts, rendering individual consent an even less effective mechanism for governing negative consequences of data processing. Different approaches have been proposed to rebalance of this asymmetry of information about the implications of personal data processing. In Rahwan (2018), the case is made for AI governance with societal level engagement through a society-in-the-loop model. This consists of a number of mechanisms including crowd-sourcing in the articulation of values related to AI, introduction of professional algorithm auditors and a new class of algorithm monitoring software, but acknowledges the need for new public institution needed to make this a reality. An alternative suggestion is to treat personal data as a valuable resource shared between the individuals providing it and the organizations collecting and using the data (Tene & Polonetsky, 2013). However, while in the EU such a feature has been included as the right to portability of GDPR (De Hert et al., 2017), it is unclear whether this will result in any compensating realizable value accruing to an individual at a level that would outweigh the transaction cost the individual incurs in managing the additional data sharing. More significantly without an enforceable legal mechanism for realizing the value of data, any transactions would be performed on terms dictated by the company capturing the data in the first place.

## Governance Based on Shared Data Ownership

The problems constraining the leverage of data supply in the governance of AI is a symptom of a broader problem identified in EU policy deliberations on facilitating data sharing to enhance the data-driven economy (European Commission, 2018) related to the legal transaction costs of sharing data. This derives from the complex and multifaceted legal nature of data as identified in report on data ownership rights in the EU (Van Asbroeck et al., 2017a). Data ownership rights to the extent they exist are therefore grounded in a complex range of different legislation, and it is this complexity that impedes the development of data sharing agreements, especially when the respective current and future value of data to the sharing parties is not well understood. There are some limited ownership rights to data through copyright when a level of creativity or originality can be demonstrated, and in Europe through the Database Directive is sufficient effort in the structuring of data can be demonstrated to claim sui generis rights. GDPR offer some specific rights, but only for personal data, and not in the form of familiar ownership rights that can be traded or bequeathed to descendants (the data portability right offers some small move in this direction). Companies may also exert ownership-like rights over information in relation to patented inventions and through trade secrets.

One proposal resulting from this analysis proposes a new EU data ownership right (Van Asbroeck et al., 2017b). This proposal suggests the non-exclusive ownership on any data for any party that can show a traceability log of having produced or contributed to the data. This is combined with an obligation on such right-holders to share data on fair, reasonable and non-discriminatory (FRAND) terms provided that this does not conflict with the normal exploitation of the data or if it unreasonably prejudices the legitimate interests of a right-holder. The non-exclusive nature of the ownership right in this proposal reflects the non-rivalrous, non-exclusive and inexhaustible nature of data and the objective to facilitate, rather than restricts its sharing and reuse. The traceability requirement provides a simple but clear opt in mechanism that can be applied to data at different scales, from an individual piece of data, such as a "like" on a social media post or a sentence translation used to train a machine translation facility, up to entire data sets. Technical standards are already available for capturing such trace logs in an interoperable format, while new distributed ledger technologies offer secure mean for sharing immutable trace logs, greatly simplifying the implementation of the traceability requirement.

Another critical analysis of the data property in the EU highlights the need for transparency in identifying to whom data property rights belong and how this may be confounded by the claims of more than one person, for example, in cases of DNA, group photos, or social media consumer behavior profiles (Purtova, 2017). This means that suggestions related to collective management of data rights, such as the notion of collective consent outlined in the study by Bygrave and Schartum (2009), may struggle to define the boundaries of the group being represented at any point in time. Separate to considerations of data ownership, Mantelero (2016) not only identifies the dynamic boundary of groups in considering data protection concerns of AI, but also the lack of awareness that groups members would have of each other or of shared risks of group discrimination, for example, of low-income worker driving unsocial hour being grouped with young people driving for socializing. This motivates a proposal for data protection regulators, rather than representative of specific group to play a role in conducting mandatory ex-ante multi-stakeholder risk assessment of any new group profiling activities. This however seems to be beyond the current remit of regulators, at least under GDPR in Europe, and would have to overcome governments' reluctance and limited technical ability to intervene systematically in commercial AI innovation activities.

Data are often referred to as the fuel on which the effectiveness of AI is based, and the impact of any AI application grows in step with the relevance and volume of data available to train the AI and to be processed by it. Our contention is therefore that the stakeholders involved in the sourcing of data for an AI system are often the most representative of those impacted by that AI system. Further if data sources act collectively, they may use their key role in data supply as strong leverage in negotiates that balance between commercially-driven AI data processing design and the reduction of risks both to the individuals providing data and to other connected stakeholders with whom they empathize, for example, family, friends, groups with specific, sometime minority, characteristics. By grounding representation broadly on data ownership, we hope representation will extend beyond those subject to the specific regulation of data protection, but other groups including knowledge workers, who also have a stake in monitoring and mitigating harms from data gathering and processing by AI.

Therefore, we propose that the obligation to share data under FRAND terms in the data ownership proposal from (Van Asbroeck et al., 2017a) be made conditional in data sharing contracts on safeguarding measures to deliver trustworthy AI. Such conditionality provides an opportunity for new forms of model contracts to be developed that can demand the implementation of trustworthy AI requirements prior to data be shared for the training of or processing by the AI system in question. This approach would enshrine trustworthy AI requirements into the legal contracts that provide access to the data, rather than relying on external regulation of the AI system during its development and operation. Therefore, the agreement on implementing trustworthy AI requirements works in concert with the flow of data that drives AI innovation and commercialization, rather than acting a separate regulatory system that may serve to impede innovation. Implementation of such a data ownership right would therefore introduce "shared data ownership" as a alternative means for governing trustworthy AI as a complement to market-based, self-organization, self-regulation and co-regulation approaches reviewed above.

However, as discussed above in relation to data protection, individuals are at severe disadvantages in agreeing terms and conditions for sharing data with organizations due to the asymmetry in the technical understanding of data processing and the asymmetry of power due to the relative value of the individual's data compared with that offered by an organization processing data at scale. This would also therefore put individuals' exercising data ownership rights at a disadvantage when agreeing trustworthy AI safeguarding terms in a data-sharing agreement. While the simple nature of the new proposed data owner right coupled with the practices developed in legislating for intelligible personal data sharing contracts under GDPR may help alleviate the asymmetry of understanding, it does not address the asymmetry in the value being exchanged, that is, individuals would have very little leverage in negotiating strong trustworthy requirements with organizations already providing value from large amounts of information from other sources. However, data right-holders may be able to redress this power asymmetry by *acting collectively*.

Analyzing mechanisms for encouraging and supporting collective action in negotiating shared data ownership therefore requires a shift in how we consider data sharing. A move from considering the sharing of data as a transfer of value to considering it as the management of a shared resource may provide a better grounding for considering the role of all the relevant stakeholders in the governance of data flows and, consequently, of the AI systems that depend on these flows.

By considering the process of developing and deploying AI systems as part of a shared data resource management activity, we are able to compare different approaches to governing this activity through the Institutional Analysis and Development (IAD) framework (Ostrom, 2011). The IAD framework is a tool that assists in designing the institutional structures for the sustainable governance of complex shared resource. The IAD can help build a shared understanding of resource dynamics, the diverse interests within the community of users and producers, and the costs and benefits of different governance structures. It can therefore also be used in comparing different institutional design for a shared resource governance setting. While originally developed for analyzing the governance of share physical resources such as fisheries, it has also been found application in the analysis of shared digital resources, referred to as knowledge commons (Hess & Ostrom, 2007). Such institutional analyses have

been used to analyze new forms of digital resource collection, aggregation, and processing, for example, image and geospatial data (Alvarez León, 2016). A common-resource pool approach inspired by Ostrom's theories is identified as a way to focus debate on the governance of data processing for AI onto the social and political dilemmas it presents (Taylor & Purtova, 2019).

We use the IAD to compare patterns of interaction and likely outcomes from potential AI governance approaches based on self-regulation, co-regulation, and shared data ownership. We do this for two action settings related to two classes of AI data-producing stakeholders: (1) digital content consumers whose behavior is profiled by AI to target further content to them, and (2) knowledge workers whose processing of data in a knowledge-rich task contributes to the development for AI to automate of that task. An example of the first action setting could be users of a video streaming platform such as YouTube, where profiles of viewing behavior guide the recommendation of subsequent video content to users. An example of the second action setting could be online customer service agents whose text chat interaction with customers together with customer service system interaction is logged to provide data for automated customer service chatbot. However, in this article, we will focus on the example of translators engaged in post-editing of machine translation output to improve its qualify, the result of which is recorded as parallel text used to further train machine translation AI, for example, in the form of statistical or neutral machine translation.

## IAD Comparison of Governance Approaches

The IAD framework is centered on the "action arena" where actors make decisions within an action situation constrained by the nature of the resource being managed, the community that use and produce the resource, and the existing rules in operation within the community. This results in specific patterns of interactions between community actors and resources that in turn lead to outcomes that deliver net costs and benefits to the actors involved.

We follow the elaborated guidelines for applying the IAD framework from Ostrom (2011), by identifying the actors involved, the positions they hold within each action situation, the actions they perform, and the rules under which the action situations operate. We also follow the approach taken in Strandburg et al. (2017), which adapts IAD to the study specifically of knowledge commons, and proposed not separating the analysis of patterns of interaction from that of outcomes, as in knowledge generating setting they are closely intertwined. To compare the three different governance approaches, we first examine the existing rule-in-use for each of the two use cases and then assess the pattern of interaction as they currently operate, comparing the cost and benefits to actors in relation to the OECD principles under the three different governance approaches.

### *Characteristics*

The characteristics in terms of the resources, actors involved in their creation and use, the roles they undertake, and the actions involved are summarized in Figure 1.

In the video content streaming use case, a video content consumer avails of the services of a video content streaming service provider. We assume the provider follows an advertising-led business model similar to YouTube, rather than a subscription-based business model. An individual acting as a video content consumer will access the service via a personal account which may provide the video streaming service provider with personal and contextual data including age, gender, location, and preferred language. Recommendation for videos to watch will be made as the result of a query or as a follow-on activity once the consumer has finished watching a video. When the consumer selects a video, it is made available with advertisements integrated into the stream. The revenue from such advertising forms the core business model. Advertisers pay to have specific advertisements inserted into video watched by people with desired characteristics and can receive detailed information on the number of impressions and click-through for advertisements against that specification. Deep learning models are developed that make video recommendations based on personal and contextual information of the consumer together with their video search and watching history (Covington et al., 2016). Creators of videos may post them and receive data on impressions, and when the content proves popular they may receive a share of the advertisement revenue attracted by that post. The shared resources in this use case are therefore the posted videos, the personal, contextual, and search and viewing history data of consumers and the AI recommender system that targets both video and the embedded advertisement to individual consumers.

In the post-editing of machine translation (PEMT) use case, a Language Service Provider (LSP) is contracted by a translation client to translate content from a source language to one or more target languages. Commercial translation (e.g., translation of content from instruction manuals, web sites, product descriptions, etc.) are typically contracted on a per-word basis. Discounts can be negotiated for leverage of translation memories (TM), that is, databases held by the client of source and target sentence pairs resulting from previous translation projects. We focus here on the common case where translators are hired on a freelance basis by an LSP to work on specific translation project (though larger LSPs and governmental translation agencies may retain full-time translator cohorts). Translators use Computer Aided Translation (CAT) tools that offer related translation suggestions from TMs where available. In this case, the proposed target language text is edited by the translator to achieve the required level of accuracy and fluency, rather than creating it from scratch. This process is known as post-editing. Increasingly, proposed translation from TMs are supplemented by proposals generated by machine translation

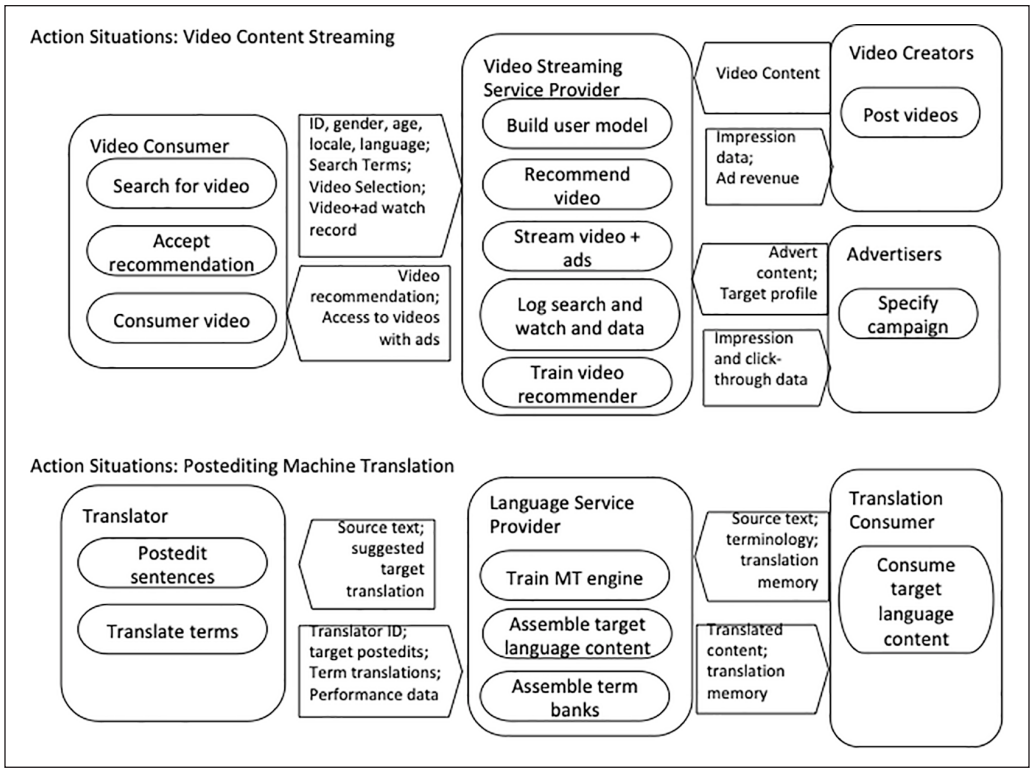


**Figure 1.** Actors, role, resource flows, and action for each use case.

(MT), which is therefore referred to as PEMT. MT suggestions are generated by statistical MT or, increasingly neural-MT using deep neural network techniques. The data needed to train these MT models may be derived from client TMs, from previous translations where copyright was constrained, or from public sources of parallel data. CAT tools may also integrate support for identifying client-specified terminology and using specific translations of those to improve the quality and consistency of translations. The shared resources in this use case are therefore the translated source-target language pairs contributed by individual post-editors and aggregated in TMs, term-bases containing the terms required by a client and their translation into target languages, and MT AI-models developed from particular client domains and language pairs.

## *Rules-in-Use Analysis*

To analyze the rule in use in the two use cases, we can use the classification of rule-in-use provided in Ostrom (2011):

- Boundary rules that determine how someone is accepted into the group of appropriators of the commons resources being considered;
- Position rules that determine that someone can undertaken a specific task or decision-making responsibility in the group;
- Scope rules that determine any aspects of resources that cannot be accessed;
- Choice rule that enforce or prohibit specific forms or configurations of harvesting technology;
- Aggregation rule related to permissions or agreements from others needed to harvest a resource;
- Information rule related to the confidentiality or publishing of information; and
- Payoff rules related to the sanctions for breaking other rules and the mechanisms by which they are monitored and enforced.

## *Rules-in-Use: Video Streaming Use Case*

Boundary rules for the video streaming use case are very permissive in terms of video consumers, video producers, and advertisers, but highly restrictive for the Video Streaming Service Provider (VSSP). This reflects the network effect of social media applied to video sharing, which promotes the dominance of the provider that hosts the most videos by maintaining the lowest barrier of entry for consumers, producers, and advertisers. Membership of the latter are only restricted in cases of contravention by producers and advertisers of social norms or, in some jurisdictions legal restrictions, on content, for example, that involves harm to minors, extreme or incitement to violence and so on, which would be defined in the service's terms of service.

Position rules are very restrictive, as apart from external regulation, decision-making positions are assigned only within the VSSP. Beyond content breaching terms of use, scope rules are administered within the VSSP, with regional and language restrictions to content being one of the features of the video recommendation AI. Choice rules are also governed by the VSSP and mainly address the consent permissions needed from consumers, creators, and advertisers with respect to personal data, for example, as stipulated in GDPR for EU residents. Similarly, aggregation rules for all members are controlled by the VSSP as specified in the terms of use, with consumers, producers, and advertisers restricted to accessing resources only though their relevant interfaces, that is, user interfaces or application programming interfaces (APIs). Information rules are also set by the VSSP, primarily to discharge its legal data protection duties, but also to ensure confidentiality of commercial data exchanges with advertisers necessary for any online platform with multiple paying customers. Payoff rules are therefore largely enforced by the Video Storage and. Processing Platform (VSPP), with the sanction of removing accounts of consumers, producers, or advertisers who to breach the terms of use. The VSSP is only subject to sanctions where strong external regulatory enforcement is in place, for example, with GDPR. The only sanction for breaches by the VSSP is for consumers, producers, and advertisers to stop using the service, but with a near monopoly position arising from strong network effects, these are extremely limited.

### *Rules-in-Use: Post-Editing Machine Translation*

We have previous applied the IAD framework to analyze possible institutional structures to treating translations as a commons resource to better balance the relative positions of post-editing translators and LSPs adopting machine translation (Moorkens & Lewis, 2019). This analysis suggested that a future commons approach administered via professional translator organizations could improve the sustainability of the translation industry by improving redistributional equity in translation value chains through royalties arising from resource contributions to MT services. This also offered the potential to reduce the legal risk exposed by current ambiguous copyright status and to offer a resource- anchored hub for improved mutual professional support activities as the translation industry and its knowledge professionals transition to an MT-centered value chain model. Several challenges to a commons-based approach remain however. Boundary rules are difficult to determine as many translators work on a contract basis outside of the structures of national professional bodies and there is a large number of small LSPs operating in the market. The lack of strong representative bodies for both translators and the large body of small LSPs means that positions of decision making related to language resource appropriation and use in training MT systems resides with management roles in the larger LSPs or translator clients that amass large collections of TMs and have the capacity to develop MT engines or to subcontract such activities. Consequently, scope rules for appropriation of language resources and MT are dictated by these managers based on competitive and commercial motivations. This includes the motivation to restrict access to language resources and MT service for competitors and also to restrict access to resources representing content that may contain confidential client information, for example, from translation of pre-release public content or internal reports, or commercially valuable content, for example, from translation of paid-for content. These restrictions should also be reflected in the choices used in appropriating resources; once a translation is in the public domain, the reuse of it in TMs and for those with technical capabilities to develop MT engines is widespread. It is, for example, not uncommon for individual translators to retain their own translations in TM for reuse on personal CAT tools, and for LSPs to retain their clients' TMs, and to use them once immediate confidentiality or commercial sensitivity issues have passed. This is also undertaken through scraping of published translations to reform the parallel translation data in a form suitable for MT. Therefore, though copyright does enable protection of source and target text, the sampling nature of their use in both TM and MT means that the barriers to appropriation of publicly available translated content are low and may be made without reference to the contributor. The widespread reuse of TMs in CAT tools has resulted in little distinction being drawn to the use of these resources for machine translation. Therefore, aggregation rules for appropriating language resources for MT are loose, despite the fact that the scope for applying the resultant MT engine may be much wider that what would be possible with a given TM, which rarely yields much reuse benefit outside of its subject matter domain. Thus, those with the technical ability to harvest translated content from a wide range of sources and use them to develop MT engines benefit from the same relaxed rules that are accepted as the norm across the long tail of those using elements of that resource for TMs. While some translated content may be subject to confidentiality constraints, in general meta-data associated with such content, for example, the identity of translators, post-editors, and translation quality assessors, is not shared with appropriators. This is in part to avoid the need to address personal data regulations, which is also a concern with performance monitoring meta-data such as key-logging, which can act as a biometric to re-identify post-editors (Lewis et al., 2017). This meta-data information rule means, however, that appropriation is extremely limited in acknowledging the contributions of individuals. In the face of limited employment of copyright protection of translation language resources when applied to MT, there are few sanctions placed on their appropriation beyond those in place in individual translation service contracts, which typically favor LSPs and their clients over individual translators.

## Trustworthy AI Interaction Patterns, Cost, and Benefits

Given the characteristics and rules in use for the Video Content Streaming and PEMT use case, we now expand on the interaction patterns and associated costs and benefits. We analyze the expected costs and benefits experienced by the different actors involved in AI governance under self-regulation, co-regulation, and shared data ownership models.

We view interactions related to governance through the generic classes of rights suggested in Hess and Ostrom (2007) for analyzing the access and governance of shared resources. This includes the rights of individuals to access, contribute, and extract value from commons resources and to remove resources that one has contributed. Further rights address the design of institutional structures and decision-making. These are the management right to regulate the patterns resource use and participate in its improvement of this resources; exclusion rights to who has rights to access, contribute, extract value, and remove resources and how those rights can be transferred; and alienation rights to sell or lease management and exclusion rights.

## Video Streaming Use Case Patterns of Interaction

Considering first the video streaming use case. For the consumer, the benefits come from free access to a large collection of video content and recommendations about content that is considered likely of interest. However, the costs, primarily in the risks of algorithmic selection, are identified in Latzer et al. (2016). The algorithmic power over behavior wielded by the VSSP via recommendation may be so well tuned to the consumers interests and impulses that they spend much more time consuming videos that they would without recommendations. The corresponding increased exposure to the targeting of adverts in videos adds a risk of manipulation in terms of purchases or opinion forming to the costs borne by consumers. Also, the prominence of recommendation and tailored search results in the user interface also raises risks of diminished variety that can lead to the formation of biased views. While this effect of individual world view my not impose a major cost on individuals, it does present a cost at a societal level when the consumer group grows large enough. The ability to target video content and adverts accurately to individual consumers is obviously a benefit for the VSPP and the advertiser in pursuing the shared goal of maximizing the impressions and click-through of advertisements. Increasing consumers' watching time contributes to these benefits and is also of benefit to the content provider who wishes to maximize views even when they do not secure sufficient to benefit from the advertisement revenue sharing. This conflict between the cost of algorithmic selection of videos to watch on consumers, individually and collectively, and the benefits of maximizing such viewing for the VSPP, advertisers, and content creators represents therefore the main arena where governance regimes can strive to achieve a sustainable balance between actors. This current regime is best characterized as self-organization.

We first considering a self-regulation regime where commercial actors collaborate to form and enforce governance regimes with the aim to better balance the benefits between actors. The most prominent self-regulation effort in the AI area is the Partnership on AI which consists of many of the major actors, including Amazon, Apple, Facebook, Google, IBM, and Microsoft. This body has demonstrated some shared intrinsic motivation of its member to self-regulate. This is visible through the definition of high-level principles of behavior in employing AI, inclusion of a wide range of non-commercial partners and initiation of studies to develop rules and possible liability mechanisms that can keep pace with the accelerating innovation and uncertainty. However, an analysis in Leenders et al. (2019) shows its members have not all conducted themselves in adherence to these principles, for example, in seeking to apply AI technologies to lethal weapon system. Furthermore, the Partnership on AI has also not yet demonstrated a mechanism for monitoring or enforcing the adherence of members to its principles and the transparency of such a process needed to establish confidence in it externally. This highlights that while video content consumers and creators benefit from rights to access, contribute, extract value, and remove their individual contributions to the shared pool, they have no say in how pattern of resource use are managed nor how individual resource rights are assigned. This means therefore that consumers have no influence on managing the collective manifestations of risk and are unable to exert control on how monitored behavior data (e.g., search and watch logs) are used in machine learning of group profiles use in search results and recommendations.

Co-regulation requires some decisions be made on which constraints on operation and governance are laid down in law, which are the responsibility of regulators, for example, data protection authorities, and which are left to VSSP and advertisers. As above, consumers and creators already benefit from individual rights to access, contribute, extract value, and remove content due to GDPR. It is the state that makes the political decisions of where the rights to manage, exclude, and alienate right over the data used in AI reside, acting on behalf of the collective of consumers and creators providing the data. The example of the social media regulation now being considered in the United Kingdom (Great Britain, HM Government, 2019) may indicate that government will be primarily motivate to regulate in areas where there are obvious real-world harm coming from AI data processing, for example, in child protection and extremism, and that the approach should be more focused on the simple control of content deemed unacceptable rather than the more complex use of data in driving algorithmic content selection. This approach is both sensitive to reacting to harms that impact public opinion over ones that are less visible to the public, such as bias or profiling. Furthermore, undertaking the

research to identify the existence, scope, and degree of such harm must be undertaken by the state or civic society, which may be relatively poorly resources and less expert than the digital platforms which may offer counter arguments. Furthermore, setting governance rule at a national level to protect the broadest population of users also means regulation has to be considered in terms of its applicability to all VSSPs, not just the largest. This will likely lead to variation in regulations between nations that reflect the relative influence of different market players in that polity, raising the prospect of regulatory shopping by multinationals which further increases their impact on government decisions on management, exclusion, and alienation rule setting.

We contrast this with a shared data ownership approach to data governance. While this still would need to reflect the opinion of the group providing the data supply, that is, video consumers and creators, negotiations on management, exclusion, and alienation will be naturally more focused on the needs of the users than general public opinion. It also means that the emergent perception of new harms can be acted upon responsively based on the leverage of shared data ownership (for example by threatening to share data collectively with advertisers), and does not need to rely on the national political cycle to influence shifts in policy, or to take into account the needs of all VSPP. A mutually beneficial aspect of this arrangement is that part of negotiations could include funding from the VSPP to the representative of the collective data owners to commission research into emerging harms. This means such research can be much more targeted and timely, benefiting both the consumers and creators to better inform them in renegotiations about rules, but also for the VSPP is they are made aware earlier of unanticipated harms and can negotiate mutually agreeable safeguards with its immediate stakeholder, rather than being subjected to safeguards designed to accommodate a much wider range of VSPPs and the general public. The national legislation required for this approach is the fairly simple proposal for non-exclusive, universal data ownership rights and protections to ensure data owners cannot be impeded by, in this case VSPP from organizing collective representation when negotiating on management, exclusion, and alienation rights. This latter legislation may be analogous to labor rights present in many countries to support collective bargaining with employers. This relative simplicity means that this approach may be more easily adopted internationally than attempting to harmonize harm-driven co-regulation rules that are more likely subject to national variations. This has the advantage for consumers and creators that they could potentially organize collective representation at an international level, increasing bargaining power with multinational VSSPs. Conversely, this offers the VSPP a potentially much reduced set of engagement points for negotiating and implementing trustworthy AI governance rule, reducing the cost while simultaneously improve trustworthiness in its products.

### *Post-Editing of Machine Translation Use Case Patterns of Interaction*

In this use case, the potential harms of AI are more relevant to the loss of intellectual property already experienced by translators who are poorly positioned as individuals to sign work contracts that retain the copyright claims to which they would otherwise be entitled. Translators therefore have few rights to access, extract value, or remove the data they provide to the action situation. They also are unable to influence rights held by LSPs and translation clients to manage, exclude, and alienate these rights. The power of AI to influence their behavior in professional practice is also a potential harm as the role of translators becomes more one of correcting and improving the performance of MT engines. A more subtle harm may be upon the consumer of MT in terms of loss of variety due to translations being mediated by a few dominant MT engines and losing the context and experience-related variation from a population of human translators. Translation consumers may also suffer constraints in communication where MT is used in language pairs for which the language data resources available are of low quality or even non-existent relative to more high-demand language pairs such as, for example, French-to-English.

Reducing these harms through self-regulation does not currently seem to be a likely prospect as the population of LSPs is characterized by a very long tail of small providers, most operating to very tight margins. The industry is therefore poorly placed to bear the administrative costs of setting, monitoring, and enforcing regulations, with industry bodies such as the Globalization & Localization Association insufficiently resourced to take such action.

Co-regulation is also unlikely to address the harms experienced by translators. Despite translation being essential to international trade and travel and the provision of public services in multilingual nations, translators are a relatively small portion of the knowledge workforce that has already experienced widespread casualization. They are therefore unlikely to benefit from direct government regulatory support, though they may benefit from the growing demand to protect the much larger population of less skilled casualized workers employed through digital platforms, for example, Uber drivers. There may be more opportunity for the harms to MT consumers to be addressed by co-regulation to address language issues where digital exclusion concerns meet multicultural objectives. After lobbying on the problem of uneven language resource availability for MT in the EU, the European Parliament (2018) in 2018 passed a resolution to actively address language equality in the digital arena. One result is continued support for efforts to collect publicly funded TMs from across Europe to develop MT services to be offered by the European Community in a wider range of language pairs. Although these MT services are made available freely to public administrations across Europe, the data suppliers have no rights over how the data are applied or reused by the Commission.

Shared Data Ownership may offer, therefore, the best route for translators to address the harms of the current system. The proposal in this article would seemingly offer them perhaps fewer rights than the exclusive right to their translation offered by copyright; it is that exclusivity that forces by LSPs and their clients to contractually obtain the copyright to publish the output of paid translations. A non-exclusive data ownership right does not interfere with the publishing of translated content is the same way. It may, therefore, be a more robust basis for securing safeguards and benefit from the use of their work in MT, especially as the AI training of MT consumes the data in fragments in a way that is as yet untested in the protection it gains from copyright. The proposed protection of collective bargaining rights, which are broad enough to support any form of data supplying population, may offer more opportunity for translators to negotiate better terms than labor collective bargaining protections that are of limited use for contractors to the small LSPs typical in the translation industry.

## Conclusions

In this article, we have compared the costs and benefits to stakeholders in two use cases across self-regulation, industry-state co-regulation and new proposal for governance through shared data ownership. Data are the fuel on which the effectiveness of AI is based, and as we see most pointedly in social media, the societal impact of any AI application grows in step with the relevance and volume of data available. The stakeholders that supply the data for an AI system, for example, the people and businesses using social media platforms, are often those most impacted by that AI. Implementing trustworthy AI requirements at the point of data supply may therefore reflect stakeholder needs more accurately, respond to emergent harms more rapidly and scale better with the number and data density of data supplying stakeholders involved, compared with enacting governance rules after the data are acquired and the AI starts impacting people's lives, which is the norm for co-regulation and self-regulation.

The necessity in this approach to engage with data-owning stakeholders at the beginning of the process may in fact help AI development teams not only take those stakeholder views into account with respect to trustworthiness safeguards, but may also promote best practice in engaging with users during the design process (Doherty & Doherty, 2018). A further benefit of a collective shared data-ownership approach to AI governance is that the involvement of data supplying stakeholders scales naturally with the increase in impact of AI applications and the risks that exposes. The more people who are the source of data to an AI system and the more data they provide, the bigger their collective ownership stake in the core data component feeding the system. This serves to balance the power currently accruing unchecked to those gathering data and developing AI systems with it. It may also help to empower the growing proportion of those employed by digital platforms who perform knowledge tasks such as translation and annotation that are needed to train and improve AI models. While some analysts confidently predict the AI revolution will create as many jobs as are lost to AI automation, the experiences of translators indicate much of this will be precarious as manual tasks are highly monitored and systemized to reliably feed AI model. These new jobs will require new protections if AI is to avoid a net decrease in the quality of work. Governance of AI through shared data ownership that applies to the fruit of knowledge work may offer one route of providing such protections in concert with AI innovation.

Many challenges remain to advance any shared data ownership solution to the governance challenges of achieving trustworthy AI. The political will to introduce a non-exclusive universal data ownership right has to be found, and potential conflicts with other areas of legislation need to be researched and overcome. Furthermore, while in the social media domain at least, the mechanisms for data supplying stakeholders to network and organize should be readily accessible through social media itself, the legal protections to allow virtual assembly and negotiation with digital platform unimpeded need to be developed. The EU charter of fundamental rights (Official Journal of the European Communities, 2018) may also offer a legal mitigation to this problem if, under the solidarity chapter, the traceability logs that confer ownership rights over data could also be ascribed as a form of work if that data are used in developing and operating an AI system. Specifically, in this event the workers' right to information and consultation within the undertaking (Art 28) could provide data rightholders with fundamental rights for securing transparency and accountability related to the processing of their data by an AI system. Furthermore, and perhaps more significantly, the right of collective bargaining and action (Art 29) could mean that the all rights-holders of data feeding an AI system have the fundamental right to collectively negotiate the conditions of the data sharing agreement, including representation to monitor the transparent and accountable implementation of trustworthy AI requirements. An organization collecting data for AI development could even be compelled to enable collective action by its data source rights-holders and to negotiate with their legitimate representative. We should also not underestimate the desire of digital platforms to find sustainable governance mechanisms to maintain public trust in their services and indeed to maintain the trust of employees in their business models. Also, this approach may allow for more transnational coordination for a specific service while simultaneously allowing more freedom for experimentation with different safeguarding designs between collectives and digital platforms. Open exchange of experiences, similar to that taken by trade unions internationally, may drive more rapid convergence to best practice that balances the societal risk of AI

with the economic imperative to innovation with AI than would be possible through self- or co-regulation alone.

This novel proposal for shared data ownership as the basis for AI governance requires much further research and analysis by legal scholars, ethicists, and practitioners. Models of representation and decision-making in data supplier collectives are needed, though current practices emerging in data cooperatives[1] and data trusts[2] may offer important insights. Our institutional analysis does however point to the potential for negotiating trustworthy AI safeguards at a point where the collective needs of relevant data supply stakeholders can be contractually aligned contractually with AI innovation in a targeted and responsive fashion. In this way, future social media platforms may be able to maintain trust in their use of AI by committing to no datafication without representation.

## Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This article is supported in part by the ADAPT Centre for Digital Content Technology, which is funded under the SFI Research Centres Programme (Grant 13/RC/2106) and is co-funded under the European Regional Development Fund.

## ORCID iDs

Dave Lewis https://orcid.org/0000-0002-3503-4644
Joss Moorkens https://orcid.org/0000-0003-0766-0071

## Notes

1. https://datacommons.coop/
2. https://theodi.org/article/what-is-a-data-trust/

## References

Access Now. (2018). *Mapping regulatory proposals for artificial intelligence in Europe*. https://www.accessnow.org/mapping-regulatory-proposals-AI-in-EU

Adamson, G., Havens, J. C., & Chatila, R. (2019). Designing a value-driven future for ethical autonomous and intelligent systems. *Proceedings of the IEEE*, *107*(3), 518–525. https://doi.org/10.1109/JPROC.2018.2884923

Alvarez León, L. F. (2016). Property regimes and the commodification of geographic information: An examination of Google street view. *Big Data & Society*, 3(2). https://doi.org/10.1177/2053951716637885

Bygrave, L., & Schartum, D. (2009). Consent, proportionality and collective power. In S. Gutwirth, Y. Poullet, P. De Hert, C. de Terwangne, & S. Nouwt (Eds.), *Reinventing data protection?* (pp. 157–173). Springer. https://doi.org/10.1016/j.clsr.2010.07.003

Calo, R. (2013). Consumer subject review boards: A thought experiment. *Stanford Law Review Online*, *66*, 97–102. https://review.law.stanford.edu/wp-content/uploads/sites/3/2016/08/Calo.pdf

Calo, R. (2017, August 8). Artificial intelligence policy: A primer and roadmap. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3015350

Covington, P., Adams, J., & Sargin, E. (2016, September). *Deep neural networks for YouTube recommendations* [Paper presentation]. RecSys'16 Proceedings of the 10th ACM Conference on Recommender Systems, Boston, MA, United States. https://doi.org/10.1145/2959100.2959190

De Hert, P., Papakonstantinou, V., Malgieri, G., Beslay, L., & Sanchez, I. (2017). The right to data portability in the GDPR: Towards user-centric interoperability of digital services. *Computer Law & Security Review*, *34*, 193–203. https://doi.org/10.1016/j.clsr.2017.10.003

Doherty, K., & Doherty, G. (2018). Engagement in HCI. *ACM Computing Surveys*, *51*(5), 1–39. https://doi.org/10.1145/3234149

European Commission. (2018). *EU Commission: Towards a common European data space, COM(2018)*.

European Commission. (2019, April). *Ethics guidelines for trustworthy AI*. https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top

European Parliament. (2018). *Language equality in the digital age*.

European Union. (2016). Regulation 2016/679 of the European parliament and the Council of the European Union (General Data Protection Regulation). *Official Journal of the European Communities*, *2014*, 1–88.

Future of Life Institute. (2017). *Asilomar AI principles*. https://futureoflife.org/ai-principles/

Great Britain, HM Government. (2019). *Online harms white paper*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/793360/Online_Harms_White_Paper.pdf

Hess, C., & Ostrom, E. (2007). A framework for analyzing the knowledge commons. In E. Ostrom & C. Hess (Eds.), *Understanding knowledge as a commons: From theory to practice* (pp. 1–54). Massachusetts Institute of Technology Press.

The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. (2019). *Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems*. https://doi.org/10.1109/MCS.2018.2810458

Kroll, J. A., Huey, J., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2016, March). Accountable algorithms. *University of Pennsylvania Law Review*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2765268

Latzer, M., Hollnbuchner, K., Just, N., & Saurwein, F. (2016). The economics of algorithmic selection on the Internet. In J. M. Bauer & M. Latzer (Eds.), *Handbook on the economics of the Internet* (pp. 395–425). https://doi.org/10.4337/9780857939852.00028

Leenders, G., Human, B., Apr, I. M., Source, F., & Policy, I. P. (2019, April 13). The regulation of artificial intelligence—A case study of the partnership on AI. *NikolaNews*, 1–26.

Lewis, D., Moorkens, J., & Fatema, K. (2017, April). *Integrating the management of personal data protection and open science with research ethics* [Paper presentation]. First Workshop on Ethics in Natural Language Processing, Valencia, Spain.

Mantelero, A. (2016). Personal data for decisional purposes in the age of analytics: From an individual to a collective dimension of data protection. *Computer Law & Security Review*, *32*(2), 238–255. https://doi.org/10.1016/j.clsr.2016.01.014

Metcalf, J., Keller, E. F., & Boyd, D. (2016). *Perspectives on big data, ethics, and society*. The Council for Big Data, Ethics, and Society. https://bdes.datasociety.net/wp-content/uploads/2016/05/Perspectives-on-Big-Data.pdf

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2). https://doi.org/10.1177/2053951716679679

Mittelstadt, B. D., & Floridi, L. (2016). The ethics of big data: Current and foreseeable issues in biomedical contexts. *Science and Engineering Ethics*, *22*(2), 303–341. https://doi.org/10.1007/s11948-015-9652-2

Moorkens, J., & Lewis, D. (2019). Research questions and a proposal for the future governance of translation data. *Journal of Specialised Translation*, *32*, 2–25.

Official Journal of the European Communities. (2018). Charter of Fundamental Rights (EU) (2000/C 364/01). In L. Schintler & C. McNeely (Eds.), *Encyclopedia of big data* (pp. 1–3). https://doi.org/10.1007/978-3-319-32001-4_35-1

Organisation for Economic Co-operation and Development/LEGAL/0449. (2019). *Recommendation of the council on artificial intelligence*.

Ostrom, E. (2011). Background on the institutional analysis and development framework. *The Policy Studies Journal*, *39*(1), 7–27. https://doi.org/10.1111/j.1541-0072.2010.00394.x

Polonetsky, J., Tene, O., & Jerome, J. (2015). Beyond the common rule: Ethical structures for data research in non-academic settings. *Colorado Technology Law Journal*, *13*, 333–368.

Purtova, N. (2017). Do property rights in personal data make sense after the big data turn? *Journal of Law and Economic Regulation*, *10*(2), 1–26.

Rahwan, I. (2018). Society-in-the-loop: Programming the algorithmic social contract. *Ethics and Information Technology*, *20*(1), 5–14. https://doi.org/10.1007/s10676-017-9430-8

Richards, N. M., & King, J. H. (2014). Big data ethics. *Wake Forest Law Review*, *49*(2013), 393–432. https://doi.org/10.1177/2053951714559253

Sax, M. (2016). Big data: Finders keepers, losers weepers? *Ethics and Information Technology*, *18*(1), 25–31. https://doi.org/10.1007/s10676-016-9394-0

Solove, D. J., & Washington, G. (2012, November 4). Privacy self-management and the consent dilemma. *Harvard Law Review*. http://papers.ssrn.com/abstract=2171018

Strandburg, K. J., Frischmann, B. M., & Madison, M. J. (2017). Knowledge commons and the road to medical commons. In K. Strandburg, B. Frischmann, & M. Madison (Eds.), *Governing medical knowledge commons* (pp. 1–8). Cambridge University Press. https://doi.org/10.1017/9781316544587.001

Taylor, L., & Purtova, N. (2019). What is responsible and sustainable data science? *Big Data & Society*, 6(2). https://doi.org/10.1177/2053951719858114

Tene, O., & Polonetsky, J. (2013). Big data for all: Privacy and user control in the age of analytics. *Northwestern Journal of Technology and Intellectual Property*, *11*, 239–272. http://ssrn.com/abstract=2149364

Tutt, A. (2016, March 15). An FDA for algorithms. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2747994

Van Asbroeck, B., Debussche, J., & César, J. (2017a, January). Building the European data economy data ownership. *Bird & Bird*. https://sites-twobirds.vuture.net/1/773/uploads/white-paper-ownership-of-data-(final).PDF

Van Asbroeck, B., Debussche, J., & Cesar, J. (2017b, March). Data ownership—A new EU right in data. *Bird & Bird*. https://www.twobirds.com/en/news/articles/2017/global/data-ownership-a-new-eu-right-in-data

Zwitter, A. (2014). Big data ethics. *Big Data & Society*, 1(2). https://doi.org/10.1177/2053951714559253

## Author Biographies

Dave Lewis (PhD, University College London) is an associate professor at Trinity College Dublin, School of Computer Science and Statistics and deputy director of the ADAPT SFI Research Centre for Digital Content Technology. His research interests include international standards for ethical and trustworthy AI and big data governance, knowledge models for compliance with data protection regulation and innovation ethics for digital technologies.

Joss Moorkens (PhD, Dublin City University) is an assistant professor at the School of Applied Language and Intercultural Studies and a researcher at the ADAPT Centre and the Centre for Translation and Textual Studies. His research interests include translation technology, translation ethics, translation quality assessment, translator precarity, and translation technology standards.