The role of the epistemic communities in the 'constitutionalization' of the Internet Governance: the case of the EU High-Level Expert Group on Artificial Intelligence.

Nicola Palladino[1]

Abstract
In the last few years, digital constitutionalism has emerged as a novel Internet Governance approach aiming at ordering and limiting the exercise of power by both states and private operators, as well as promoting people's control over digital technology development. Although digital constitutionalism is encountering growing popularity, the prevalent non-binding nature of its initiatives, the discrepancy between the jurisdictional border and transnational digital processes, and the technological embedment of governance mechanisms have hindered its effectiveness. This paper identifies epistemic communities as a crucial factor for the digital technologies' constitutionalization since they have the necessary technical expertise and policy commitment to translate normative provisions into architectural design.
In order to illustrate this nexus, the article performs an in-depth investigation of the case of the European Commission High-Level Expert Group on Artificial Intelligence (HLEG-AI). The analysis revealed that the HLEG-AI effectively gathered an epistemic community strongly committed to fundamental rights promotion and capable of influencing the following policy-making activities of the European Commission, as well as of other non-state stakeholders, whose involvement is essential in order to embed digital constitutionalism principles within the design of AI systems.

Keywords: Digital Constitutionalism, Internet Governance, Artificial Intelligence, Epistemic community, Data Ethics, Human Rights by Design.

1. Introduction

The development and deployment of the Internet and new digital technologies raises several constitutional concerns. The more digital technologies become crucial in our everyday life and social relations, the more they could empower people and foster human wellbeing, as well as endanger an increasing number of fundamental rights.
Since the beginning, Internet governance has been characterized by the opposition between neo-liberalist and sovereigntist approaches (Mueller 2010, Goldsmith and Wu 2006, Kurbalija 2016, Hofmann 2007). The former promoted private self-regulation in order to foster technological innovation and development. However, it also favored the concentration of a huge and almost unrestricted power in the hands of a few corporations. The latter claimed states' authority in cyberspace in order to safeguard national interest, but often ultimately led

[1] Internet and Communication Policy Centre, Department of Political and Social Studies of the University of Salerno, Via Giovanni Paolo II, 132 - 84084 Fisciano (SA), Italy. npalladino@unisa.it; nicolapalaldino@gmail.com.

to security and surveillance policies. Despite these differences, both approaches eventually put in danger the integrity and safety of people and communities, by subverting the potential of digital communication technologies (Zittrain 2008, Jørgensen 2011, De Gregorio and Radu 2020).

Neither multistakeholderism, a compromise between the two based on the dialogue and cooperation among different stakeholders has proven to be able to address these concerns, remaining a fuzzy approach characterized by bland procedural requirements, reproducing already existing power asymmetries and unable to achieve an effective normative framework (Palladino and Santaniello 2020, Carr 2015, Doria 2014).

However, since the Nineties, a further approach, namely "digital constitutionalism", has emerged. It consists of numerous attempts to establish a universal set of Internet-related governance principles and standards for the protection of human rights, aiming at ordering and limiting the exercise of power by both states and private operators, as well as promoting people's control over digital technology development (Redeker et al. 2018, Zalnieriute 2015).

Digital constitutionalism does not necessary require taking the form of traditional constitutional law; rather it consists of the development of widely accepted principles, instruments, tools and practices capable of promoting fundamental rights protection and empowering net-citizens.

Although digital constitutionalism is encountering growing popularity and leads to some relevant outcomes[2], several factors hinder its effective affirmation.

First, for the vast majority, digital constitutionalism initiatives consist of non-binding documents, being drafted by civil society organizations, multi-stakeholder fora, private actors or international organizations.

As noted, "the general lack of binding instruments among interventions in digital constitutionalism is a common source of skepticism regarding the effectiveness of their impact" and "exposes digital constitutionalism to the risk of being turned into a rhetorical device" (Padovani and Santaniello 2018: 4). Although in the last few years, we can witness the increasing number of legislative initiatives following a digital constitutionalism approach (Santaniello et al. 2018), the translation of digital constitutionalism principles and value into binding legislative acts remain a quite limited, and most of the time a cumbersome, phenomenon[3].

Even more challenging difficulties for the affirmation of common rules ordering and limiting the power in the Internet Governance field are due to the transnational nature of the digital processes.

In many regards, the Internet and digital technologies represent one of the more interesting cases of transnational governance (Rosenau and Czempiel 1992, Fuchs 2008, Rioux and Fontaine-Skronski 2015), that is "new modes" of governance (Bäckstrand et al. 2010; Koenig-Archibugi and Zürn 2006; Risse 2006), concerning cross-border processes and phenomena, in which authoritative state-centric control is replaced by more horizontal and decentralized form of steering or coordination systematically including non-state actors.

The transnational nature of the Internet and new digital technology governance challenges the constitutional perspective in many ways.

In general terms transnational governance, blurring the boundaries of jurisdictions as well as that between public and private powers, make state constitutions no longer capable to "regulate

[2] I refer to Parliamentarian initiatives such as the Brazilian Marco Civil, the Italian Declaration on Internet Rights, the Magna Carta for Philippine Internet Freedom, New Zeland Internet Rights and Freedom Bill, Nigerian Digital Rights and Freedom Bill. Also the European Union legislation (De Gregorio 2020) and the activities of UN, OECD, CoE, AU are increasingly moving toward a digital constitutionalism paradigm.

[3] For example, of the Parlamentarian initiatives quoted in note 1, only the Brazialian Marco Civil has been defenively approved.

the totality of governance in a comprehensive way" (Peters, 2006:5 80, se also Grimm 2010), undermining fundamental rights' claim of universality.
Then, also the formalization of digital constitutionalism principles within national or international law risk not being a definitive solution.

Moreover, the Science Technologies and Society approach has highlighted how most of the Internet and digital technology governance take place within day-to-day, operational, yet politically salient practices that keep the Internet and digital technologies functioning (Musiani et al. 2016). It deals with an infrastructural and often invisible layer informing "the adoption and (re)appropriation by users, regulation, and organizational forms" of digital networks (De Nardis and Musiani 2016:6), which is, in its turn, shaped by coordination of a myriad of operators, standard-setting processes, infrastructural design, and operative routine (Hofmann 2017, van Eaten and Mueller 2013).
Although the capability of states to control the Internet and digital technologies has been reconsidered in the light of increasingly common national "cyber-sovereignty" strategies (Goldsmith and WU 2006, Nye 2011), the transborder and distributed nature of digital technologies operation and interactions challenges the effectiveness of national law.
Further, since many technical and public functions in the digital realm are carried out by transnational private companies (DeNardis 2014), they could not be directly affected by international law and international human right law, without calling into question "horizontal" or "third party" effects of international agreements (Teubner 2017, Gardbaum 2008).
It is worth noting that, also when escaping national law, digital networks are not unregulated spaces: the realm of liberty and freedom imagined by Internet founding fathers. Their infrastructural and invisible layer are a means of regulation and control in itself.
Regardless of whether they are shaped by governments, private corporations or unintentional interactions, software and hardware architectures are "law" in itself (Lessig 2006), in the sense that they control and regulate people's behavior, by tracing and identifying, as well as allowing or inhibiting certain courses of action. Digital technology design could then affect humans' integrity and autonomy and its "invisible work" (Stars 1999, DeNardis and Musiani 2016) could make fundamental rights violations hard to perceive and then to be remedied (e.g. the international community has become aware of NSA mass surveillance programs only because of an insider's disclosures, see Laprise 2016).

This work argues that the action of an epistemic community strongly committed to digital constitutionalism principles could overcome these difficulties and contribute to "constitutionalize" the governance of digital technologies.
To this purpose, in the next paragraph I provide a theoretical foundation for this argument by combining Gunther Teubner's societal constitutionalism theory, Science Technologies and Society approach and Haas' work on epistemic communities, highlighting how epistemic communities are pivotal elements to embed digital constitutionalism principles and values into digital code and architectures.
In the following paragraph, I illustrate the role of an epistemic community in the constitutionalizing Internet and digital technology governance, through an in-depth analysis of the work of the High-Level Expert Group on Artificial Intelligence (HLEG-AI), established by the European Commission in 2018.
This case appears particularly favorable to observe an epistemic community promoting a digital constitutionalism framework.
Artificial intelligence is a novel and complex technical issue in its early-stage; conditions that seem to maximize epistemic community influence. Moreover, the topic has raised a lively debate around human rights implications of this technology, and members of the HLEG-AI

have been selected also to carry out the task to draft AI ethical guidelines. So, the European call is likely to have gathered experts concerned with fundamental rights promotion.
Thus an empirical analysis has been performed combining both qualitative and quantitative methods in order: 1) to verify if the set of experts composing the HLEG-AI effectively correspond to one or more epistemic communities and to explore its composition and belief system; 2) to verify if the HLEG-AI effectively gave rise to a convergence of expectations around digital constitutionalism views; and 3) to explore how and to what extent the HLEG-AI, as an epistemic community, contributed to the spreading of digital constitutionalism principles and values.

2. Digital constitutionalism and Epistemic Community: a necessary nexus?

Gunther Teubner's societal constitutional theory could provide useful insights to re-frame the digital constitutionalism issue in order to overcome the above mentioned difficulties.
Building upon the systems theory (Luhmann 2004, Sciulli 1992) Teubner conceives constitutionalization as a process related to the dynamics of differentiation among social sub-systems. In this view, the more a social sub-system becomes autonomous, the more it develops its own systemic logic, based on a specific "communicative media". This is the universal value allowing interaction within the subsystem (power for politics, money for economy, knowledge for science, and so on). The more a subsystem become crucial for the whole society, the more it gives rise to "expansionistic" and "totalizing" tendencies, subordinating other social spheres to its logic in order to guarantee its self-preservation and self-reproduction (Teubner 2011, 2012). As noted, "the integrity of institutions, persons and individuals" are endangered by "anonymous matrix of an autonomized communicative media", resulting in "institutions, discourses, systems" (Teubner 2011:210; 2012:142ss; 2006).
The constitutionalization of a social sub-system then consist of a process capable "to liberate the potential of highly specialized dynamics by institutionalizing it and, at the same time, to institutionalize mechanisms of self-restraint against its society-wide expansion" (Teubner 2004:12), by the means of fundamental rights conceived as "social and legal counter-institutions" (Teubner 2011:210). On the one hand fundamental rights include people within the subsystem, ensuring the access to its own medium, and on the other hand they are supposed to give "individuals and institutions outside of the social sphere guarantees of autonomy in regard to the latter's expansionist tendencies. (Teubner 2011:201).
According to Teubner (2011:202) "social fundamental rights must begin with exactly the forms that will be used to communicate in the respective medium", meaning that constitutionalization must operate within the specific systemic logic of a subsystem.
For Teubner Internet and the digital technologies represent one of the most interesting cases of societal constitutionalism in the making (Teubner 2004). Following Lessig, Teubner identified in the 'code' the communicative media of the Internet, conceived as "the digital incorporation of behavioral norms in the architecture of cyberspace" (Teubner 2004: 26).
Teubner assigns a crucial role to private arbitration tribunal and national courts in the social positivization of fundamental rights (Teubner 2011), paying little attention to the operation occurring at the operational level. This choice reflects Teubner's preference for private constitutional ordering, as he already discussed in his work on lex mercatoria (2002). Nevertheless, it seems to contradict its own principle according to which "fundamental rights must be readjusted to the rationality and normativity of different sub-areas" (Teubner 2012: 135). If "code" is the communicative media of the digital subsystem, then its

constitutionalization should rely primarily on the embodiments of fundamental rights within the technical design of digital technologies.
This kind of theoretical solution seems to find its empirical application in the spreading of the "human rights by design" approach, which is increasingly promoted by engaged technologists, academics and activists, as well as being incorporated in public telecommunications policy, such as the European Union General Data Protection Regulation (Suzor 2019, Milan and Oever 2017; Cath and Floridi 2017, Braman 2011).
This approach moves from the acknowledgment that technology is not neutral. Whether consciously or not, "the decisions embedded within protocols embed values and reflect the socioeconomic and political interests of protocol developers" (Denardis 2013:10).
It follows that standard setting bodies "are increasingly becoming arenas for tussles over value-sensitive design, and the moral as well as legal responsibility of technologists to protect human rights-by-design" (Cath and Floridi 2017: 450). The same could be said of other organizations involved in technological development such as ethical committees, professional and trade associations. All of them are relevant sources of norms production which could secure fundamental rights effectiveness. Indeed, as noted: "the technical community will not only be best placed but will have the sole ability to protect human rights standards […] precisely because they are the only community able to see the human rights issues that have been hard-wired into the very way in which the Internet operates" (Liddicoat and Doria 2012:15).
Embedding fundamental rights into code or design could be considered as a necessary complement to human rights law in the digital field, the technological side of that multilayered and "compensatory" constitutionalism filling the void and shortcoming of classical constitutionalism challenged by transnational regimes. (Peters 2006, Wet 2006).
Of course, lack of awareness and responsibility by technical operators about the negative externality and the danger to the integrity of individuals, community and institutions related to digital development could foster fundamental rights violations. As Lessig noted: "This code presents the greatest threat to liberal or libertarian ideals, as well as their greatest promise. We can build, or architect, or code cyberspace to protect values that we believe are fundamental, or we can build, or architect, or code cyberspace to allow those values to disappear" (Lessig 2006:6).

If constitutionalizing the digital sphere requires embedding fundamental right into the technological design, engaged epistemic communities could play a crucial role in this process. According to Haas (1992:3),

"An epistemic community is a network of professionals with recognized expertise and competence in a particular domain and an authoritative claim to policy-relevant knowledge within that domain or issue-area. Although an epistemic community may consist of professionals from a variety of disciplines and backgrounds, they have (1) a shared set of normative and principled beliefs, which provide a value-based rationale for the social action of community members; (2) shared causal beliefs, which are derived from their analysis of practices leading or contributing to a central set of problems in their domain and which then serve as the basis for elucidating the multiple linkages between possible policy actions and desired outcomes; (3) shared notions of Validity- that is, intersubjective, internally defined criteria for weighing and validating knowledge in the domain of their expertise; and (4) a common policy enterprise-that is, a set of common practices associated with a set of problems to which their professional competence is directed, presumably out of the conviction that human welfare will be enhanced as a consequence".

Epistemic communities are supposed to exert a notable influence in policy-making, in particular as regards novel, high complex and technical issues, due to their ability to "elucidating cause-and-effect relationships and providing advice on the likely results of various

courses of action; shedding light on the complex interlinkages between issues, and helping to define the self-interest" (Dunlop 2012:230)
Usually their influence is related to their ability to influence policy-makers conceived in the terms of public authority, more in detail bureaucrats and politicians. However, this traditional view of epistemic community influence could sound reductive, not considering the relevant role that they can play in the context of transnational governance.
Carayannis et al. (2012:135) highlighted how "epistemic communities are also relevant to the consideration of a global mode of regulation in that they are formed by international technical expertise rather than national political dynamics", inasmuch they are able to create "uniform systems of rules and norms that will form the basis of a global policy making agenda" (ibidem 132). According to Cogburn epistemic communities might be critical in international regime formation, favoring the "convergence of multi-stakeholder (e.g., governments, private sector, and civil society) expectations around the principles, values, norms, rules and decision-making procedures in a specific area, or interrelated areas, of international affairs" (Cogburn 2017:28).
The relevance of epistemic communities appears particularly relevant within the Internet Governance field, and particularly for the development of a discourse on human rights (Padovani, Musiani, Pavan, 2010). Indeed, since the beginning, Internet policy-making has been strongly shaped by the shared beliefs and visions of a cohesive technical community (Hofmann 2007, Drake 2004). Still today, the possibility to exert influence on the development of the information policy regime in a transnational landscape relies on the capability to manage the creation and dissemination of accepted knowledge and beliefs (Braman 2004, Cogburn 2004, 2017).
To our purpose, it is worth noting that they not only influence governments, but also other non-state actors with decision-making power. Epistemic communities "are at the forefront of recognised trends towards transnational governance, and they are a major means by which knowledge translates into power […] it is clear that epistemic communities not only seek to persuade states, but also a wide variety of non-state actors. They are not only underpinning specific government policies, but also shaping governance more broadly" (Cross 2013:138-9).
Another important point stressed by the literature is the capability of an epistemic community to build alliance with other distinct kinds of organizations such as advocacy networks, communities of practices, professional associations (Sebenius 1992, Dunlop 2012).
Epistemic communities with a strong commitment toward digital rights then could act as a hub, spreading digital constitutionalism standards and principles. They could influence public authority law-making and at the same time technological shaping by non-state actors such as standard organizations, technical communities, and private firms, from which they often come.

3. Case Background, Research Questions, Data and Methods

In order to explore the role that epistemic communities could play in the constitutionalization of the digital subsystem this paper focuses on artificial intelligence.
Indeed, among the new emerging technologies Artificial Intelligence (AI) is probably the one raising the most intense debate. Further, it is the one that more than any other makes visible the expansionistic logic of the digital subsystem logic and its impact on society.
Currently, mainstream AI's definitions revolve around a goal-oriented notion of rationality, according to which AI are systems capable of autonomously weighting different values to reach a choice while interacting with an external environment (for an exhaustive discussion on AI definition please refer Russel and Norving 2009, Turner 2019, Renda 2019).

It should be said that AI systems' logic corresponds to a pervasive process of datafication or digitalization of society (van Dijck 2014): the external environment (the society) is exploited to extract digital data, to be processed in order to elaborate added-value data (automated decision).
This extraction-processing-elaboration process deploys its logic irrespective of its consequences on human life and then is likely to produce negative externalities endangering the integrity of people, institutions, and communities. It raises several constitutional concerns, which range from privacy and data protection, to discrimination, to manipulation, misinformation and endangering of democratic institutions, to the effects on civil liberties, the rule of law and other procedural safeguards with regard their application in security and judiciary field, to the effects on human jobs and rights on the workplace (Latzer and Just 2020; Turner 2019, O'Neil 2016, Brownsword and Yeung 2018). Not by chance, the AI field has given rise to a flourish of ethical initiatives (Berkman Center 2020), and thus to a lively debate warning about the risk that this "ethification" corresponds to "ethical washing" practices masking the usual self-governance paradigm, and to what conditions, instead, it could lead to effective fundamental rights promotion (Floridi 2019, van Dijck and Casiraghi 2020, Yeung et al. 2019).

In more detail, this inquiry focuses on the work of the European Commission High-Level Expert Group on Artificial Intelligence (AI-HLEG).
This case appears particularly favorable to observe how an epistemic community could influence policy-making toward the adoption of a digital constitutionalism framework.
First, artificial intelligence is a novel and complex technical issue in its early-stage, conditions that seem to maximize epistemic community influence (Haas 1992, Radaelli 1999).
Second, knowledge and expertise have always been pivotal in the European Union policy-making practices (Radaelli 1999, Majone 1996), and even more for the Commission.
The Commission is specialized in the production of regulatory politics rather than redistributive (Majone 1996). As noted, "knowledge, rather than budget, is the critical resource in regulatory policy-making" making them "suitable for discussion and negotiation in expert circles" (Ibidem). Experts could also provide "information that is scarce and valuable to member states", transcending "the bias of information imbued with national interests" (Gornitzka and Sverdup 2011: 52), contributing to the integration process. Moreover, experts' advice could foster the Commission's capability to anticipate objections and concerns coming from the Parliament and the Council, raising the possibility for its proposal to be accepted (Tsebelis and Yataganas 2002). As Mets summarized, the European Commission could pursue a "problem solving" use of experts, to gain information and evaluate options; as well as "a politically substantiating use, so as to gain expert support for a predefined position against other actors; or a "political consensus-building use, so as to generate consensus among relevant actors in prospect of formal decision making" (Mets 2015: 10). For all these reasons committees, experts groups or other advisory bodies revolving around some DG or agencies are environments particularly conductive to the formation, or co-option, of epistemic communities (Radaelli 1999, Gornitzka and Host 2015, Cross 2015).

Third, the European Union has proven to be fully aware of the potential and concerns raised by AI. Besides addressing the matter within broader documents such as the General Data Protection Regulation, Digital Single Market Program or the Horizon 2020 program, EU institutions have devoted a specific attention to Artificial Intelligence in the last few years. The 2016 the European Parliament resolution (approved in 2017) "Civil Law Rules on Robotics" points out that the potential of these novel technologies "is nuanced by a set of tensions or risks and should be seriously assessed from the point of view of human safety, health and security;

freedom, privacy, integrity and dignity; self-determination and non-discrimination, and personal data protection" (European Parliament 2016:7). Further, the resolution calls for the responsibility and liability of developers and proposes to complete the existing EU legal framework with sets of ethical principles to inform research, design and usage of artificial intelligence and robotics.
The following Commission communication "Artificial Intelligence for Europe", released on April 25th 2018, incorporates the Parliament view in a more strategic plan. The document indeed assumes that Europe could aspire to a global leadership in the artificial Intelligence field building on its excellence in research and education, as well as on a strong and balanced regulatory framework capable of promoting innovation while protecting fundamental rights. This latter aspect is deemed "crucial for people to accept AI", building "trust" on AI, and overcoming the concerns related to its disruptive impact on society. In so doing, Europe could "set the global standard for a sustainable approach to this technology" (European Commission 2018a).
In this context a call for the establishment of a High-Level Expert Group on Artificial Intelligence was launched in spring 2018[4], with the tasks to 1) "advise the Commission on next steps addressing AI-related mid to long-term challenges and opportunities"; 2) "propose to the Commission AI ethics guidelines"; 3) "support the Commission on further engagement and outreach mechanisms". For this last purpose a broad multi-stakeholder platform, the European AI Alliance, was established.
Then, DG-CNECT selected and appointed 52 members on the basis of their expertise, of which 19 individuals appointed in their personal capacity (Type A members); 1 appointed to represent a common interest shared by stakeholders, namely the European DIGITAL SME Alliance (Type B members); 30 members representing organizations in the broad sense of the word, including academic institutions, companies, consumer organizations, trade unions, civil society interest groups (Type C members); two members representatives from other public entities, namely (Type E members), namely the European Union Agency for Fundamental Rights and the The European Economic and Social Committee (EESC).

Before setting out the research questions addressed by this paper, some further consideration on the concept of an epistemic community is needed.
It is worth noting that in their theorization, epistemic communities are not necessarily composed of scientists, and they could gather experts from different backgrounds (Haas 1992, Dunlop 2012).
As Haas clearly indicated, the essential feature of an epistemic community lies in the sharing of a set of normative and causal beliefs, an agreed notion of validity and a common policy enterprise, the combination of which allow us to distinguish epistemic communities from other kinds of groups involved in policy making (Haas 1992:18), such as interest groups (which have a common policy enterprise but not a casual belief and a validation notion) or professions and disciplines (which could have casual beliefs and validation notion, but generally miss normative beliefs and a common policy enterprise).
In my view, therefore, epistemic communities could be understood as *discursive networks*, or more specifically networks of experts held together by the production and reproduction of a specific policy discourse based on validated knowledge. An epistemic community could surely arise around some professional or academic organizations or advisory body (Dunlop 2010; Cross 2013, 2015) but this does not set its boundaries. Rather, epistemic communities cross institutional borders and spread their belief system from one organization to another. As noted,

[4] https://ec.europa.eu/digital-single-market/en/news/call-high-level-expert-group-artificial-intelligence

"the issue is not where community members sit but instead what they say" (Drake and Nicolaïdis 1992:39).
Of course, past professional or educational backgrounds could favor an ideational alignment (Schneiker and Joachim 2018), and collaboration among experts, common affiliations or attendance to a same set of organizations, conferences or initiatives could be all important indicators of an epistemic community at work (Haas 1992).

In the light of these considerations, the research questions addressed by this paper could be specified as follows:

RQ1) Has the HLEG-AI given rise to an epistemic community?
Which means:
1.1) Investigate the affiliation and backgrounds of HLEG-AI's members in order to detect indicators of previously existing epistemic communities.

1.2) Did the members of the HLEG-AI share a set of normative and causal beliefs, an agreed notion of validity and a common policy enterprise? Did this shared belief system represent a digital constitutionalism framework?

RQ2) To what extent the outputs of the HLEG-AI represent a convergence of expectations around a core set of digital constitutionalism principles and values?

RQ3) What is the role of HLEG-AI, as an epistemic community, in spreading digital constitutionalism views within both private and public policy-making processes, as well as into technical design?

With the purpose of verifying if the HLEG-AI could correspond to an epistemic community; biographic notes and curricula of the members of the expert-group have been retrieved form their personal pages or institutional websites and have been employed to perform an affiliation network analysis.
Affiliation network analysis is a particular kind of social network analysis, designed to explore two-mode networks (namely networks composed by two different sets of entities), in this case members and organizations (Borgatti et al. 2013; Carrington et al. 2005).
Data on co-affiliation are usually employed as a "proxy of social relation" on the assumption that "co-affiliation provides the conditions for the development of social ties of various kinds", and then data on "social proximity" could give us some insight also into "social similarity" (Borgatti and Halgin 2011:420).
Having retraced current and past affiliations of each member of the HLEG-AI and built-up a member-organization matrix, I extracted both a member-network (where members are tied by co-affiliations to the same organizations), and organizations-networks (where a tie between two different organizations corresponds to a common member), and related graphs and statistics through the software Gephi.
In this case, affiliation network analysis is supposed to provide information on three different aspects relevant to this inquiry:
i) if the HLEG-AI's members have some common professional or educational background, where they could have acquired similar ideas, approaches and culture;
ii) if effectively HLEG-AI's members worked together in some organizations or initiatives, which could indicate an epistemic community in action;

iii) the number and the variety of organizations to which each member of the HLEG-AI is related, which could give some insight about the potential for the spread and dissemination of the whole group (this point is particularly relevant for RQ3).

In order to investigate if, and to what extent, the members of the HLEG-AI effectively shared common beliefs, I realized a second two-mode network, this time relating actors with concepts. This kind of analysis could be conceived as a simplified version of a Discursive Network Analysis, as described by Leifeld (2017).
To this purpose I collected texts (public statements, articles, blog posts, reports, etc.) drafted by the HLEG-AI members concerning Artificial Intelligence, and if possible, its ethical and fundamental rights implications. Preferably, I selected materials released prior to the establishment of the HLEG-AI. When this was not possible, I resorted to later items. In some cases, regarding members appointed as representatives of an organization, statements and position papers released on behalf of the organization have been included in the analysis even if they could not be surely attributed to the individual member joining the HLEG-AI.
Key concepts contained in this textual corpus have been hand-coded through the NVIVO software. Then I built a matrix of members-concepts and performed network analysis through the software Gephi.
The analysis then allowed us to map arguments and beliefs of the HLEG-AI members, to identify the most shared concepts and ideas, and members of HLEG-AI grouped around them, making it possible in the end to verify if HLEG-AI members shared an epistemic community belief system.

The results of this latter analysis are useful also to answer RQ2 and to explore to what extent the outputs of the HLEG-AI represent a convergence of expectations around a core set of digital constitutionalism principles and values. Comparing the content of HLEG-AI outputs with the socio-conceptual network of its members could indeed provide insights into to what extent divergences among members (if any) have been resolved, leading to a higher synthesis, and how this process signalled a move towards digital constitutionalism. Unfortunately, transcripts of the HLEG-AI meetings were not made available to the public, so it was not possible to re-trace the discussion and discursive interactions within HLEG-AI that led to the outputs.

With regard to RQ3, as mentioned above, results of affiliation analysis could provide some insight into potential for the spread and dissemination of the HLEG-AI. Furthermore, EU institutions and Members States' policy documents concerning Artificial Intelligence, as well as the policies of non-state actors (such as private firms, research centers, especially the one involved in the European AI Alleance initiative), released after the HLEG-AI submitted its output have been analyzed in order to investigate the impact of HLEG-AI on European policy-making.

4. Results and Discussion

4.1 Background and affiliations of HLEG-AI members

Affiliation analysis pointed out the existence of two different components within the HLEG-AI, that is two different "set of nodes in which every node can reach every other by some path" (Borgatti et al. 2013: 26), in our case a web of shared or overlapping membership with a set of organizations (Fig.1). The first component gathers 29 members, and the second 7, while the

remaining 16 members of the HLEG-AI have no co-affiliations with the other members of the expert group.

Fig. 1 Affiliation Network - Members

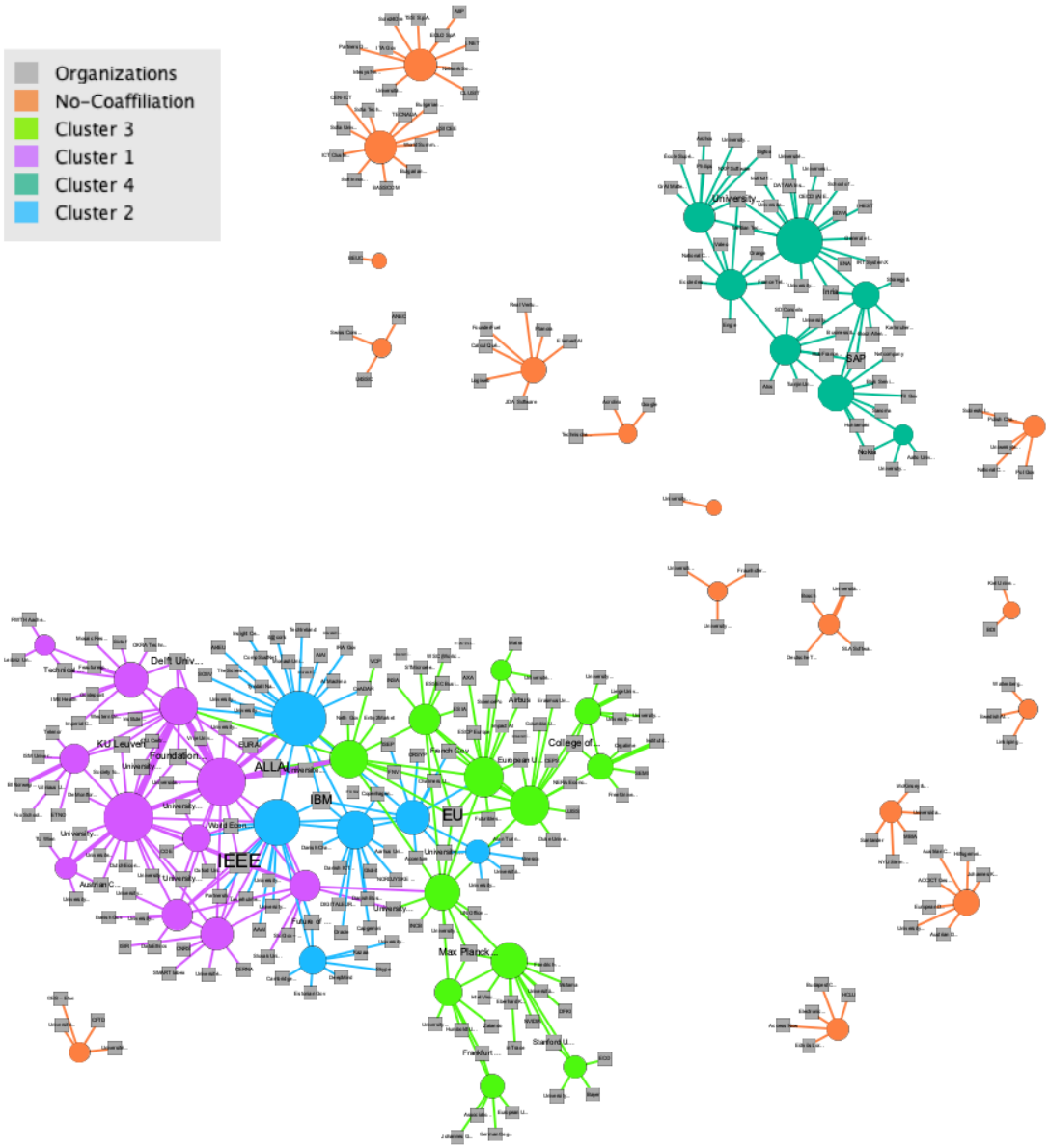


Fig.1 shows also that the first component could be further broken down into three clusters, based on the density of their interconnection[5]. Cluster 1 has a clear Tech-Academic profile. Most of its members belong to the Institute of Electrical and Electronics Engineers (IEEE), and enjoyed the same sub-groups or initiative such as the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems or the IEEE Robotics and Automation Society. Further, some of them share training or teaching experience in tech-focused universities. Members of Cluster 2 for the vast majority had or have some professional experience in IBM, have taken

[5] The cluster have been obtained throught the "Modularity" function of Gephy. For more detail please consult Blondel et al. 2008.

part in advisory bodies of national European governments and two of them are members of the Future of Life Institute. Members of Cluster 3, instead, share overlapping affiliations to some Europe-wide organizations. The cluster indeed gathers members, or former members, of EU agencies, committees or expert groups, College of Europe, European University Institute, Airbus, European trade and business associations. Some of them, further, share training or teaching experience at the Max Plank Institute (even if in different sites).
All the members of the second component (Cluster 4) belong to private firms and are scarcely related to other stakeholder groups.
The members of the HLEG-AI who do not have co-affiliations with other members constitute the more variegated group and show the highest concentration of civil society associations.

Tab.1 Clusters Composition

	Component 1	Cluster 1	Cluster 2	Cluster 3	Cluster 4	No Co-Affiliation
Private	16	2	5	8	7	9
Academic	22	10	4	8	2	5
Government	9	1	3	5	1	3
Civil Society	3	2	0	1	0	5
Multi Stk	8	3	2	3	1	1
Tech	7	5	2	0	1	3
Ios	1	1	0	0	0	1
Other	0	0	0	0	1	1
N. of Members	29	11	7	12	7	16

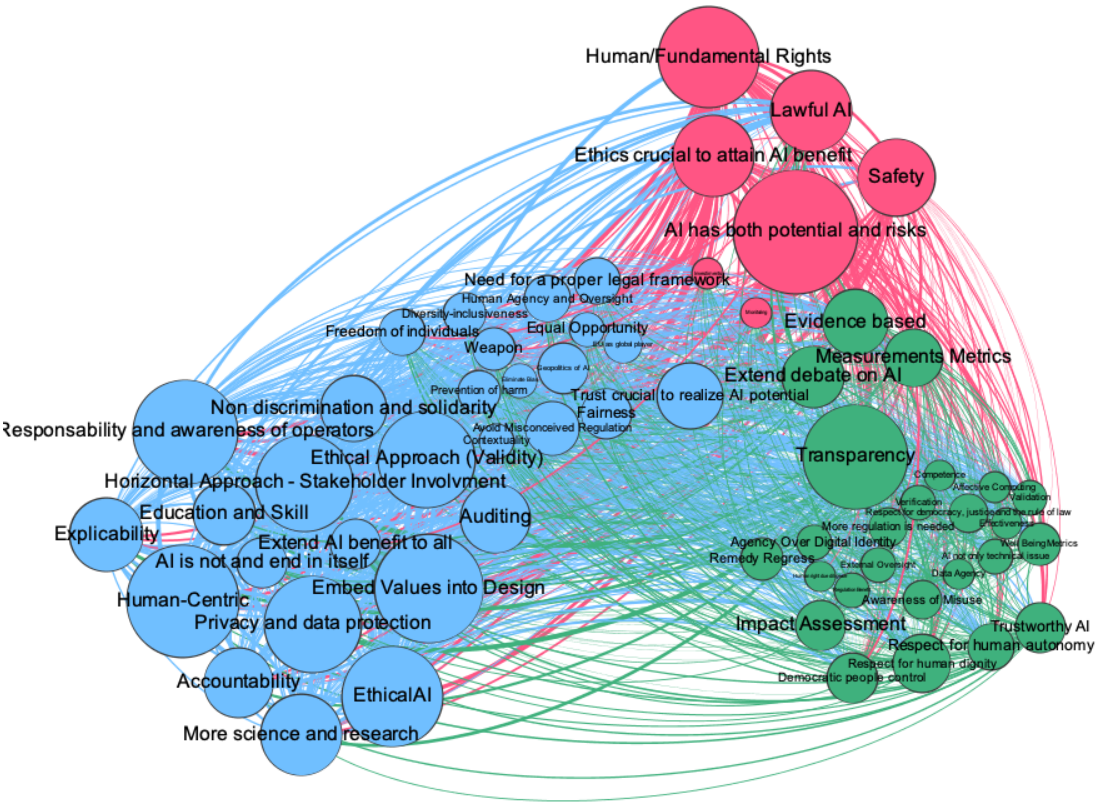
These findings suggest that the HLEG-AI could have gathered members from one or more pre-existing epistemic communities, developed around specific professional, academic and technical organizations.
Further evidence of the existence of already existing epistemic communities within the HLEG-AI, is provided by the fact that many of its members co-authored, or subscribed, relevant texts in the field of artificial intelligences ethics and principles, such as the IEEE “Ethically Aligned Design”, or the Future of Life Institute “Asilomar AI Principles”, or the AI4People “Ethical Framework for a Good AI Society”. Members of the HLEG-AI then, not only have a shared background in terms of training and professional experiences, but they effectively cooperated in some initiatives.
It is worth noting that, although most of the authors of IEEE “Ethically Aligned Design” belong to Cluster 1 and some of the authors of the Future of Life Institute “Asilomar AI Principles” are in Cluster 2, all the co-authored documents have been drafted or subscribed by members of different clusters.
These findings suggest that at least the first component could be considered as the manifestation of a unique and epistemic community developed around several poles and held together by members and initiatives acting as bridge and hub among different parts of the network[6]. A network composed for the most part of academic or technical organizations, or people who had some academic experience in the past, capable of finding a place within governmental advisory bodies, research departments of private firms or multi-stakeholder policy-making arenas.

[6] Data on betweeness centrality confirm this special role played by a limited number of members, organization and initiatives. Betweenness centrality measures "how often a given node falls along the shortest path between two other nodes”, being interpreted as an indicator “of the potential for controlling flows through the network (Borgatti et al. 2013:185)

4.2 The discursive network

Discourse network analysis confirms that members of the HLEG-AI shared a common discourse on principled artificial intelligence, and that this discourse corresponds to the epistemic community belief system highlighted by Haas.
Fig. 2 shows the discursive network of HLEG-AI[7]. Nodes represents arguments or concepts; their connection indicates they have been expressed by the same HLEG-AI members. The size of each node corresponds to the number of HLEG-AI members who shared that argument/concept.

Fig. 2 Discursive Network



It is worth noting that the HLEG-AI discursive network gives rise to a single component, meaning that the all the members of the HLEG-AI share, partially and in an overlapping manner, the same set of concepts, regardless of their affiliation or occasions to collaborate with the other members of the group.
The analysis allows us also to distinguish three different clusters within the network. However, we need to pay heed and point out that the clusters indicate that certain groups of concepts are

[7] The analysis concerns only 40 members on 52 of the HLEG-AI, since for 12 of them it was not possible access to relevant statement on AI.

most frequently referred to by the same members, but all the members of HLEG-AI refer to concepts belonging to all clusters.
The first cluster gathers some of the most quoted concepts and could be considered a sort of minimum common agreement of the expert group. Most of the examined contributions start with the assumption that AI raises both potential and risks. It might sound trivial, nevertheless this stance contrasts both "cyber-enthusiastic" and "cyber-apocalyptic" narratives that informed the debate around new and digital technologies in the last few decades; the former emphasizing the benefit and minimizing the risk claimed for "innovation without permission", while the latter related new technologies to dystopian scenarios and called for strong regulation and precautionary approach. Rather, to unleash the benefit of AI is deemed necessary to guarantee safety, to make proper investment in its development but also in monitoring its effects, and above all to adopt an lawful ethical approach, rooted in fundamental rights, meaning that ethics aims at complete rather than substitute law.

Tab. 2 20 Most Shared Concepts

Concepts	N.
AI has both potential and risks	28
Human-Centric AI	25
Embed Values into Design	24
Transparency	23
Responsibility and awareness of operators	23
EthicalAI	22
Human/Fundamental Rights	22
Horizontal Approach - Stakeholder Involvment	21
Privacy and data protection	21
Validity of Ethical Approach	21
Ethics crucial to attain AI benefit	17
More science and research	17
Lawful AI	17
Safety	16
Explicability	15
Auditing	14
Accountability	14
Trust crucial to realize AI potential	13
Non discrimination and solidarity	13

The second cluster refers to a strategic and geopolitical discourse. In this view, ethics is considered crucial to attain AI benefits because ethics could generate the trust necessary for AI development and full acceptance by society at large. Moreover, from a European perspective, ethics it is deemed crucial to create AI standards safeguarding EU values and sovereignty. Further, in this view ethics figure as the added-value of a human-centric European AI, liable to make EU a global leader in this field.
The third cluster focuses on the necessity to develop instruments and procedures (such as verification, metrics, impact assessment) to make compliance with human rights and democratic values effective.
The views represented by these two latter clusters do not conflict. Rather, members of HLEG-AI carry forward both. The only point on which experts' opinions differ, echoing the opposition between "cyber-enthusiastic" and "cyber-apocalyptic", or between neoliberal and sovereigntist mentioned at the beginning of the article, is regulation.
People from business warn against the impact of heavy-handed regulation on AI development. On the contrary civil society and more engaged members claim for *ad hoc* legislation. However, positions are much more nuanced and closer rather than in past debates on technology. HLEG-AI members agree on the need for a proper legislative framework providing

both trust among citizens and certain for the operators. An ethical approach, rooted in fundamental rights and liable to be institutionalized through a certification system or a regulatory body oversight, is generally deemed the best solution.

Above all, it is important to consider how the set of arguments, ideas and concepts shared by the members of the HLEG-AI, as a whole correspond to the belief system that, according to Haas, defines an epistemic community.
Cluster 1' concepts could be included among the **normative and casual beliefs**.
Members of the HLEG-AI indeed seek to transcend both "cyber-enthusiastic" and "cyber-apocalyptic" views in a higher synthesis. On the basis of a growing body of evidence and their in-depth understanding of the functioning of AI technologies, they argue about its unprecedented disruptive potential, both in positive and negative terms.
HLEG-AI members agree that AI is much more than a technical issue and its governance could not be limited to efficiency, innovation and economic growth. Rather, most of the HLEG-AI conceive AI as a "socio-technical" system and maintain that "the full benefit of these technologies will be attained only if they are aligned with society's defined values and ethical principles" (IEEE 2016-19:3), as well as fundamental rights. The rationale is twofold. First, ethical and fundamental rights compliance helps to avoid damages caused by misuse of AI. Second, it also contributes to create trust around this technology, favoring its acceptance in the society at large and avoiding the loss of opportunities and benefit due to AI underuse (Floridi et al. 2018).
Another key concept is that, due to the nature of this technology, making fundamental rights and ethical considerations effective requires their embedment within the AI systems' design, as well as the awareness and responsibility by engineers, developers, managers and operators of AI system.
This is a task that goes far beyond allowing or inhibiting some behaviors in the code, or adopting an open source code to permit public scrutiny, since the partial autonomy of these systems and the "black box" among input and output that characterize "deep-learning" and "neural-network", could undermine these measures. Rather, the "by design" approach in these cases call into play the socio-technical nature of AI, requiring awareness and responsibility by all the operators involved and consideration in all AI application lifecycle (Renda 2019; Yeung et al. 2019).

Members of HLEG-AI then share a common **policy enterprise**, consisting in promoting and realizing an "Human-centric AI", also referred as "Beneficial AI", "Well-Being AI", or "AI for All".
These *formulae* indicate that AI "is not an end in itself" (Villani et al. 2018: 7), and that "individual, not the value of their data, must be at the centre of any discourse" (O'Sullivan, 2017:2). Then, "these systems must be developed and should operate in a way that is beneficial to people and the environment, beyond simply reaching functional goals and addressing technical problems […], A/IS creators shall adopt increased human well-being as a primary success criterion for development" and AI systems should be "designed to respect human rights, align with human values, and holistically increase well-being while empowering as many people as possible" (IEEE 2016-19:3-10, the same words have been used also in the Alisomar AI Principles and AI4People Framework).
Among HLEG-AI members there is a strong consensus among the key principles, which should drive the development of a Human-Centric AI.
In particular, most of them empathized a triad of strictly related principles, namely transparency, explicability and accountability, which are deemed play to a pivotal role in creating a trustworthy environment for AI. Briefly speaking, within AI context, transparency

needs to provide information about "to the data used, the fact that a user interface is not human, as well as to the identity of those that trained the system, and basic information on how the system was trained" (Renda 2019:53), as well as about its risks (Access Now 2018:14). Explicability could be understood as a sub-dimension of transparency focusing on the explanation of the decision-making process carried out by the AI. Both of them are a pre-condition for accountability, that in this context is mainly concerned with the identification of responsibility in the case of harmful output and the related remedy to/redress procedures.

Another cluster of related principles includes fairness, equality, non-discrimination, and solidarity. All of these principles are concerned with the capability of AI systems to produce exclusion and inequalities, with regard to at least three different dimensions. First, the possibility that AI systems are designed embedding bias, or are trained with biased data, in such a way that they systematically damage certain groups of people. Second, even when they are not biased, AI systems could exclude vulnerable and marginalized groups from the access to basic services (e.g. health insurance). Third, in a broad sense, the equal distribution of AI benefits the whole society.

Privacy and data protection are also often mentioned, mainly referring to individuals' access to, and control over their data and digital identity, as well as the guarantee that personal data are properly secured.

This last aspect leads to the consideration of another shared principle: safety, which refers to the safeguard of human integrity and prevention of harm through technical robustness in order to prevent malfunctioning, errors or hacking.

The policy enterprise of this epistemic community also includes the development of procedures and instruments to effectively embed the abovementioned values and principles within the design of an AI system, such as impact and risk assessment, auditing and tests. Acknowledging that this process is still in its infancy, most members of the HLEG-AI call for fund and investment on research.

Members of HLEG-AI have also shared **notions of validity**. In this regard, three different strands could be distinguished.

First of all, for what concerns the definition of the principles that should inform AI systems, the vast majority of the experts group agree on the adoption of an ethical approach. Another widely held belief is that there is not one ethical methodology or tradition to be preferred. IEEE Ethically Aligned Design refers to virtue, deontological, utilitarian and care ethics, as well as to the contribution of non-western ethical traditions. Floridi (2016) recall the legacy of computer and information ethics to found a novel data ethics.

Despite this diversity, as seen above, members of HLEG-AI converged on a relative short list of principles. Rather, this assumption has as corollary that any list of principles must not be considered definitive (AI4People initiative talk about a "living document"), and those principles could conflict with each other. Thus, each ethical consideration should be placed in its context and based on evidence.

This point implies also a stance for ethics as an ongoing process, a mind-set that should permeate each AI related activity, rather than a fixed reference system.

Another shared belief on validation criteria is that ethical consideration and ethical design should be a driven evidence-based approach and scientific research. A sub-set of these beliefs assumes that the impact of AI on humans and society could be quantified through proper measurement and metrics, and these metrics could be embedded within AI design or used to monitor, evaluate and modify AI systems. For example, IEEE talks to this purpose of "Well-Being Metrics", referring to a wide range of already existing indicators of well-being to be adapted and included in AI development processes; AI4People initiative of "metrics for the

trustworthiness of AI", Renda refers to "quantitative fairness" metrics already in use in machine learning systems.
Further, several members of HLEG-AI agree on the necessity to adapt tools already employed in software development to ensure human rights and ethical standards compliance. In particular they refer to verification and validation. The former consists in formal or mathematical checking that the system meets some specified requirement. Validation instead provides empirical evidence about the capability of the system to fit the purpose for which it was created. Usually it requires stakeholders' involvement and it includes techniques such as penetration tests and user experiences design.
Usually, HLEG-AI members consider these techniques as part of a broader process of auditing or human right/ethical due diligence, including also regulatory governance techniques such as requirement analysis, risk and impact assessment, testing, post-implementation monitoring, data and documentation requirement (for example logs of system operations). Further, they are also related to certification or external oversight processes, carried out by ad hoc governmental agencies or international standard bodies such as IEEE itself or ISO.

4.3 Convergence around a core set of digital constitutionalism principles and values.

The HLEG-AI released a first draft of the "Ethical Guidelines for Trustworthy AI" in December 2018, open to stakeholders' consultation, and a final version in April 2019.
Generally speaking, the Ethical Guidelines recall almost all the arguments abovementioned, that is, it seems to reflect the epistemic community beliefs system as it has been traced back to the previous paragraph.
The guidelines indeed draw on the "human centric" narrative, transcending both cyber-enthusiastic and cyber-apocalyptic approach; their declared aim is "set out a framework for achieving Trustworthy AI", to foster AI development "seeking to maximise the benefits of AI systems while at the same time preventing and minimising their risks".
As a whole, the guidelines recall all the principles and validation criteria mentioned above, as well as geopolitical considerations, and the belief about the necessity to embed principles within socio-technical design.
For the purpose of this work is is particularly relevant to highlight how through the work of the HLEG-AI, the commitment toward digital constitutionalism principles, already present in many of its members, reaches an even more broad consensus and a clearer definition, becoming foundational in the HLEG-AI view. Further, also the principle of the embedment of digital constitutionalism principles within socio-technical design, finds in the experience of the HLEG-AI a pivotal moment of development and puts it into practice.
Indeed, the guideline states: "We believe in an approach to AI ethics based on the fundamental rights enshrined in the EU Treaties, the EU Charter and international human rights law. Respect for fundamental rights, within a framework of democracy and the rule of law, provides the most promising foundations for identifying abstract ethical principles and values, which can be operationalised in the context of AI".
Moreover, the guidelines clarify the relationship between fundamental rights, ethical principles and law or regulation.
The Trustworthy approach is grounded in fundamental rights and composed of three components: Lawful AI; Ethical AI; Robust AI. As pointed out: "Understood as legally enforceable rights, fundamental rights therefore fall under the first component of Trustworthy AI (lawful AI), which safeguards compliance with the law. Understood as the rights of everyone, rooted in the inherent moral status of human beings, they also underpin the second

component of Trustworthy AI (ethical AI)", while the third component, robust AI, refers to the expectation that AI systems perform their activities in a safe, secure and reliable manner, from both a technical and societal perspective.
From fundamental rights stem four ethical imperatives for the AI environment (Respect for human autonomy; Prevention of harm; Fairness; Explicability), translated in "seven key requirements that AI systems should implement and meet throughout their entire life cycle" by "developers, deployers and end-users, as well as the broader society": human agency and oversight, technical robustness and safety, privacy and data governance, transparency, diversity, non-discrimination and fairness, environmental and societal well-being and accountability.
The ethical approach is then conceived as a complement, not a substitute, of law and regulation, which could overcome the difficulties encountered by law in terms of jurisdiction, technological pace and specificity.
Unlike most of other AI ethical codes, the HLEG-AI Ethical Guidelines also introduces a set of technical and non-technical methods to implement ethical requirements, and provide a pilot version of a concrete assessment list aimed at operationalizing Trustworthy AI and check the fulfillment of its requirements.

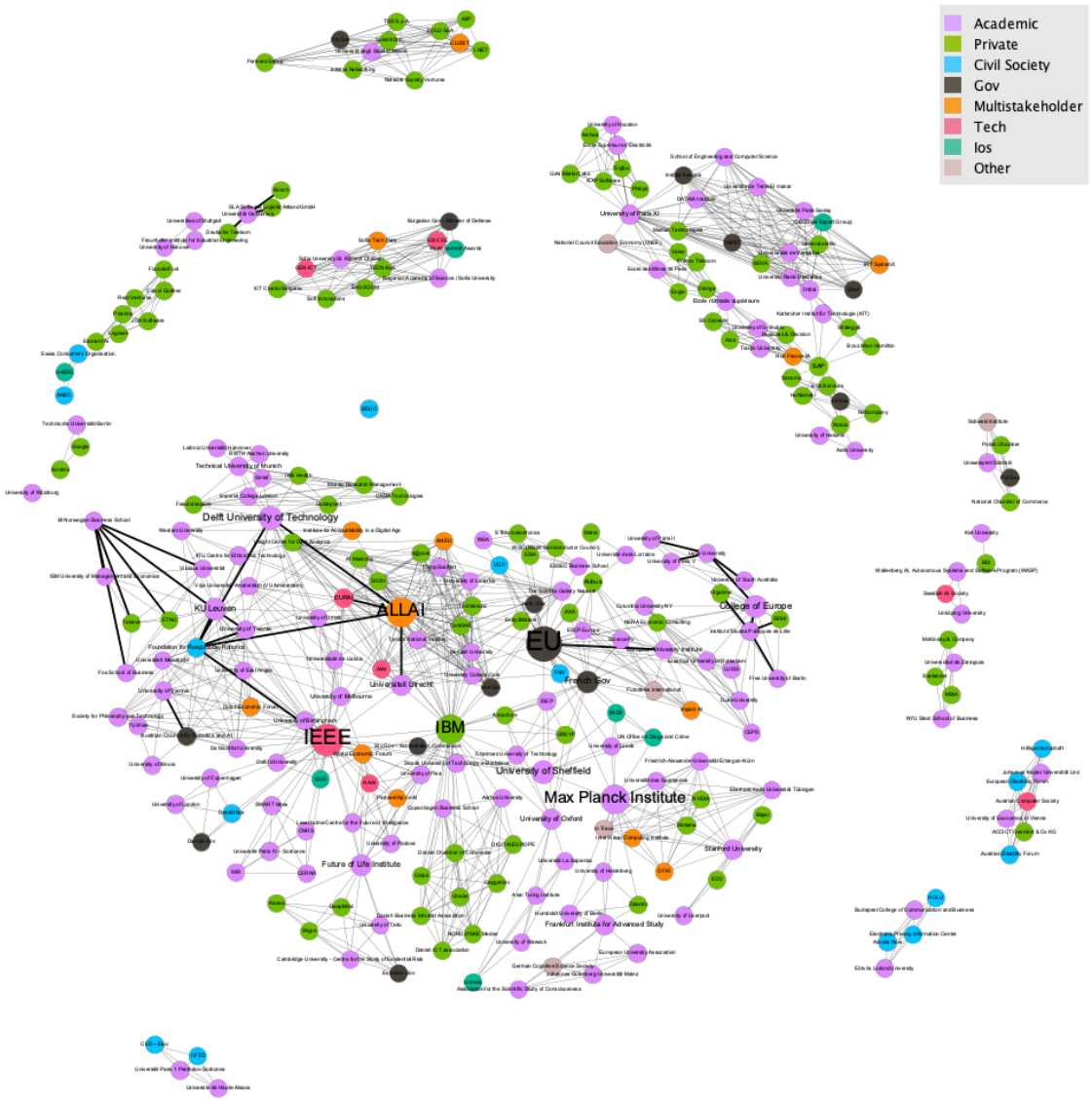
4.4 The role of HLEG-AI in spreading digital constitutionalism views

A first insight into the potential for the HLEG-AI to disseminate and spread digital constitutionalism principles enshrined in its Ethical Guideline can be drawn by the affiliation network, focusing this time on the organizations rather than the members.
In Fig. 3, nodes are organizations to which members of HLEG-AI are or were affiliated: their link means they have a shared member, their size represents their betweennesses centrality.
This picture represents surely the background of the HLEG-AI, the environment from which its ideas could have arisen, but at the same it could figure the narrow circle where the output of the HLEG-AI could be disseminated, the ensemble of the organizations with which HLEG-AI's members already have a relation.

Tab. 3 Composition and Statistics Affiliation Network for Stakeholder Type

	N.	Degree Sum	Degree Mean	Bet. Centr. Mean	Eig. Centr. Mean
Academic	126	1009	8,01	108,00	0,12
Civil Society	15	78	5,20	15,72	0,06
Gov	15	169	11,27	411,25	0,18
Ios	7	45	6,43	0,00	0,08
Multistakeholder Initiative	13	158	12,15	259,53	0,24
Private	96	778	8,10	39,75	0,11
Tech	8	117	14,63	528,57	0,31
Other	5	32	6,40	0,00	0,04

Fig.3 Affiliation Network – Organizations



As shown in Tab.3, not surprisingly, the network is composed for the most part of academic institutions (we also considered affiliations related to members' educational background), followed by the private sector (including some big companies such as IBM, AXA, Bosch, Airbus, Zalando, SAP, Nokia, and many little software-houses, consulting firms or think-thanks, often founded by the members themselves), while other categories of organizations have a more limited presence. However, centrality measures[8] reveal that together with academic institutions, technical, multi-stakeholders, and governmental (mainly advisory bodies) have the higher and most strategic number of connections and then are best placed to spread their views across the network.

[8] Centrality measures provide quantitative scores on the relevance of each node as regarding the connectivity and the flow of information within the network. Degree centrality counts of the number of connections of a node. Eigenvector centrality weights the ties of a node based on the degree centrality of the node to which it is connected. For betweeness centrality please consult note 4. For more detail see Borgatti et al. 2013.

Within the HLEG-AI, we can find a network composed mainly of people with a research background capable of entering the private and governmental sectors. The HLEG-AI itself offered a further opportunity to contact and share views and initiatives with other actors, especially with NGO and civil society organizations.
Just considering its members' already known affiliations, the epistemic community within the HLEG-AI appears capable of spreading its belief system to a large and diversified number of organizations.

However, observing the effective influence of the epistemic community gathered by the HLEG-AI, requires taking into account how the expert group outcomes have been incorporated and adopted by organizations more directly involved in policy-making.
First, one must look at the following European Union policy documents. Even before the expert group released its guidelines, they have been recalled in the communication COM(2018) 795, "Coordinated Plan on Artificial Intelligence" of December 2018[9], as a key factor to "ensure that the EU as a whole can compete globally", becoming "a global leader in developing and using AI for good and promoting a human-centric approach and ethics-by-design principles".
As one can see from the very title, the HLEG-AI Guidelines are considered "valuable input for policy-making" also in the Commission communication COM2019-168 "Building Trust in Human-Centric Artificial Intelligence"[10]. In this document, the Commission embraces the guidelines' concepts, arguments and requirements, welcome their "the practical nature and the "concrete guidance they offer to developers, suppliers and users of AI on how to ensure trustworthiness", and stresses "its efforts to bring the Union's approach to the global stage and build a consensus on a human-centric AI".
Guidelines are also recalled by the recent Commission's White Paper on Artificial Intelligence[11]. The document is particularly interesting inasmuch as it clarifies the envisaged legal status of ethics guideline requirements. The White Paper indeed, based on the input obtained during the piloting phase of the Ethics Guidelines, notes that most of its requirements are already covered by existing "European legislation on fundamental rights (e.g. data protection, privacy, non-discrimination), consumer protection, and product safety and liability rules". Instead, further "mandatory legal requirements" (with regard to training data, data and record-keeping, human oversight), related to AI technologies specificities, could be foreseen for "high-risk" application only, to ensure the enforcement of European fundamental and consumer rights. For low-risk application instead, these further requirements could give rise to a voluntary labelling scheme.
Ethics guidelines are also recalled in more than 12 national and international policy documents released between 2019 and 2020, for the most part national AI strategies drafted under the impulse of the abovementioned "Coordinated Plan on Artificial Intelligence"[12].
As mentioned above, the communication "Artificial Intelligence for Europe", besides the expert group, also provided for the establishment of the "European AI Alliance", a multi-stakeholder platform to "to gather input, exchange views, develop and implement common measures to encourage the development and use of AI".

[9] https://ec.europa.eu/knowledge4policy/publication/coordinated-plan-artificial-intelligence-com2018-795-final_en
[10] https://ec.europa.eu/digital-single-market/en/news/communication-building-trust-human-centric-artificial-intelligence
[11] https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf
[12] More in detail, they are recalled in the AI Strategies of Catalonia, Switzerland, Belgium, Italy, Finland, Norway, Serbia, Malta, Lithuania, as well as in the Eurocontrol, (2020), FLY AI Report.

AI Alliance stakeholders took an active part in the work of the HLEG-AI. They submitted more than 500 contributions to the first draft of the Ethical Guideline released in December 2008. Further, over 450 stakeholders participated to the piloting phase (from 26 June to 1 December 2019) of the Ethics Guidelines' assessment list[13], providing feedback which will be employed to draft an improved version of this instrument to be released in the next few months. Some stakeholders, including AXA, IBM, SAP, have already officially endorsed or recalled HLEG-AI guidelines in their policies[14].

Conclusions

This study assumed that epistemic communities play a crucial role in the constitutionalization of the Internet and digital technologies since they have the necessary technical expertise and policy commitment to translate normative provisions into architectural design.
In order to illustrate this nexus, I performed an in-depth investigation on the case of the European Commission High Level Exert Group on Artificial Intelligence.
The analysis revealed that the HLEG-AI effectively gathered an epistemic community sharing normative and casual beliefs, notions of validity and a policy enterprise. This epistemic community has proved to be strongly committed with fundamental rights, conceived as instruments to both unleash the potential and limit the negative impact on the integrity and autonomy of individual and communities of AI systems.
The inquiry has also pointed out how the work of this epistemic community was able to influence the following policy-making activities of the European Commission and Member States, as well as of other non-state stakeholders, whose involvement is essential in order to embed fundamental rights within the design of AI systems and to realize a workable ethical approach.
The case of the HLEG-AI illustrates how the activities of an epistemic community could, on the one hand, inform the law-making process and, the other hand, raise awareness and accountability of the operators, offering them a normative framework and practical guidance to implement digital constitutionalism principles and values.
In this regard, the recent IBM stance on facial recognition deserves to be mentioned. Following the George Floyd case, IBM announced it would "no longer offer general purpose IBM facial recognition or analysis software. IBM firmly opposes and will not condone uses of any technology, including facial recognition technology offered by other vendors, for mass surveillance, racial profiling, violations of basic human rights and freedoms", asking the U.S. Congress "to begin a national dialogue on whether and how facial recognition technology should be employed by domestic law enforcement agencies[15]". Shortly after IBM's decision Amazon and Microsoft followed suit. Further, IBM asked the Department of Commerce to include facial recognition system allowing for facial matching in the list of export controlled technology[16].

---

[13] https://ec.europa.eu/futurium/en/ethics-guidelines-trustworthy-ai/stakeholders-registered-pilot-ethics-guidelines-trustworthy-ai; https://ec.europa.eu/futurium/en/ethics-guidelines-trustworthy-ai/register-piloting-process-0.
[14] https://www.ibm.com/blogs/policy/ai-ethics-eu/; https://www.ibm.com/downloads/cas/J2LAYLOZ; https://news.sap.com/2019/12/sap-ethical-ai-guiding-principles-one-year-later/; https://www.axa-research.org/en/news/AI-research-guide.
[15] https://www.ibm.com/blogs/policy/facial-recognition-sunset-racial-justice-reforms/
[16] https://www.ibm.com/blogs/policy/facial-recognition-export-controls/

This story may indicate that if an epistemic community engaged with digital constitutionalism values and principles spreads its views and succeeds in raising the awareness and sense of responsibility of the people working on digital technologies, it may result in professional and organizational culture capable of affecting the behavior of transnational non-state actors[17]. Insofar as it occurs, transnational "constitutional" norms and rules could be implemented by the means of socio-technical systems, or at least a convergence could be favored around global rules.
However, it should be added that the IBM stance, after having raised up expectations of human right defenders, it shortly after caused their disappointment whereas the company did not clarify in which concrete acts its commitment has been translated[18], fueling the suspicion that "the organisation is motivated by public relations, instead of fundamental rights"[19].
Evaluating the truthfulness of IBM commitment is out of the scope of this piece, nevertheless this last observation provides an opportunity to recall the warning against ethical washing practices and remind that digital constitutionalism, resulting in the interplay between technical design and law-making, requires external independent oversight, if not the "shadow of hierarchy" (Heritier 2008), in order to ensure accountability and effectiveness.

References

Access Now. The Toronto Declaration: Protecting the rights to equality and non-discrimination in machine learning systems. (2018). https://www.accessnow.org/the-toronto-declaration-protecting-the-rights-to-equality-and-non-discrimination-in-machine-learning-systems/ Accessed 1st September 2020.

Bäckstrand, K., Khan, J., Kronsell, A. and Lövbrand, E. (2010). *Environmental Politics and Deliberative Democracy. Examining the Promise of New Modes of Governance*. Cheltenham, UK: Edward Elgar.

Berkman Center. Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI. (2020). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3518482. Accessed 1st September 2020.

Blondel, V. D., Guillaume, J., & Lefebvre, E. Fast unfolding of communities in large networks. (2008). https://arxiv.org/abs/0803.0476; https://arxiv.org/pdf/0803.0476.pdf. Accessed 1st September 2020.

Borgatti, S. & Halgin, D. (2014). Analyzing affiliation networks. In J. Scott & P. J. Carrington (eds), *The SAGE handbook of social network analysis* (pp. 417-433). London: SAGE Publications.

Borgatti, S., Everett, M., Johnson, J. (2013). *Analyzing Social Networks*. London: Sage.

[17] Even if not related to AI, consider for example how Facebook employees recently succeeded in changing the company policy on content governance with regard to Trump posts, https://www.forbes.com/sites/siladityaray/2020/08/18/facebook-will-remove-trump-posts-that-violate-standards-says-sandberg/#66b1a1372a02

[18] https://edri.org/our-work/open-letter-edri-calls-on-ibm-to-clarify-stance-on-facial-recognition/

[19] https://edri.org/our-work/ibm-facial-recognition-solution-cannot-be-left-to-companies;

Braman, S. (2004). *The Emergent Global Information Policy Regime*. New York: Palgrave MacMillan.

Braman, S. (2011). Privacy by design: Networked computing, 1969–1979. *New Media & Society*, 14,(5), 798-814.

Brownsword, R. and Yeung, K. (eds) (2008). *Regulating Technologies*. Oxford: Hart Publishing.

Carayannis, E.G., Pirzadeh, A. and Popescuet, D. (2012). *Institutional Learning and Knowledge Transfer Across Epistemic Communities*. New York: Springer.

Carr, M. (2015). Power Plays in Global Internet Governance. *Millennium*, 43(2):640-659.

Carrington, P.J., Scott, J., Wasserman, S. (2005). *Models and Methods in Social Network Analysis*. Cambridge: Cambridge University Press.

Cath, C. and Floridi, L. (2017). The Design of the Internet's Architecture by the Internet Engineering Task Force (IETF) and Human Rights. *Sci Eng Ethics*, 23, 449–468.

Cogburn, D. (2017). *Transnational Advocacy Networks in the Information Society*. New York: Palgrave MacMillan.

Cogburn, D.L. (2004). Elite Decision-Making and Epistemic Communities: Implications for Global Information Policy. In Sandra Braman, *The Emergent Global Information Policy Regime* (pp. 154-178). New York: Palgrave MacMillan.

Cross, M.K. (2013). Rethinking epistemic communities twenty years later. *Review of International Studies*, 39, 137-160.

Cross, M.K. (2015). The Limits of Epistemic Communities: EU Security Agencies. *Politics and Governance*, 3(1), 90-100.

De Gregorio, G. and Radu, R. Fragmenting Internet Governance: Digital Sovereignty and Global Constitutionalism. (2020). http://www.medialaws.eu/fragmenting-internet-governance-digital-sovereignty-and-global-constitutionalism/. Accessed 14th September 2020.

DeNardis, L. (2013). *Protocol politics: The globalization of Internet governance*. Boston: MIT Press

DeNardis, L. (2014). *The global war for Internet Governance*. London: Yale University Press.

DeNardis, L. and Musiani, F. (eds). (2016). From Governance of Infrastructure to Governance by Infrastructure, in Musiani, F.; Cogburn, D.L., DeNardis, L.; Levinson N. S., *The Turn to Infrastructure in Internet Governance*. Ney York: Palgrave Macmillan.

Doria, A. (2014). Use [and Abuse] of Multistakeholderism in the Internet. In Radu R., Chenou JM., Weber R. (eds), *The Evolution of Global Internet Governance* (pp. 115-140). Berlin: Springer.

Drake W. and Nicolaıdis K. (1992). Ideas, interests, and institutionalization. *International Organization*, 46(1), 37–100.

Drake, W.J. (2004). Reframing Internet Governance Discourse: Fifteen Baseline Propositions. In: D. MacLean (ed.). *Internet Governance: A Grand Collaboration* (pp. 122–161). New York: United Nations Information and Communication Technology Taskforce.

Dunlop, C.A. (2012). Epistemic communities. In M. Howlett, S. Fritzen, W. Xun, and E. Araral (eds), *Routledge Handbook of Public Policy* (pp. 247-261). London: Routledge.

European Commission. Artificial Intelligence for Europe. (2018). https://ec.europa.eu/digital-single-market/en/news/communication-artificial-intelligence-europe. Accessed 1st September 2020.

European Parliament. Civil Law Rules on Robotics. (2016). https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html?redirect. Accessed 1st September 2020.

Finnemore, M. and Sikkink, K. (1998). International Norm Dynamics and Political Change. *International Organization*, 52(4), 887–917.

Floridi, L. (2019). Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical. *Philos. Technol*. 32, 185–193.

Floridi, L., Cowls, J., Beltrametti, M. et al. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds & Machines,* 28, 689–707.

Fuchs, C. (2008). *Internet and Society*. New York: Routledge.

Gardbaum, S. (2008). Human Rights and International Constitutionalism. In Dunoff, J. and

Trachtman, J. (eds), *Ruling the World? Constitutionalism, International Law and Global Government* (pp.233–57). Cambridge: Cambridge University Press.

Goldsmith, J., Wu T. (2006). *Who Controls the Internet? Illusions of a Borderless World*. Oxford: Oxford University Press.

Gornitzka, Å and Host C. (2015). The expert-executive nexus: An introduction. *Politics and Governance*, (3), 1, 1-12.

Gornitzka, Å. and Sverdrup U. (2011). Access of experts: Information and EU decision-making. *West European Politics,* 34(1), 48–70.

Grimm, D. (2010). The Achievement of Constitutionalism and Its Prospects in a Changed World. In P. Dobner and M. Loughlin (eds), *The Twilight of Constitutionalism?* (pp. 3–22). Oxford: Oxford University Press.

Haas, P. M. (1992). Introduction: Epistemic communities and international policy coordination. *International Organization*, 46(1), 1–35.

Hofmann, J. (2007). Internet Governance: A Regulative Idea in Flux. In R. K. J. Bandamutha (Ed.), *Internet Governance: An Introduction* (pp. 74-108). Icfai: University Press.

Hofmann, J., Katzenbach, C. e Gollatz, K. (2017). Between coordination and regulation: Finding the governance in Internet governance. *New Media & Society*, Vol 19, Issue 9, 1406 – 1423.

IEEE. Ethically Aligned Design. (2016-19). https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead1e.pdf. Accessed 1st September.

Jørgensen, R.F. (2011). Human Rights and Their Role in Global Media and Communication Discourses. In A. Sreberny, R. Mansell and M. Raboy (eds), The Handbook of Global Media and Communication Policy (pp. 95-112). Oxford: Blackwell.

Khagram, S., Riker J.V. and Sikkink K. (eds.). (2002). *Restructuring World Politics: Transnational Social Movements, Networks and Norms*. Minneapolis: University of Minnesota Press.

Koenig-Archibugi, M. and Zürn, M. (2006). *New Modes of Governance in the Global System*. NY: Palgrave MacMillan.

Kurbalija, J. (2014). *An Introduction to Internet Governance*. Diplo Foundation.

Laprise, J. (2016). Exploring PRISM's Spectrum: Privacy in the Information Age. In Musiani, F.; Cogburn, D.L.: DeNardis, L.; Levinson N. S., *The Turn to Infrastructure in Internet Governance* (pp. 203-2019). New York: Palgrave MacMillan,.

Latzer, M., Just, N. K. (2020). Governance by and of algorithms on the internet: impact and consequences. In Nussbaum, Jon F. (eds), *Oxford Research Encyclopedia of Communication*. Oxford: Oxford University Press.

Leifeld, P. (2017). Discourse network analysis: policy debates as dynamic networks. In Victor, J. N., Lubell, M. N. and Montgomery, A. H. (eds.), T*he Oxford Handbook of Political Networks*. Oxford University Press.

Lessig, L. (2006). *Code 2.0*. New York: Basic Books.

Liddicoat, J., and Doria, A. (2012). Human rights and Internet protocols: Comparing processes and principles. https://www.unesco-ci.org/cmscore/sites/default/files/2013wsis10/human_rights_and_internet_protocols-_comparing_processes_and_principles28129.pdf. Accessed 1st September.

Luhmann, N. (2004). *Law as a Social System*. Oxford: Oxford University Press.

Majone, G. (1996). *Regulating Europe*. London: Routledge.

Mets, J. (2015). *The European Commission, Expert Groups, and the Policy Process*. London: Palgrave MacMillan.

Milan, S. and Oever, N. (2017). Coding and encoding rights in internet infrastructure. *Internet Policy Review*, 6(1). DOI: 10.14763/2017.1.442

Mueller,L.M. (2010). *Networks and States: The Global Politics of Internet Governance*. Cambridge MA:MIT Press.

Musiani, F.; Cogburn, D.L.: DeNardis, L.; Levinson N. S., (2016*), The Turn to Infrastructure in Internet Governance*. New York: Palgrave MacMillan.

Nye J. S. (2011). *The Future of Power*. New York: Public Affairs.

O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown Publishers.

O'Sullivan B. Towards a Magna Carta for Data, 2017. https://www.ria.ie/sites/default/files/ria_magna_carta_data.pdf. Accessed 1st September.

Padovani ,C., Musiani, F., Pavan, E. (2010). Investigating Evolving Discourses On Human Rights In The Digital Age: Emerging Norms And Policy Challenges. *The International Communication Gazette*, 72(4–5): 359–378

Padovani, C. and Santaniello, M. (2018). Digital constitutionalism: Fundamental rights and power limitation in the Internet eco-system. *International Communication Gazette,* 80, (4).

Palladino, N. and Santaniello, M. (2020). *Legitimacy, Power, and Inequalities in the Multistakeholder Internet Governance*. London: Palgrave MacMillan.

Peters, A. (2006). Compensatory Constitutionalism: The Function and Potential of Fundamental International Norms and Structures. *Leiden Journal of International Law,* 19, 579–610.

Radaelli, C. M. (1999). The public policy of the European Union: Whither politics of expertise? *Journal of European Public Policy*, 6, 757–774.

Redeker, D., Gill, L. and Gasser, U. (2018). Towards digital constitutionalism? Mapping attempts to craft an Internet Bill of Rights. *International Communication Gazette,* 80(4).

Rioux M. and Fontaine-Skronski K. (eds). (2015). *Global Governance Facing Structural Changes New Institutional Trajectories for Digital and Transnational Capitalism*. New York: Palgrave MacMillian.

Risse, T. (2006). Transnational Governance and legitimacy. In A. Benz and Y. Papadopoulos (eds.). *Governance and Democracy Comparing National, European and International Experiences* (pp.179-99). New York: Routledge.

Rosenau, J. and Czempiel E.O. (1992). *Governance Without Government: Order and Change in World Politics*. New York: Cambridge University Press.

Russel S. and Norvig P. (2009). Artificial Intelligence: A Modern Approach. Prentice Hall.

Santaniello, M., Palladino, N., Catone, M.C., Diana, P. (2018) The language of digital constitutionalism and the role of national parliaments. *International Communication Gazette*, 80(4), 320-336.

Schneiker, A. and Joachim, J. (2018). Revisiting Global Governance in Multistakeholder Initiatives: Club Governance Based on Ideational Prealignments. *Global Society*, 32:1, 2-22,

Sciulli, D. (1992). *Theory of Societal Constitutionalism*. Cambridge: Cambridge University Press.

Sebenius, J.K. (1992). Challenging Conventional Explanations of International Cooperation: Negotiation Analysis and the Case of Epistemic Communities. *International Organization*, 46:1, 323–65.

Star, S. L. (1999). The ethnography of infrastructure. *American Behavioral Scientist*, 43(3): 377–391.

Suzor, N., Dragiewicz, M., Harris, B., Gillett, R., Burgess, J. and Van Geelen, T. (2019). Human Rights by Design: The Responsibilities of Social Media Platforms to Address Gender-Based Violence Online. *Policy & Internet*, 11: 84-103.

Teubner, G. (2002). Breaking Frames: Economic Globalisation and the Emergence of Lex Mercatoria. *European Journal of Social Theory*, 5, 199-217.

Teubner, G. (2004). Societal constitutionalism: Alternatives to state-centred constitutional theory? In Joerges, C., Sand, I.Y. and Teubner, G. (eds), *Transnational Governance and Constitutionalism* (pp. 3–28). Oxford: Hart Publishing.

Teubner, G. (2011). Transnational fundamental rights: Horizontal effect? *Netherlands Journal of Legal Philosophy,* 40(3), 191–215.

Teubner, G. (2012). *Constitutional Fragments: Societal Constitutionalism and Globalization*. Oxford: Oxford University Press.

Teubner, G. (2017). Horizontal Effects of Constitutional Rights in the Internet: A Legal Case on the Digital Constitution. *The Italian Law Journal*, (03) 01, 193-205.

Tsebelis, G., and Xenophon Y. (2002). Veto Players and Decision-making in the EU after Nice: Policy Stability and Bureaucratic/Judicial Discretion. *Journal of Common Market Studies*, 40, 2, 283–307.

Turner, J. (2019). *Robot Rules – Regulating Artificial Intelligence*. London: Palgrave.

van Dijck J. (2014). Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & Society*, 12(2), 197-208.

van Dijck N. and Casiraghi S. (2020). The “Ethification” Of Privacy And Data Protection Law In The European Union. The Case Of Artificial Intelligence. Brussels Privacy Hub Working Paper Vol. 6, N. 22.

Van Eeten, M.J. and Mueller, M. (2013). Where is the governance in Internet governance? *New Media & Society,* 15,(5), 720–736.

Villani C., Bonnet Y., Berthet C., Cornut AC, Levin F., Rondepierre B. (2018). For a meaningful artificial intelligence. https://hal.inria.fr/hal-01967551/document. Accessed 27 July 2020.

Wet, E.D. (2006). The International Constitutional Order. *International and Comparative Law Quarterly*, 55, 51-76.

Yeung K., Howes, A. and Pogrebna G. (2019). AI Governance by Human Rights-Centred Design, Deliberation and Oversight: An End to Ethics Washing, In M. D. Dubber, F. Pasquale, and S. Das, *The Oxford Handbook of AI Ethics*. Oxford: Oxford University Press.

Zalnieriute, M. (2015). An international constitutional moment for data privacy in the times of mass-surveillance. *International Journal of Law and Information Technology*, 23, 2, 99–133.