Manifesto from Dagstuhl Perspectives Workshop 18262

10 Years of Web Science — Dagstuhl Manifesto

Edited by

Bettina Berendt[1], Fabien Gandon[2], Susan Halford[3], Jim Hendler[4], Katharina Kinder-Kurlanda[5], Eirini Ntoutsi[6], and Steffen Staab[7]

1 **KU Leuven, Belgium.** Bettina.Berendt@cs.kuleuven.be
2 **Inria, Université Côte d'Azur, I3S, CNRS, France.** fabien.gandon@inria.fr
3 **University of Bristol, UK, susan.halford@bristol.ac.uk**
4 **Rensselaer Polytechnic Institute,** hendler@cs.rpi.edu
5 **GESIS - Leibniz-Institut für Sozialwissenschaften, Cologne, Germany** katharina.kinder-kurlanda@gesis.org
6 **Leibniz Universitaet Hannover & L3S Research Center,** ntoutsi@l3s.de
7 **Institute WeST, Universität Koblenz-Landau, Germany and WAIS, University of Southampton, UK,** staab@uni-koblenz.de

Abstract

This is the Manifest of the Dagstuhl Perspectives Workshop 18262 on "10 Years of Web Science". At the Workshop, we revisited the origins of Web Science, explored the challenges and opportunities of the Web, and looked forward to potential futures for both the Web and Web Science. In this Manifesto, we address those futures through the lens of Artificial Intelligence, which will be key to the evolution of the Web over the coming decade. We begin by outlining the key ambivalences that lie at the heart of the Web. For all the enormous benefits that the Web offers - for information sharing, collective organization and distributed activity, social inclusion and economic growth - the negative consequences are only too apparent. Artificial Intelligence has the potential to exaggerate these outcomes - in both directions - and we explore these possibilities, situating them within the wider debate about the future of regulation and governance for the Web. Finally, we outline the need to extend Web Science as the science that is devoted to the analysis of Web, to strengthen our role in shaping the future of the Web and present the key strands of capacity building that are necessary to achieve this.

Perspectives Workshop 25.–29. June, 2018 – www.dagstuhl.de/18262
2012 ACM Subject Classification Dummy classification
Keywords and phrases Dummy keywords: Please provide 1–5 keywords
Digital Object Identifier 10.4230/DagMan.1.1.1

A Manifesto Sample, *Dagstuhl Manifestos*, Vol. 1, Issue 1, pp. 1–28
Editors: John Q. Open and Joan R. Access

Dagstuhl Manifestos
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Executive Summary

Web science is the interdisciplinary study of the World Wide Web. We begin from the premise that the Web operates as a complex and dynamic sociotechnical system and - as such - demands integrated expertise from the engineering and social sciences and the humanities if we are to understand its past, present and potential futures. In this Manifesto, we address those futures through the lens of Artificial Intelligence, which will be key to the evolution of the Web over the coming decade. We begin by outlining the key ambivilences that lie at the heart of the Web. For all the enormous benefits that the Web offers - for information sharing, collective organization and distributed activity, social inclusion and economic growth - the negative consequences are only too apparent. Artificial Intelligence has the potential to exaggerate these outcomes - in both directions - and we explore these possibilities, situating them within the wider debate about the future of regulation and governance for the Web. Finally, we outline the need to extend Web Science, to strengthen our role in shaping the future of the Web and present the key strands of capacity building that are necessary to achieve this.

Table of Contents

1 Introduction

30 years have passed since Tim Berners-Lee proposed a new architecture for information sharing for Physicists working with the high energy Physics laboratories at CERN. Shortly afterwards, this came to be known as the World Wide Web. And the rest is history. The Web has evolved from a draft proposal to a global infrastructure, with over 4 billion regular users and shapes the lives of those who do not even know of its existence [footnote: for example influencing commodity prices, global politics and climate change research].

[centre with lines above and below]The Web has become the nervous system for our planet.

This was an unexpected journey. What was originally imagined as a technical system is now deeply embedded in and shaped by the social world. Web architecture and infrastructure are technical, but the Web cannot be understood as only technical. The Web is made by social activity — by individuals, communities, governments and businesses — but it cannot be understood as only social.

In 2006, Berners-Lee[4] proposed a new interdisciplinary field of study 'Web Science', to research how the Web was evolving and what might be done to protect its future. Already, it was clear that the Web was implicated in some fundamental social and economic transformations — and posed some unexpected challenges, for example cybercrime, hate-speech and the increasing centralisation of content and infrastructure.

Over the past decade, Web Science has grown and spread across the world[1], integrating expertise from across the engineering sciences, social sciences and humanities. Web Scientists have generated new knowledge and understanding of how the Web has changed the world, and how the world has changed the Web.

The World ⇌ The Web

For example, Web Scientists have documented atrocities in information warfare [28, 29], identified how fake news is influencing our political landscapes [25, 26] and developed methods to identify individual cybercriminals on the dark web CITE MISSING(van hardeveld et al 2018).Web Scientists have traced how the Web generates new business models [add reference], enabled the effective use of big data and web infrastructures for peace-building in fragile states (Gaskell 2018) and built tools to analyse web data at scale and speed for a wide variety of purposes (ref).

And still the Web evolves. In the years since Web Science was established, we have seen the emergence of social media, a fully fledged data economy, revelations of mass surveillance and interference in democratic elections. And now, a new wave of Artificial Intelligence, spurred on by the phenomenal data resources created (in large part) by the Web, has begun a new round of transformations that mark a step-change in the Web of the future.

The Web ⇌ Artificial Intelligence

The Web has both facilitated the current and rapid growth of Artificial Intelligence as a field and - in turn - will be increasingly shaped by a variety of artificial intelligences, developed by researchers in academia, in business and in government. Looking towards the future, we see a Web that will be comprised of a broader range of actors than ever before,

[1] http://wstnet.webscience.org.

as humans, artificial intelligences and other technologies interact shaping each-other and shaping outcomes across the globe.

Exactly how this will happen is, as yet, undetermined. On the one hand, there are very real concerns, for example AI is radically extending the propagation of misinformation, of biases in information distribution and decision making and further extend social polarization, harassment and hate speech. On the other hand, there are also opportunities for AI to mobilise against these activities and support a Web that is more inclusive and collaborative than ever before.

The future of the Web is a deeply sociotechnical question. Web Science is essential if we are to address these concerns and harness the opportunities for an AI Web.

Web Science ⇌ The Web of Humans and Artificial Intelligences

This Manifesto for Web Science addresses the future of the Web as follows. In Section 2 below, we outline the current challenges facing the Web, in general, and with particular reference to Artificial Intelligence. In Section 3, we outline how Web Science can address these challenges, by shaping the development of AIs for the Web in a socially responsible way. In Section 4, we widen out to consider the wider landscape of governance and regulation that will be required in order to support a fair and inclusive AI Web for the future. The interdisciplinary capacity of Web Science is critical to these endeavours. Without well-developed collaboration that draws together all the necessary forms of expertise, we cannot expect this to succeed. We have learnt a great deal over the past decade about how difficult this is to do and what it takes to do this effectively. In Section 5 we explore how we can build on this, to support research practices that integrate expertise across the disciplines to deliver the next decade of Web Science. In Section 6, we turn attention to how this capacity should be supported by academia, funding authorities and governments.

2 Ambivalences of The Web

Whilst the Web is a mighty instrument for individual, institutions and society, as an instrument it can be put both to beneficial or to detrimental use. For all the enormous benefits that the Web offers for information sharing, collective organization and distributed activity, social inclusion and economic growth, the negative consequences are only too apparent. It is a core task for Web Science to analyse, understand and navigate these ambivalences of the Web, amplifying the good and countering the bad.

2.1 Information Freedom vs. Information Quality

There is a very long history behind the idea that the sharing of ideas and of evidence-based information is positive for social inclusion, productivity and well-being in society.[2] The printing press greatly reduced the costs of information dissemination and allowed the sharing of ideas and evidence that supported the development of modern science. The invention of the Web further reduced these costs by many orders of magnitudes democratizing access to formal knowledge (open science, free textbooks, historical sources) and informal knowledge

[2] http://www.rogerclarke.com/II/IWtbF.html

(recipes, reviews, etc.). Both the production and consumption of information such as political news and scientific results has been greatly widened, accelerated, which is *prima facie* a positive outcome.

In the pre-Web era, publishers were gatekeepers. Their role could be ambivalent, too, as they could exert quality control or censorship on content. In the era of the Web, it is inexpensive to become a publisher. While it is easy to publish facts and judgments, it is equally possible to widely disseminate mis- and disinformation, unfounded beliefs and prejudices.

Mis- and disinformation existed before the Web, but the Web has facilitated its distribution, at scale and speed. Artificial Intelligence will further amplify the problem. It has been a core principle of modernity to question printed information, but to believe in original sources, such as photography, audio or video. Artificial Intelligence allows for the creation of *deep fakes*, *i. e.* video or audio that cannot be distinguished from true video or audio, but that is completely made up by its creator.

2.2 Personalization vs. Privacy

Core to the success of the Web is the convenience that it offers. Using our personal devices, information consumption and production has become frictionless. One or two clicks are sufficient to buy, to like or to share, and service providers know about our preferences facilitating selection of music, audio, restaurants or other products from swathes of offerings. Such *personalization* is made possible through *online behavioural tracking mechanisms* able to harvest minute details of online activity. A variety of actors aggregate online activity across populations in order to analyze, profile and serve targeted information. The more detailed the tracking, the smoother our experience of the Web, and the higher the revenues for the corporates.

While every small improvement derived from closer tracking and more precise analytics turns into a sizable business benefit for the trackers motivating them to ever more comprehensive monitoring of our lives, there are diminishing returns yielded to the individual. The consequence is a deep intrusion into the privacy of individuals, what they do and like, what their political beliefs are, whom they do business with, and whom they spend their days and nights with. At best this intrudes people's privacy and exploits it without fair retaliation, at worst it may leads to misuse manipulating individuals or deriving damaging conclusions about them.

Consider a case of manipulation in 2014. Facebook faced a storm of protest after it was revealed that via an *online psychological experiment* they were attempting to make users feel happier or sadder by controlling their feeds. This incident provoked widespread public concern regarding the effect of such experiments and interventions and the lack of agreement on expertise and ethics knowledge about how to do Web experimental research. Consider the topic of possibly damaging conclusions as demonstrated in early experiences with prototypes of a planned Chinese social scoring system. There have been reports that individual's political opinion and even the political opinion of their friends influence the score and hence freedom of movement, ability to work, or chances to date other people.[3]

[3]Cf., e.g., https://www.wired.co.uk/article/china-social-credit-system-explained on the ambivalences of the Chinese social scoring systems.

Artificial intelligence may be prone to nurture both types of damages. People who move houses more often or who have a less regular telephone usage behaviour are considered worse credit risks[27]. Telematics are used to calculate our car insurance. Thus, Artificial intelligence may be used for Big Nudging. Could there be positive outcomes of the same technology? Yes, maybe Big Nudging would increase cycling and walking and, thus, increase our health, too.

2.3 Influence by the Masses vs. Manipulation of the Masses

The Web offers the opportunity for countless people to connect across geographical boundaries in ways that were difficult to imagine in the context of previous media systems. The resulting network allows for a scale and speed of information dissemination that has since the inception of the Web been seen as a huge opportunity for democratization, for example by allowing to widely publicize injustices. A current example is the Rezo video on Youtube which allowed a Youtuber to influence the German political discussion on climate change and political inactivity. https://www.spiegel.de/wissenschaft/mensch/rezo-video-die-politik-lullt-ein-und-die-jugend-wehrt-sich-a-1269173.html).

The Web also allows to tap the 'wisdom of the crowd' by sourcing its seemingly infinite variety of connected entities. However, this network has from its beginning always held surprises and developed in unexpected ways and while we have made huge inroads in understanding phenomena such as homophiliy, there is still a lot to learn about how online crowds behave. This is especially the case if we look at the less positive aspects of the Web as a public network. Crowd sourced judgment can be good or bad, depending on how you want to see the world [23].

For example, the way in which crowd behaviour may be manipulable or at what point a righteous public outcry may turn into undesirable and vindictive pursuit are not well understood. And both the positive and the negative sides of the Web may be amplified by the possibilities that algorithmic analyses and AI bring. For example, in 2018 it was revealed that Cambridge Analytica had harvested the personal data of millions of people's Facebook profiles without their consent and used it for political purposes (*the Facebook–Cambridge Analytica data scandal*), like targeting voters in the US presidential election [4]. This was a turning point in the public understanding of personal data and caused many calls for tighter regulation of tech companies' use of data. It also showed that there is a potential to influence large amounts of people - and uncertainty as to how exactly such influencing works, may already be at work and who may be trying to influence people.

These uncertainties also became clear in connection with the Brexit elections when it turned out that bot accounts may have been used to boost follower numbers for parties on Twitter. (Researchers Say Many Of The Brexit Party's Twitter Followers Aren't Behaving Like Genuine Voters https://www.buzzfeed.com/alexspence/nigel-farages-brexit-party-twitter-following) This example highlights the uncertainty about the influence of bots: their actual impact is uncertain as is their power to influence opinions - there certainly seems to be large potential for it.

The same mechanisms and possibilities that the interconnected networks of the Web offer may allow to publicise both desirable and undesirable content, to manipulate for the 'good'

[4] https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election

or for the 'bad' depending on your viewpoint and give influence to people who may not have had a voice without the Web.

2.4 Inclusive and Fair vs. Exploitative

We want the Web to be inclusive. This means that all individuals, but also all institutions — companies, associations, governmental or non-governmental organizations — can make best use of it, contributing their content and data and being able to access it.

While this may sound simple enough, it is not. Even low costs of internet access might be too high for some. In this context, Facebook's internet.org initiative proposes to offer for free a limited version of the Web with access to specific Web sites like Wikipedia as well as to Facebook and Facebook partner businesses. However, the Indian government has banned this initiative, which gives Facebook the power to act as a monopolistic gatekeeper to the internet, such that only businesses who partnered with Facebook would be able to thrive in India. At the same time, this initiative would have given many people access to the internet opportunities they seek and any alternative will be more expensive and therefore more exclusive.

The Web as a system of people, platforms, and economic affairs naturally tends towards such a monopoly that internet.org wants to create. Indeed, other monopolies or oligopolies exist in the Web, e.g. for mobile phone operating system, search, social networking, business networking, shopping, or video hosting, where companies have created free offers that attracted a majority of people rendering further competition economically unfeasible. While it seems fair enough that these companies draw a profit from their investments, once these monopolies/oligopolies have grown large enough, they play the role of gatekeepers that can exploit their customers as well as the existing infrastructure and business partners almost at will.

There are fundamental problems in defining what constitutes fair behaviour and what does not and the fairness of economic behavior cannot be disentangled from the other questions raised before, which were the ambivalences of personalization versus privacy violations or wisdom of the crowd versus mass manipulation.

Artificial Intelligence seems to amplify this problem. Progress in AI depends on the availability of outstanding researchers and availability of data, both of which are found at the Web behemoths more than anywhere else, and AI allows for those value-added services that are likely to attract even more customers thus nurturing the spiral that strengthens existing monopolies. There is an active discussion as to when AI systems are fair.[5]

Regulations that aim at creating fairness for all, thus should not only consider the individual, but they should consider inclusiveness and fairness for all groups. Small companies and social groups may suffer most from regulations that large companies can deal with. These authors were briefly involved in running a truly free social networking site, Metalcon, with more than 10,000 users. That social network could not be maintained in spite of its attraction to new visitors and in spite of voluntary work, as legal compensations for copyright breeches provoked by its users far exceeded its minor income.

Further discussions concerning fairness versus exploitation are ongoing considering issues like *net neutrality* (especially vivid in USA) or *upload filters* (especially vivid in Europe).

[5] E.g. consider the ACM FAT* conference series on Fairness, Accountability, and Transparency in socio-technical systems.

This should not come as a surprise. While our societies may have improved fairness over the last centuries at large, there is an ongoing need for improving understanding and discussing as to what constitutes (un-)fairness — also in the Web.

2.5 Growing vs Sustainable

The Web happened during the anthropocene, the current geological age during which human activity has been strongly influencing the climate and the environment. As a global artefact the Web needs to grow to cover and connect ever more people, usages, cultures and resources. But leading the web to its full potential requires us to strike a balance between its growth and its costs of all sorts: the economical cost of the infrastructure it requires (e.g. Internet, telco networks) and the maintenance of its resources (e.g. maintaining and evolving Wikipedia), the ecological cost of its deployment and use (e.g. Google post on powering a search[6]), the educational cost of the experts it requires (technical aspects, content production, legal framework, etc.)
And as we will see in section 3, the encounter of the Web with AI can both amplify or alleviate this problem as some AI techniques can be used to downscale the resources needed to operate Web applications [22] and at the same time some of the AI techniques are consuming a lot of energy [31]. This is also the case with other technology the Web is linking to such as blockchain, the internet of things, mobile and ubiquitous computing, to name a few.
The emergence and scaling of new capabilities and usages of the Web are not only questioning our ability to sustain the evolution of the Web in terms of energy and money but in every domain. In the educational domain the Web curricula keep growing with new topics, standards, frameworks, tools, etc. In the legal domain each evolution requires the legislator to monitor and control it closely. In other words it is important to not only consider the full potential of the Web also the one we can afford[7].

2.6 Ambivalences without an End

In this section, we have discussed ambivalences of the Web today and how they will be amplified by Artificial Intelligence. It is germane to the Web as a powerful instrument that such ambivalences that affect human well-being for the better or the worse will never disappear. However, we, using Web Science as a discipline, must ask ourselves how to deal proactively with these ambivalences. In the following section, we will try to address this question.

3 The Web linking all Forms of Intelligence

Half of humanity uses the Web directly, and nearly all people are affected by it. But this is only a fraction of the agents who contribute to the Web. More and more human and artificial forms of intelligence are connected to the Web and participate in its content, structure,

[6] https://googleblog.blogspot.com/2009/01/powering-google-search.html
[7] https://www.w3.org/community/wwca/

services and all the activities it supports. This section focuses on the potential and challenges of a Web that links all forms of intelligence.

The first section introduces this evolution as a continuation of the historical trend toward a Web of everything and on everything. We then divided the challenges in evolving Web Science to study a Web that includes AIs into three sections: challenges at the individual level (e.g. a user, a software agent), challenges at the level of collaborations of different forms of intelligence (e.g. hybrid societies over social media), and cross-domain challenges to highlight the inter-disciplinary nature of the research needed. Before closing this section we stress the specific potential and challenges in using AI to study and manage the Web and help us conduct Web Science research at Web scale.

3.1 A historical trend toward a Web of everything on everything

When the Web started to spread at the beginnings of the 90s, it was essentially perceived as a hypertext system distributed over an open set of servers, and in particular over the Internet. Susan: not by social scientists - I realise that its too late now, but social scientists saw the Web as a culture, or as a community - different metaphors It was described using metaphors of a universal library formed by linked HTML documents called Web pages and primarily for human consumption. But, as soon as it started to spread, it also started to reveal all the potential of the distributed hypermedia software architecture it had globally standardized [18]. Although the Web was initially open to contribution (HTTP PUT and POST) this was not implemented in the first browsers and we had to wait until the mid-90s, with wikis and forums, for the Web to be re-opened in read-write mode. This paved the way to numerous new social media applications. In parallel, the initial use of programs to generate pages on the fly (e.g. CGI used for first search engines or Web applications) opened the way to a second major evolution of complex Web applications and dynamic pages leading to a deep Web far exceeding the initial librarian metaphor and effectively linking programs [15]. As programs became more and more connected to the Web the need to exchange other data than pages became clear [3], and the Web first separated form (CSS) and content (HTML) to then be able to exchange different types of content (XML, RDF, JSON) and ultimately became a place for publishing and linking datasets. Both trends made the Web more machine-friendly and supported the development of Web-connected devices starting with mobile access to the Web and leading to the more recent notion of a Web of Things where the Web is now in position to provide a universal software architecture and framework to access the Internet of Things (IoT) and program its applications and interfaces.

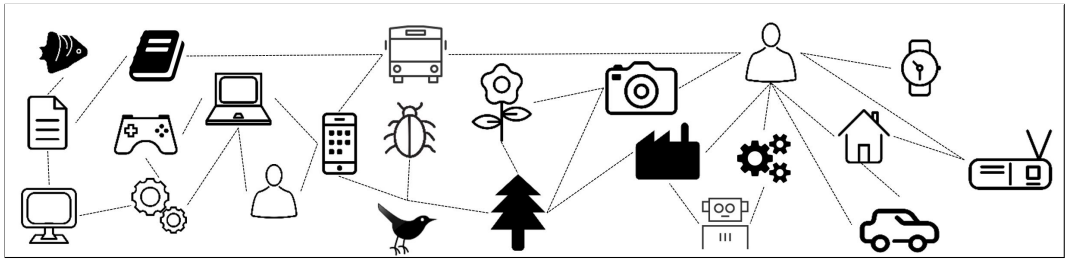

Figure 1 a Web of everything and on everything

Despite these different evolution trends, it is remarkable to see that the Web architecture remains one and unique. One Web is now linking documents, people, data, programs,

and virtually anything that can be identified and one Web is accessed via multiple means: computers, smartphones, watches and a growing heterogeneity of connected things. Together "Web of everything" and "Web on everything" are two viral and synergistic evolution trends of the Web: every new device connected to the Web provides more data and more resources to access and to link and, in return, every new piece of data linked to the Web supports new applications and provides resources for the connected devices. Fabien: In section 2.1 we identified the tension between the freedom to contribute information and the need to ensure the quality of the information we find on the Web and we can now extend that challenge to everything the Web will connect raising the problems of searching and finding relevant resources on the Web while ensuring security, certification and authentication in these connections we make.

As a result the Web has become a universal collaborative space for actors that consume and produce information including all different forms of natural and artificial intelligences. Web Science was set up to understand the Web and evolve it for the benefit of society, including the generation of opportunities and the protection of the vulnerable. The linkage of diverse forms of intelligences necessitates accelerated development of Web Science as a discipline. As Artificial Intelligence augments the means of the rich and powerful, Web Science must be developed to leverage opportunities for society as a whole and to protect weaker member of societies against negative ramifications of a Web of linked intelligences.

3.2 Challenges at the individual level: a variety of intelligence forms hosted on the Web

Figure 2 connecting and linking a variety of intelligence forms on the Web

A first challenge we can identify is the one of designing artificial Web intelligence i.e. the study of the specific problems in designing AI agents that are connected to the Web – its resources and users – with all the heterogeneity, different scales and speeds this involves. AI connected to the Web must be robust and benevolent by-design as they are potentially in contact with billions of users.

Challenge – Benevolent: to ensure AI brings no harm to Web users and to make Web AI web bots benevolent by design.

A short term objective is the study of new AI approaches to traditional Web tasks that can benefit from AI, such as indexing, searching, or browsing but also more complex tasks such as fact checking, fraud detection, or protection enforcement to support privacy. Fabien: To the technical challenges identified in this section we should add upfront the extension of the tension identified in section 2.2 between the ability to adapt and the need to respect privacy and confidentiality: the more AI techniques are deployed on the Web and its data the more important it becomes to monitor and control the usage of data they consume and produce. Inversely but still on a short term perspective, the Web has the opportunity to become the most important source of datasets to feed data-hungry AI and the challenge here

is to ensure the Web can provide high quality datasets with the metadata required to ensure trust and proper usages.

Challenge – Web intelligence: include more intelligence in classical Web tasks and more Web resources in classical AI methods.

Steffen: The argument 'different AI techniques cater to a diversity aware Web' falls short. I could easily imagine various agents being misogynistic, while they might use diverse techniques of ML or KR in the background. I do agree that combining different techniques might be a measure to contain, e.g., misogynistic AI. Fabien: Here the diversity of the Web refers to the heterogeneity of the Web resources and actors in general and the fact a diversity of methods is needed to manage them in a reference to the Law of Requisite Variety of Ashby The diversity of AI techniques in existence is an asset as it allows to address many types of diversity we find on the Web in terms of content, users, contexts, tasks, usages, resources, etc. Doing so will require the adaptation and extension to the Web of different forms of AI – machine learning, knowledge representation and reasoning, constraint solving, etc. All the domains of AI – symbolic AI, connectionist AI – should be considered in relation to the Web: the next revolution may come from a currently very quiet field and the broad picture of the Web requires all types of expertise.

Challenge – Web-robust: design a variety of AIs robust to the Web, its heterogeneity, different scales and speeds, its noise, open-world nature and uncertainty.

Another challenge is to harness the multidisciplinary domains of Web Science to identify new ways of simulating, reproducing or engaging intelligence including emotional intelligence, communication skills, or imagination. A diversity of intelligent skills and new ways to simulate them in the open-world context of the Web are needed. On this point the Web is both raising questions and providing first answers by proposing a universal space to link different forms of intelligent processing such as knowledge-based systems, statistical learning, learning by reading, wisdom of crowd, etc. Many different sources of intelligence can be found or connected to the Web.

Challenge – Intelligence diversity: understand and harness the diversity of intelligent skills and behaviours that we need on the Web.

Vice versa, the advent of AI on the Web requires Web Science to evolve toward the studying of a Web that includes AI agents as first class citizens. The first challenge is to make the Web "AI-friendly" at the core of its architecture by providing knowledge level means to connect an AI to the Web. Beyond semantic Web, linked data and knowledge graphs formalization and models are needed that support the publication and exchange of the many types of data an AI may want to obtain or provide on the Web including: input data, training data, output data/results, parameters, configurations, schemata, embeddings, etc. All kinds of semantic networks can potentially be exchanged by AIs on the Web including complete neural networks, layers or sub-networks.

Challenge – AI-friendly: augment the Web architecture and standards to turn the Web into an environment fully supporting the hosting of AI agents.

Compared to other AI domains an important specificity of AI on the Web is that it will systematically be in contact with a huge variety of human users. Human-AI interaction on the Web is an important challenge for interaction design and HCI. In Web-based interaction

with AIs many users can interact with many different AI agents using and combining multi-modal means such as written chat, voice recognition, gesture recognition, multi-touch, etc. Morerover the Web also holds the potential to augment interactions in particular by providing background knowledge and contextual data that can feed AI assisting smarter interactions and supporting, for instance, context awareness [20]. In other words, the challenges are both to use the Web for better AI interaction and to use AI for better interactions with the Web.

Challenge – user experience: to design advanced human-AI interaction on the Web and to assist human-Web interactions with AI support.

Finally a special case of interaction with AI is the generation of explanations and justifications to support trustable human-centered IA on the Web. The challenge is not only to generate single explanations but to weave a Web of explanations and justifications allowing users to follow there noses in linked traces possibly distributed over the Web (distributed data sources, distributed processing) to understand and take ownership of the results and to achieve decision making.

Challenge – build trust: provide user-friendly explanations, justifications, and traces of AI activities and results to foster trust and support decision making.

3.3 Challenges at group level: collaborations of different forms of intelligence on the Web

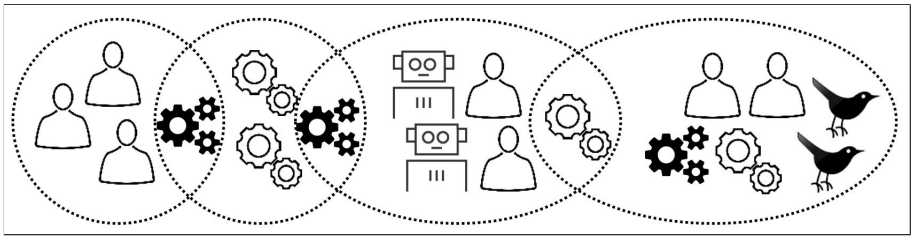

Figure 3 forming hybrid societies of intelligence forms on the Web

The simple fact that complex tasks combine several intelligent abilities – learning, reasoning, planning, solving, analyzing, extracting, communicating and collaborating – shows that Web Science will have to consider the Web evolution toward a multi-AI system. This evolution is both required and allowed by the distributed Web architecture leading us to the challenges of building and managing collective intelligence on the Web.

The long term challenge for Web Science is the study and design of human-agent collectives and social machines including AIs that form hybrid Web societies which include very different types of intelligence: people, connected animals, connected plants, artificial intelligence that reasons, that learns, or that induces etc. Consequently we need to study on the Web various combinations of possibly collective, interactions and their combinations: human-human, human-AI and AI-AI. We also already mentioned that these hybrid communities will require new means to bridge the semantics used by machines and the semantics used by humans in their interactions [19]. The challenge will also be to study the interactions of Web AIs with the resources of the Web (linked pages, linked data, connected objects, etc.) forming the environment of these forms of intelligence. Moreover, beyond individual interactions, a

challenge will be to study and design the orchestration of these multiple interactions. Web Science will have to face the problem of this massive interaction design with the Web and everything it links and AI will have to face the problem of engaging in very different types of interactions with different forms of intelligence including different kinds of AIs.

Challenge – hybrid communities: support the formation of groups of human and artificial agents and the interactions within these groups.

Web science must also consider the different types of groups that will be formed, from swarms to complex societies with their normative rules, roles, social constructs, etc. Fabien: In section 2.3 we mentioned the tension between the wisdom we can gain from the social activity of the Web and the risk of mass manipulation. This perspective of hybrid communities is amplifying that challenge of fostering the beneficial social phenomenon and preventing the dangerous ones as we are putting together billions of users and millions of software agents. For instance one could study the role AI can play in human-agent teaming and the role of a personal assistance to team members. Moreover the notion of norms raises the question of having "laws of Web AI" since a Web AI is at least as dangerous as a robot in terms, for instance, of harming users by learning and disclosing sensitive knowledge for instance. A complex question for Web science would then be to incentivize the development of benevolent Web AI through ethical principles of Web Science research and the study of normative systems for hybrid societies. Going even further, one could envision the design of Web AI to observe other Web AI, watchdogs checking the behaviours of other AIs to detect and protect against malicious or defective AI (e.g. AIs with a bias).

Challenge – social contracts: support the establishment and enforcement of social rules governing human and artificial agents behaviours within their hybrid communities.

To help communities life cycle and routines and to help users in participating in large scale groups with large collections of resources, Web Science should study Web-dedicated AI taking part in online social activities. These goal-driven agents could actively participate to the online activity and, for instance, foster linkage, interactions and convergence, bridge, translate, check, or augment our posts and maintain for us an overview of our social context and activity. They could also prevent or report problems such as bullying, harassment and polarization. Each of these task require dedicated intelligent processing that would need to be resilient to the variety of users, (mis)usages and content one can find on the Web. Ultimately, a very ambitious goal would be to have AIs with social and psychological skills to help humans face humanity on the Web and help us, individuals, scale to the world-wide web scale.

Challenge – online intelligence augmentation: support human users online by helping them understand the complexity of the Web and, behind it, the world and humanity.

This macro perspective also shows that Distributed AI (DAI), in particular, has a *rendez-vous* with the Web and its sciences because the distributed nature of the Web calls for distributed AI approaches. Multi-agent systems and distributed AI blackboards are examples of distributed AI architectures which, if merged with the Web architecture would allow for many different kinds of AIs to collaborate worldwide to the benefit of the Web. The AIs and the multi-agent systems would also in return benefit from the Web, its resources, its standards, its users and its methods. A conceptual challenge is therefore to study the

merging of these architectures to make the Web DAI-friendly and vice-versa as, so far, these approaches are more often juxtaposed than merged [6]. Fabien: Moreover, as we saw in section 2.5 we need to address the sustainability of the growth of the Web and of AI techniques and we can add to the need for new social contracts a need for the Web to respect a natural contract [30] with our environment.

Challenge – distributed intelligence: bring together the architectures of distributed AI and the architecture of the Web.

4 Governance and Regulation of People and AIs in the Web

The question of how the just mentioned challenges can be addressed at a global level leads to questions of governance. Steffen: I am slightly skeptical that this is true for most of the above mentioned challenges. Governance is a big question but not necessarily for these agreed - a slight change in language would help – the challenges raise questions about governanananceWeb governance is complex because moral values, jurisdictions and even fundamental rights differ across the globe and also change over time. Different values may have to be balanced against each other. What is more, governance itself is a complex undertaking, especially as much that happens on the Web is not governed by laws alone.

4.1 Landscapes of governance

Steffen: How about putting this sentence into the very beginning of section 4 ? Governance comprises all of the processes of governing, whether by a government or other actors, over a social system, through laws, norms, or power (based on [5]).

Internet Governance is an important component of governance for the Web and it comprises "the development and application by Governments, the private sector and civil society, in their respective roles, of shared principles, norms, rules, decision-making procedures, and programmes that shape the evolution and use of the Internet." (https://publicadministration.un.org/en/internetgovernance) Within this wide realm, Web-related questions center on how much influence each sector of society should have on the development of the Web and how issues such as freedom of speech and censorship, attempts to influence elections, misinformation, hate speech and cyber-bullying, data protection, privacy violations and surveillance should be approached. However, with the rise of collaborations of different forms of intelligence, the laws, norms, powers and languages governing AI are more and more relevant for how the Web evolves.

Laws have been and are being implemented by states and supra-national bodies such as the European Union. These laws govern what is legal (or not) on the Web and how violations of these rules can be sanctioned. An example is the General Data Protection Regulation (GDPR) that came into effect in the European Union in May 2018 and that regulates also non-EU actors if they process data of EU citizens (this includes search and social-network giants). At the same time, some EU member states, the US, China and other countries have laws that, e.g. in the stated interest of national security or to help citizens to trust one another legalise or mandate mass surveillance based on big networked data.

Decentralisation, openness, and freedom of speech have been founding principles and values of the Web community. This tradition by itself frequently clashes because a) certain

restrictions on speech are key elements of most jurisdictions [11] and b) these restrictions differ between countries and cannot easily be enforced globally.

It therefore appears reasonable for states to try to ensure that the laws that govern society anyway ("in the offline world") are also followed "in the online world". The German Network Enforcement Act (Netzdurchsetzungsgesetz (NetzDG) in force since 2018) is a principled approach along these lines designed to enforce a functioning complaints management system in large social-network platforms, imposing heavy fines if these do not remove content that has been flagged to them as violating a law (such as various laws against hate speech that predated the Internet). Other countries and the EU itself have implemented or are debating laws specifically against misinformation, hate speech, or cyber-bullying [13]. In the wake of allegations concerning election manipulation, some countries have decided to regulate speech in pre-election times, and to address specifically influence by foreign-national actors [8].

These content-restricting laws are often condemned not only by free-speech Web advocates, but also by voices from the US, a country that treats freedom of speech as a near-absolute fundamental right, e.g., [16]. European countries, based on historical experiences, balance this right more strongly against other fundamental rights in the interest of protecting individuals and groups. Restrictions on which content may be distributed via the Web may also be based on a balancing against other interests and rights, such as intellectual property and copyright. Laws that have created wide-scale protests such as the EU Directive on Copyright in the Digital Single Market (2019, not yet entered into force) attempt to enforce proportionality in these balancing acts.

Norms are developed and enacted at different levels. One is corporate governance and self-regulation. For example, with the help of human fact-checkers and automated classification, social media companies have tried to combat the spread of misinformation across their platforms, sometimes in efforts to obviate legislation.

One advantage of non-legal governance processes is that ethics-related desiderata that are difficult or impossible to formalise into law can be addressed. This may not be instigated by management, but also or only by workers and employees. Recent examples are the protests by Google employees against the company's participation in weapons development, or the price surges created by Uber drivers in an attempt to raise their earnings to acceptable levels and raise public awareness of their working conditions.

NGOs and not-for-profit organisations are further types of governance actors. The focus may be on how to govern the organisation itself (an example is the Wikimedia Foundation Governance system) or on monitoring other actors (an example is AlgorithmWatch).

Ethics-related desiderata and governance norms based on them may also arise via codes of conduct that professional associations give themselves (relevant examples for Web Science include the AoIR Recommendations for Ethical Decision-Making and Internet Research and the IEEE Ethically Aligned Design Guidelines) or that national or supra-national bodies create (such as the EU's 2019 Ethics Guidelines for Trustworthy AI [7]).

It appears likely that any global, Web-scale agreement on rights and on the need for weighing different rights against each other in governance processes will only be possible at the level of the UN. The ongoing quest for a declaration or even bill of digital human rights [21] may realise the ROAM Principles, endorsed in 2015 by UNESCO's General Conference, which state that the internet should be '(i) Human Rights based (ii) Open, (iii) Accessible to all, and (iv) Nurtured by Multistakeholder participation'.

[8] France's law concerning the fight against the manipulation of information (loi relative à la lutte contre la manipulation de l'information) is an example of this

Power relations and perceptions change over time, and this impacts possibilities and impact of governance. For example, the GDPR is now, after much initial industry lobbying and claims to the effect that it would hamper innovation, hailed by many as an example for legislation protecting personal data and privacy. Conversely, after more than a year of NetzDG application in Germany, there appears to be no evidence of the over-blocking by social-network platforms that had been feared by many civil-society advocates.As the Snowden revelations have illustrated, laws on national security passed by dominant nations govern how individuals around the world can (or cannot) effectively exercise their fundamental rights and freedoms, including those afforded to them by their own countries' laws. Tech giants wield enormous power over citizens and conceivably even over elections and thus the fabric of democracy. It has also been argued that the blocking of content that violates economically powerful companies' copyright enters laws and is enforced faster and more thoroughly than the blocking of content that violates vulnerable individuals' and groups' personality rights, even though the technical processes to detect and block such content may be similar. Within companies, ethics-related desiderata formulated by management become binding rules for governance more easily than those formulated by workers and employees.

4.2 The role of Web Science in governing the Web

Challenge – Action: to further values by concrete steps in the design and operation of Web systems

The complex governance processes open up possibilities for intervention for individual researchers, platforms, or the Web as a whole. Methods for intervention can be classified by the phase of the life-cycle they focus on. They include methodologies for embedding ethical concerns and values in the design and revision process, descriptive/analytic studies, affirmative/corrective action, and the raising of awareness and nudging.

One important role of Web scientists is to study, describe and analyse the socio-technical system of the Web, for example to contribute a scientific view to supplant media and policy discourse that is often carried out in ideological and emotional terms. The gathering and analysis of a strong evidence base is an important task for Web Science. Principled evaluations of the processes and successes of legal as well as self-regulation efforts are another important area for Web-Science research. A further key contribution are critical analyses of the ideals invoked by laws and other professional, national and international agreements, such as the investigation of multi-stakeholderism in Internet Governance by [12]. Especially given the pervasive and dynamic power structures across the Web, *critical* research on governance and its evolution is another key task for Web-Science research and policy advice.

In hybrid human-agent environments, often unexpected situations arise that may require affirmative or corrective actions. Risk analysis during the design phase and before deployment may help preview some possible risks and to be pro-active as well as reactive. Existing laws may provide effective means against some risks, and unexpected means against other risks (cf. the proposal to use copyright law to combat revenge porn [**?**]). Steffen: Taybot is no longer mentioned in section 2 In emergencies (with Taybot being a rather harmless example), it should be possible to have a kill-switch under human control, and it is necessary to act fast. Operators of Web content and services should be responsive to advice and, where possible, draw also on research findings. A positive example is the reaction of the social network platform nextdoor.com to the occurrence of racial profiling of neighbours by neighbours: Users who referred to "race" when posting to "Crime Safety" forums were prompted to provide additional

information, such as hair, clothing, and shoes, which helped reduce the occurrence of racial profiling. (https://hbr.org/2018/05/how-nextdoor-addressed-racial-profiling-on-its-platform)

In different roles, we are also responsible for informing users, the public, and other stakeholders. These activities should include, as a minimum, to raise awareness about technical possibilities and facts (such as the presence and activities such as bots [7]) and about what is regulated by laws (e.g., due diligence to clarify liability questions, bot content production in Wikipedia as a case of public speech). Information activities can aim at raising awareness, and they can also aim at changing behaviours in certain directions, from nudging to forcing. Nudging towards behaviours may operate via information about psycho-social effects and nudging towards attitudes such as empathy. Well-known examples from the Web include differences between online and offline behaviour (in terms of social checks and balances that are missing online), and the reductions in empathy often associated with cultural or spatial distance.

However, most informing and certainly all nudging or forcing rests on values and choices, and neither can be assumed to be "good" in and of itself [10]. Self-reflection about these values and choices, and about whether and how to integrate which other parties' values and choices, is as important as the attempt to be "ethical" itself. This has impacts on the definition and implementation of the concept of "benevolence" that was introduced above. At the same time, a belief in democracy necessitates that we limit the integration of some values and choices: There is ample historic evidence (as well as current evidence, see the example in Section **??**) that democracy may give its enemies the means to destroy it. We recommend that the future governance of the Web draws on lessons learned on how to build a "militant democracy" to better defend itself [2].

4.2.1 Designing networked systems in an ethical way: What does "benevolent" mean?

Steffen: I do not quite understand the structure of section 4.2. Simple indicator: why have a 4.2.1 if there is no 4.2.2? Further: my feeling is that as the text progresses, it slowly, slowly leaves Web/internet governance and moves on to other - very relevant - issues. It is not clear to me where to draw the boundary. I wonder if e.g. 4.2.1 should not rather be merged into section 3?

Challenge – Benevolent (revisited): to find out what "harm" and "benevolence" means, and to design in the face of conflicts between different notions and interests

A key challenge for ethical thinking and acting is that every technology is "good" for someone and that the temptation to think that one is "good" and knows what "is good for" others is large. Striving for the aims of being benevolent and avoiding harm is important and necessary to create not only the AI, but also the Web we want. However, this goal is associated with many pitfalls, and even a simple analysis shows how difficult it is to even define what it would mean for AI (or Web, or other) computational technology to be "for the Common Good". Even basic questions relating to substantive, procedural and distributional facets of the Common Good are routinely not asked in today's research; instead, existing power relations tend to be reinforced [2].

There is a rich literature with long traditions on the philosophical and practical problems of how to design "in an ethical way" (i.e., in order to bring about "good"), and this includes studies on the design of computational and information systems. Families of techniques such as value-sensitive design focus on how to bring values as requirements to bear, participatory design and democratic methods for gathering and negotiating requirements can counteract

overly narrow notions of values. On the algorithmic and HCI level, design can and should draw on the rich body of literature, methods and tools for protecting against specific challenges such as privacy violations and discrimination.

Bettina: How much depth and breadth do we want in the references here? I could add some overviews of these vast fields.

While there is wide agreement that stakeholder involvement is key, a purely utilitarian foundation of design decisions may not always be sufficient [2]. This may not only ignore or contravene deontological values, but also non-human stakeholders such as animals or the environment.

Reflective methods such as [8, 9]) structure moral reasoning and usually proceed through stages such as identifying the functional requirements of the existing or to-be-built systems, identifying the stakeholders of the system, identifying the values of different stakeholders and examining conflicting interests, and making a decision on each functional requirement reflecting on the above. Typical stakeholders in Web-related systems will include users and non-users, advertises and marketeers, and also the Web Science community, and typical conflicts of interest will be between profit-seeking, growth, and interaction amplification on the one hand, and the limitation of socially undesired interactions such as hate speech.

In spite of these fundamental (and partially unresolvable) issues, we agree with such design techniques and methods in considering a constructive approach that generates awareness as the best that can be done. A sub-goal is to generate awareness on societal and ethical concerns that are not pre-fixed but emanate from any stage of the project activities. It is from a situation of such awareness that “better” design decisions can be made, that “bad” choices can be perceived and hopefully rectified more easily.

5 Extending Web Science

Understanding the web and its current evolution through AI demands a parallel evolution in the methods, ethics and epistemology for Web Science. As the object of our study evolves and the necessity for collaboration across disciplines and with researchers in industry and government becomes ever-greater, it is urgent that we innovate our methods and research infrastructures accordingly.

Key challenges for Web Science at present are as follows:

- How can we unlock the promise of interdisciplinarity inherent in the vision of Web Science?
- How can we harness AI for web science research?
- How to share data, algorithms and tools between researchers and with industry?
- How to create inclusive and robust public debate about the Web and its futures?

In the following we explore how these issues might be addressed.

5.1 Visualization for interdisciplinarity

Academic disciplines work with different objectives and have developed different, often very different epistemologies. Especially, there are very different ideas withing the social sciences and computer sciences as to what constitutes a method or how knowledge should be derived. Often there is a misalignment of interests, for example, if social scientists would like computer scientists to become "epistemic partners“ (Knorr-Cetina 1999) in the collaborative

production of theories and knowledge while computer scientists would like social scientists to inform the design and building of content and infrastructure and thus expect a particular type of results from the research (Kinder-Kurlanda 2014). Web Science has been calling for interdisciplinarity since its beginning and has been aiming to provide platforms for interdisciplinary work for a long time. There is clearly a large potential in integrating a comprehensive understanding of the data with sophisticated computational methods (e.g. Halford and Savage 2017). However, transcending established knowledge frameworks to build new understandings is difficult and demands not only to take risks and to deal with uncertainty but also requires plenty of patience and indulgence.

One area where we see a potential for researchers from various disciplines collaborating more closely to better understand the web is visualization. The visualisation community in computer science possesses a wealth of techniques and tools to interactively explore data and find patterns. If these visualisations became approachable and understandable across different web science communities this would allow to trace the impact of computational methods on the results. As 'boundary objects' (Bowker and Starr 2000) visualizations would be focus points for a meeting of different types of experts.

Visualization can enable an interdisciplinary 'interviewing' of data – of enabling a dialogue between data, method and theory – using iterative visualisations to support abductive reasoning. The complexity of data makes formulating hypotheses very difficult, and one way of allowing such formulating could be through a dialogue where we query the data and refine ideas in an iterative process. This could allow computer scientists and social scientists to work together, to bring theory and data into dialogue much earlier in the process, by looking at the same data, asking questions, visualizing the data in different ways. An example is Visual Analytics / Visual Data Mining where questions are derived from the data, to derive new hypotheses.

Questions could then become how we might be able to map interviewing methods used in the social sciences (e.g. expert interviews, semi-structured interviews employing an interview guide, biographical interviews) to how we might interview data; can we map social science methods of interviewing to different computation techniques? Can visualization allow to annotate quantitative data with qualitative information?

Collaborating in crafting and interrogating visualizations would allow to innovate in interdisciplinarity, to find new ways to collaborate on an eye-to-eye level, especially between computer scientists and social scientists.

Visualization can also be a way to represent outputs/findings to the academy, to government, industry and the public. Visualisation is about communicating results – it is required to make results understandable. However, it needs to be a concern from the beginning, not just at the end of a research project.

5.2 Collaborative research and shared infrastructures

How can Web Scientists do valid and ethical social science research with Web data towards understanding and impact – drawing together strong social research questions with engineering approaches. Three benefits (i) new sources of data for web science research (ii) critical analysis of ethics, methods, effects (iii) knowledge and understanding of industry practice (by inference at least).

- new models for access to data and tool sharing

- algorithm sharing across the academy and through dialogue with industry partners to develop innovative practices and mechanisms for collaboration
- building web science data archives, data management tools (for documenting research processes throughout a project's lifetime) and infrastructures
- actively seeking collaboration across domains and keeping the door open for including new disciplines in web science

The Web Science Observatories [32] of tomorrow must benefit from the latest AI techniques to support mixed research methods at a Web scale. AI techniques that could help the Web Scientist perform a number of tasks including : collect and compile datasets; ensure the quality of data; detect bias; support the data analysis; enforce privacy, anonymity, etc. in the data; apply scientific best practices; support reproducibility; etc.

5.3 Harnessing AI for Web Science Research

Steffen: Who turns this into running text? Susan: I have added the text from Oshani and David to this section too, it covers much of the same ground

Challenge – tooling: Harness AI tools for Web Science

Over the coming years, Web Science will focus closely on the emergence of Artificial Intelligence and its implications for the Web. However, it is also the case that AI tools can and should be harnessed for Web Science, enabling us to observe, analyse and intervene in the evolution of the Web.

- key role of data and their management in the relation between AI and Web Science ; The open and linked data facet of the Web is a special case of particular importance when considering the links between AI and Web (science) data (science)
- the use of AI to produce the datasets for Web Science studies and inversely the study and design in Web Science of high quality datasets to be used by AI for training, reasoning, etc.
- data science to help solve problems of datasets of AI (bias, etc.) : automate scrutiny on the Web and generate reports for Web scientists ; Web exploration powered by AI;
- Intelligent agent can help us produce, curate, share and maintain corpora and datasets. For instance AI techniques could be designed to check the quality of a dataset and look for bias in it.
- AI simulations helping us avoid privacy implications and issues;
- AI can be also used for simulations producing data and to generate synthetic data; item Inversely, Web Science could produce multidisciplinary methods and tools to certify the quality and characterize training sets to improve the quality of the learning and conclusion made by AIs using them
- Reproducible research and Dataset sharing should be a core feature in Web Science research and the need to use provenance to understand how content is generated
- Data sharing architectures/ data trust archives/reserves, met searching, search across sources; Web of archives and search across that Web
- Reproducibility and secondary use of the datasets
- specific relation between Machine Learning and Semantic Web to support Web science

we could say that in Web Science, we should build our research program as a joint effort with two other research fields born in the 50s: “AI” for Artificial Intelligence [24] and “IA”for Intelligence Amplification [1] and Intelligence Augmentation [14]. While AI focuses on methods to simulate and scale automated intelligent behaviours, IA focuses on methods to amplify intelligence with systems where the natural intelligence is at the center from the start. As a universal platform linking all kinds of intelligence, the Web has the potential of being the meeting point for AI and IA. But many open questions will have to be addressed in Web science to reconcile these two historical trends.

Susan: David and Oshani’s text below

item the challenges of big data (volume, variety, velocity and veracity) meet the complexity of AI algorithms; some aspects are a blessing (e.g. volume and deep learning) some are harder (eg. velocity, veracity, etc.)

The Web already is populated by Web bots but they usually are restricted to certain realms while they could be generalized. For instance we could generalize the bots as the ones of Wikipedia to bots on the open Web designed to monitor and preserve certain characteristics of the Web.

AI to detect and counter-attack some undesirable network effects: staring by defining the Web we want, how can AI Help? e.g. consider the "Giant attractors" of the Web either with global view (crawling) to maintain metrics and then intervene or agents with simple rules pushing emergent behaviours such as an agent purposefully posting links to other platforms to foster linking, decentralizing.

from the infrastructure point of view, we could imagine Web farms for Web AIs hosting autonomous agent that would study, monitor and report on the Web.

benevolent AI for the Web (e.g. watchdogs for the Web) and also helping humans face humanity in terms of diversity and scale but also in preventing unwanted behaviours (e.g. bullying) ; Problems that could be targeted by these Web bots include: the detection of metrics manipulation, cross-language plagiarisms, centralization or digital divide; the prevention of vandalism or spamming; the generation of links, back links, navigational content beyond search results; etc. These agents would be based on policies and values important to the philosophy of the Web (e.g. seek decentralization, equality of access) to improve its resilience and quality. Web agents working to improve users’ experience, understanding, awareness and control of their participation and contributions to the Web. For instance, educational AI could help educate Web users in many domains including Web literacy or ethical thinking. Agents could also provide customized descriptions of the context in which a user is, including security, neutrality and privacy notices or his human-computing participation when it occurs. AI could also help users burst our filter bubbles and foster serendipity. On the longer term, benevolent AIs could actively help enforce (human) rights on the Web and be scrutiny agents for important values of the Web. Workflows and data are key drivers of Web Science; The multidisciplinary nature of Web Science also puts it in an ideal position to explore and expand the forms of intelligence on the Web. First, both Web Science and AI are highly multidisciplinary [17] and the multiple disciplines that are common to both fields are as many bridges to make them interact. AI could also be used to operationalize the expertise from each domain into agents that help us providing assistance, reporting or training from the domains they represent. Thes agents could help us find and support a massively multidisciplinary method and allow us to scale to the multi-disciplinary interactions required by the design and study of the Web. One possibility, for instance, would be for these AIs to produce and maintain boundary artifacts at the frontiers of disciplines. Web science AIs should be citable in Web Science papers; Have AIs involved in the management

of our Web Science community (propose programs, topics, etc.) AI to perform good science research and the required characteristics for that such as explainable AI and transparent algorithms Following the wiki-way, AIs could be created edited, crossed, and bred on the Web, socially maintained, copied and versioned: the Web way applied to AI with, for instance, "copy-paste-customize" based contribution to the population of agents. For this to happen, and just as it was the case for the Web, we would need a public domain Web-based AI architecture.

5.4 Participatory methods: a Web 'for everyone'

We opened this manifesto with the observation that the Web is now the nervous system for our planet. Four billion people use the Web regularly, and this number will rise rapidly over the coming years. Even those who do not use the Web are deeply shaped by activities that take place online. What started as a tool for Physicists is now 'for everyone': all our futures are entangled with the future of the Web. This means that 'web science' is not just for researchers. It is for entrepreneurs with an eye for a new business proposition, corporations seeking to innovate. It is for governments wanting to make the most of the Web for services and engagement, whilst assuring their citizens that they have robust regulation and governance in place to manage the Web. And Web Science is for individuals and communities, whose everyday activities make up so much of the content online, whose data are driving the data economy, and whose lives are increasingly unimaginable without the Web.

In thinking about the future of the Web, it is important that these voices are taken into account. Web Science must engage the full range of 'stakeholders' above, to ensure that the full range of voices is heard as we build our understanding of the Web and seek to shape its future. As Web Scientists, we are committed to a developing participatory methods that allow us to gain insight of diverse perspectives and build dialogues between these.

Questions of ethics and responsible innovation are critical here. These have become pressing questions for Web Science. How do we ensure that web technologies are ethical? that they taken into account potential harms? This is not simply a question of 'ethical approval' for Web Science research but a wider 'ethics of care' towards the kind of society we are making, as we work to shape the Web towards the future. This means working with citizens as well as governments, community groups as well as businesses, to ensure that the future of the Web takes into account the diverse views, feelings and needs of all its' stakeholders.

There are already some good models for collaboration with government and business - co-funding of research, knowledge exchange partnerships, even joint-research - and it is important that we extend these in the coming years. However We also need to extend collaboration with citizens and communities to enable meaningful participation and effective voice in shaping the future of the Web. These methods may include citizen science, whereby individuals participate in distributed models of data collection and analysis Susan: ref Simperl. Similarly citizen social science can be used to encourage people to participate in research on areas of particular concern to their lives and communities Susan: ref Ryan. We should also consider how new and emerging forms of data can be used for community purposes, not just for recommender systems and targeted advertisements. For example, Web Science research has already developed environmental pollution apps to inform asthmatics which parts of cities to avoid Susan: ref, and built platforms for open data sharing and community-relevant analytics Susan: ref and participatory budget making Susan: add ref.

Beyond these excellent examples of 'tech for good' we should also think about empowering citizens to be active in debates about the future of the Web, through education in schools, and through concerted efforts to build 'anticipatory capacity' Susan: Appadurai 2012; Levitas 2014; Amsler and Facer 2018. Specific mechanisms such as online citizens Susan: Smith ref assemblies that actively seek to build widespread participation Susan: add something on what AI can do here

5.5 Conclusion: Inclusive Web Science

As Web Science evolves with the Web, so too it becomes apparent, and indeed pressing, that our methods evolve. At the beginning of Web Science this meant integrating robust social science methodologies with the expert methods of Computer Science. This has proven both more difficult and more rewarding than originally imagined, and still we continue to learn about how to do it better. As we look towards Web Futures, we must innovate again. This will include harnessing AI for Web Science research. And it will also mean adjusting Web Science research to the emergent AI Web. As the Web becomes ever more heterogeneous, to include animals, plants, ships, rocks (and so it goes on) we need to think about how we research these new networks, their activity end effects, the opportunities and the challenges that they present, not only for Web Science but for the future of the Web. And we must do so in a context where the continued growth of the Web is a pressing question for the sustainability of the planet. Environmental Web Science is as yet in its infancy and we will need to draw in new disciplines, from the life sciences in particular, and new approaches that help us to research, understand and intervene in this truly inclusive Web.

Steffen: The following was put into the capacity section, but clearly it does not belong there

5.5.1 Knowledge Infrastructure and governance of Web: *extending the Web observatory vision*

— *Editors: Oshani and David*

6 Capacity Building

— *Editors: Steffen*

6.1 Supporting Ethical and Value-based Thinking

Teachers: teach interdisciplinarity
Researchers: Data science ethics
Universities: New subjects. E.g. there are CDOs, but aside of data scientists, where are the data officers that do not only do analytics but have a picture like https://theodi.org/article/data-ethics-canvas/ in mind
Industry: new roles
Policy makers:
Funding agencies:

6.2 Supporting Interdisciplinarity

understanding the socio-technical system

Teachers: teach interdisciplinarity
Researchers: Talk to the domain experts
Universities:
Policy makers:
Funding agencies:

6.2.1 Funding

funding bodies do not fund interdisciplinary work
scientists need to come up with better tools for evaluation producing checklists
two-stage approaches * funding for preparing a full proposal

6.3 Resource Sharing

Teachers: teach interdisciplinarity
Researchers:
Universities:
Policy makers:
Funding agencies:

6.4 Global Collaboration

Teachers: teach interdisciplinarity
Researchers:
Universities:
Policy makers:
Funding agencies:

6.4.1 International Centers for the Web

Fabian
W3C
US pushing, China pushing; Europe not
if the web is not international, then it is not

7 Participants

- Steffen Staab
Institute WeST, Universität Koblenz-Landau – Koblenz, Germany, & WAIS, University of Southampton – Southampton, UK
- Fabien Gandon
Inria, Université Côte d'Azur, I3S, CNRS – 2004 rt des Lucioles, BP93, 06902 Sophia Antipolis, France
- John Doe
Somewhere University – City2, DC
- John Q. Open
Dummy University Computing Laboratory, DC

Acknowledgements

We thank John Doe and Jane Doe for their valuable contributions.

References

1 Ross Ashby. Design for an intelligence-amplifier. *Automata studies*, 400:215–233, 1956.

2 Bettina Berendt. AI for the common good?! pitfalls, challenges, and ethics pen-testing. *Paladyn*, 10(1):44–65, 2019. URL: https://doi.org/10.1515/pjbr-2019-0004, doi:10.1515/pjbr-2019-0004.

3 Tim Berners-Lee, Robert Cailliau, Ari Luotonen, Henrik Frystyk Nielsen, and Arthur Secret. The world-wide web. *Commun. ACM*, 37(8):76–82, August 1994. URL: http://doi.acm.org/10.1145/179606.179671, doi:10.1145/179606.179671.

4 Tim Berners-Lee, Wendy Hall, James Hendler, Nigel Shadbolt, and Daniel J Weitzner. Creating a science of the web. *Science*, 313(5788):769–771, 2006.

5 M. Bevir. *Governance: A very short introduction*. Oxford University Press, Oxford, UK, 2012.

6 Andrei Ciortea, Simon Mayer, Fabien Gandon, Olivier Boissier, Alessandro Ricci, and Antoine Zimmermann. A Decade in Hindsight: The Missing Bridge Between Multi-Agent Systems and the World Wide Web. In *AAMAS 2019 - 18th International Conference on Autonomous Agents and Multiagent Systems*, page 5p, Montréal, Canada, May 2019. URL: https://hal-emse.ccsd.cnrs.fr/emse-02070625.

7 European Commission. Ethics guidelines for trustworthy ai, 2018. https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai.

8 Loy D. Watley. 23(1), 2014.

9 Mathieu d'Aquin, Pinelopi Troullinou, Noel E. O'Connor, Aindrias Cullen, Gráinne Faller, and Louise Holden. Towards an "ethics by design" methodology for AI research projects. In Jason Furman, Gary E. Marchant, Huw Price, and Francesca Rossi, editors, *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, AIES 2018, New Orleans, LA, USA, February 02-03, 2018*, pages 54–59. ACM, 2018. URL: https://doi.org/10.1145/3278721.3278765, doi:10.1145/3278721.3278765.

10 R. De Wolf, E. Vanderhoven, B. Berendt, J. Pierson, and T Schellens. Self-reflection on privacy research in social networking sites. *Behaviour & Information Technology*, 36(5):459–469, 2017.

11 Ronald J. Deibert, John G. Palfrey, Rafal Rohozinski, and Jonathan Zittrain. *Access Denied: The Practice and Policy of Global Internet Filtering (Information Revolution and Global Politics)*.

12 L. DeNardis and M. Raymond. Thinking clearly about multistakeholder internet governance. In *GigaNet: Global Internet Governance Academic Network, Annual Symposium 2013*, 2013. Available at SSRN: https://ssrn.com/abstract=2354377 or http://dx.doi.org/10.2139/ssrn.2354377.

13 Deutscher Bundestag – Wissenschaftliche Dienste. Regulierung von hate speech und fake news in sozialen netzwerken durch eu-mitgliedstaaten [regulation of hate speech and fake news in social networks by eu member states], 2018. WD 10-3000 -032/18, https://www.bundestag.de/resource/blob/566942/a5eb997872bbe5dbca3f47112eb04c46/wd-10-032-18-pdf-data.pdf.

14 Douglas C Engelbart. Augmenting human intellect: A conceptual framework. SRI Summary Report AFOSR-3223, Stanford Research Institute, 1962.

15 Roy T Fielding and Richard N Taylor. Principled design of the modern web architecture. *ACM Transactions on Internet Technology (TOIT)*, 2(2):115–150, 2002.

16 B. Fox. Campaigners lock horns with eu justice chief on 'hate speech', 2018. EURACTIV, https://www.euractiv.com/section/freedom-of-thought/news/campaigners-lock-horns-with-eu-justice-chief-on-hate-speech/.

17 Fabien Gandon. The three 'W' of the World Wide Web call for the three 'M' of a Massively Multidisciplinary Methodology. In Valérie Monfort and Karl-Heinz Krempels, editors, *WEBIST 2014 - 10th International Conference*, volume 226 of *Web Information Systems and Technologies*, Barcelona, Spain, April 2014. Springer International Publishing. URL: https://hal.inria.fr/hal-01223236, doi:10.1007/978-3-319-27030-2.

18 Fabien Gandon. For everything: Tim Berners-Lee, winner of the 2016 Turing award for having invented... the Web. *1024 : Bulletin de la Société Informatique de France*, (11):21, September 2017. URL: https://hal.inria.fr/hal-01843967.

19 Fabien Gandon, Michel Buffa, Elena Cabrio, Olivier Corby, Catherine Faron Zucker, Alain Giboin, Nhan Le Thanh, Isabelle Mirbel, Peter Sander, Andrea G. B. Tettamanzi, and Serena Villata. Challenges in Bridging Social Semantics and Formal Semantics on the Web. In Hammoudi, S., Cordeiro, J., Maciaszek, L.A., Filipe, and J., editors, *5h International Conference, ICEIS 2013*, volume 190 of *Lecture Notes in Business Information Processing*, pages 3–15, Angers, France, July 2013. Springer. URL: https://hal.inria.fr/hal-01059273.

20 Fabien Gandon and Alain Giboin. Paving the WAI: Defining Web-Augmented Interactions. In *Web Science 2017 (WebSci17)*, pages 381 – 382, Troy, NY, United States, June 2017. URL: https://hal.inria.fr/hal-01560180, doi:10.1038/nbt0609-508.

21 L. Gill, D. Redeker, and U. Gasser. Towards digital constitutionalism? mapping attempts to craft an internet bill of rights. Technical report, Berkman Center Research Publication No. 2015-15, 2015. Available at SSRN: https://ssrn.com/abstract=2687120 or http://dx.doi.org/10.2139/ssrn.2687120.

22 Christophe Guéret, Victor de Boer, and Stefan Schlobach. Let's" downscale" linked data. *IEEE Internet Computing*, 18(2):70–73, 2014.

23 Annette Markham. Fabrication as ethical practice: Qualitative inquiry in ambiguous internet contexts. *Information, Communication & Society*, 15(3):334–353, 2012.

24 John McCarthy, Marvin Minsky, Nathaniel Rochester, and Claude Shannon. A proposal for the dartmouth summer research project on artificial intelligence, august 1955. URL: http://raysolomonoff.com/dartmouth/boxa/dart564props.pdf.

25 Panagiotis T. Metaxas and Eni Mustafaraj. Panagiotis t. metaxas and eni mustafaraj. In *Proc. of the WebSci'10 Conference*.

26 Eni Mustafaraj and Panagiotis Takis Metaxas. The fake news spreading plague: Was it preventable? In *Proceedings of the 2017 ACM on Web Science Conference, WebSci 2017, Troy, NY, USA, June 25 - 28, 2017*, pages 235–239, 2017. URL: https://doi.org/10.1145/3091478.3091523, doi:10.1145/3091478.3091523.

27 José San Pedro, Davide Proserpio, and Nuria Oliver. Mobiscore: Towards universal credit scoring from mobile phone data. In *User Modeling, Adaptation and Personalization - 23rd International Conference, UMAP 2015, Dublin, Ireland, June 29 - July 3, 2015.*

Proceedings, pages 195–207, 2015. URL: https://doi.org/10.1007/978-3-319-20267-9_16, doi:10.1007/978-3-319-20267-9_16.

28 S. Saad, S. Bazan, and C. Varin. Asymmetric cyberwarfare between israel and hezbollah: The web as a new strategic battlefield. In *Poster Proceedings WebSci'11, June 14-17 2011, Koblenz, Germany*, pages 1–4, 2011.

29 Sabrine Saad, Muriel Chamoun, and Stéphane B. Bazan. Infowar on the web: When the caliphate goes online. In *Proceedings of the ACM Web Science Conference, WebSci 2015, Oxford, United Kingdom, June 28 - July 1, 2015*, pages 53:1–53:3, 2015. URL: https://doi.org/10.1145/2786451.2786487, doi:10.1145/2786451.2786487.

30 Michel Serres. *The natural contract*. University of Michigan Press, 1995.

31 Emma Strubell, Ananya Ganesh, and Andrew McCallum. Energy and Policy Considerations for Deep Learning in NLP. *arXiv e-prints*, page arXiv:1906.02243, Jun 2019. arXiv:1906.02243.

32 Thanassis Tiropanis, Wendy Hall, Nigel Shadbolt, David De Roure, Noshir Contractor, and Jim Hendler. The web science observatory. *IEEE Intelligent Systems*, 28(2):100–104, 2013.