Report from Dagstuhl Perspectives Workshop 18262

10 Years of Web Science: Closing The Loop

Edited by

Susan Halford[1], James A. Hendler[2], Eirini Ntoutsi[3], and Steffen Staab[4]

1 **University of Southampton, GB,** `susan.halford@soton.ac.uk`
2 **Rensselaer Polytechnic Institute, Troy, US,** `hendler@cs.rpi.edu`
3 **Leibniz Universität Hannover, DE,** `ntoutsi@kbs.uni-hannover.de`
4 **Universität Koblenz-Landau, DE,** `staab@uni-koblenz.de`

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 18262 "10 years Web Science: Closing the Loop" that took place in Schloss Dagstuhl from 25-29 June 2018. In total, an interdisciplinary team of 22 researchers from computer science, sociology, philosophy and law gathered and discussed on the past, present and future of Web Science and what sort of actions the community should take to stay faithful to its initial mission for societal good. The role of Web Science is more critical than ever given the ever growing impact of the Web in our society.

Seminar June 24–29, 2018 – http://www.dagstuhl.de/18262
2012 ACM Subject Classification Information systems → Collaborative and social computing systems and tools, Information systems → World Wide Web
Keywords and phrases Dagstuhl Report, Web Science, Dagstuhl Perspectives Workshop
Digital Object Identifier 10.4230/DagRep.8.6.173
Edited in cooperation with Fabien Gandon, Bettina Berendt, Katharina Kinder-Kurlanda, Pinelopi Troullinou

1 Executive Summary

Eirini Ntoutsi

This Dagstuhl Seminar aimed at bringing together researchers from different disciplines related to Web Science, namely computer science, sociology, philosophy and law to discuss on future of Web Science and how it can stay faithful to its initial mission for societal good. Several recent incidents like the online psychological experiment by Facebook have provoked widespread public concern regarding the effect of such experiments and interventions and there is no agreement on expertise and ethics knowledge about how to do Web experimental research.

The Web is a complex sociotechnical system where humans and (intelligent) machines interact in unexpected ways; such hybrid societies of natural and artificial intelligence raise new challenges for Web Science which go beyond technical challenges into ethical, legal and societal implications. The role of Artificial Intelligence in these developments was discussed extensively in terms of both opportunities and risks.

Based on the discussions and inputs from all participants, we have split the discussion into three main working groups:

- Working group on innovative methods for Web Science
- Working group on values
- Working group on Web Science and Artificial Intelligence

The group will continue its work in the aforementioned topics and a manifesto is foreseen to be ready by the end of the year.

2 Table of Contents

3 Overview of Talks

3.1 10 years of Web Science

Susan Halford (University of Southampton, UK, Susan.Halford@soton.ac.uk)

In opening this Dagstuhl seminar this talk focusses on the origins, ambition and achievements of Web Science and – building on theses – the pressing challenges that Web Science must face in the coming decade. Tracing our origins in the early 2000s to both a paradox – despite the phenomenal growth and impact of the Web it was rarely a topic of research in its own right for Computer Science – and an epiphany – that whilst the Web had begun as a set of technical protocols and standards for global information sharing it was rapidly becoming much more, as individuals, communities, businesses and governments took up the opportunities it presented, turning these into unanticipated outcomes that were raising profound economic, social and political questions. The Web was changing the world and the world was changing the Web. Working with, understanding and shaping the future of the Web demanded new ways of working across traditional academic silos, a new form of interdisciplinary expertise. Our success over the past decade has been remarkable: collaborative research grants, major doctoral training programmes, an annual Web Science conference, the Web Science Summer School, the Web Science Trust and a growing network of partner research groups in the Web Science Trust network. The interdisciplinary skills, concepts, and forms of collaboration that we have built over the past decade are essential if we are to face the challenges ahead. Just as in the early 2000s we face a rapidly morphing sociotechnical system, as the Web has led to Big Data, Data Science and the resurgence of Artificial Intelligence, raising profound questions not only about privacy, security and ethics – critical as these are – but also about the future of work, economic and social inclusion, risk, sustainability and global governance, about the kind of society we want for the future. Put this way, there can be no doubt that Web Science is needed now, more than ever!

3.2 Why data science needs web science: reflections from recent research

Elena Simperl (University of Southampton, GB, e.simperl@soton.ac.uk)

Data science emerged as a term in its own right less than five years ago. Since then it has proven tremendously popular across sectors and organisations around the world looking for new ways to innovate their products, services and operations. The demand for data scientists has never been greater – to be able to source and make sense of data using advanced data science machinery, organisations need new capabilities, drawing upon statistical and computational methods.

The data science community developed a wide understanding of the challenges and limitations their methodological tools. Data science approaches are often much more than a mix of mathematical models – they rely heavily on having access to relevant data, on the quality of this data, and on means that help people use understand their implications in

practical situations. Web science can provide a context and emerging solutions to some of these questions. The Web is one of the largest sources of data ever created, including large repositories of labelled data, for example in the form of user-generated content, system logs, or social media posts, that can be used to train data science models.

Web scientists have worked with these training corpora for many years and have developed an unparalleled understanding of the underlying systems and communities. They were among the first to debate critical issues around the quality of large data sets and the implications of the associated data collection and processing methodologies on the validity and usefulness of analytics. These experiences could and should inform the design of data science algorithms, guide the interpretation of their outcomes, and enrich ongoing discussions around responsible data science and FAIR.

By the methods it uses, data science draws upon insights from different disciplines. Web science is one of the best examples of recent date that show how interdisciplinarity could work – the lessons learned since the seminal article in Communications of the ACM that founded the field in terms of research methods, education and impact should be transferred into data science to develop a similarly rich understanding of the opportunities and challenges of true cross-disciplinary work.

We live in a time when data science is plagued by concerns about responsible data sharing and use. These are key challenges that can be addressed only through a concentrated effort that is mindful and serious about interdisciplinarity and defines a new code of values in the digital world. Data scientists should collaborate with Web scientists to learn more about the biases and limitations of Web datasets and platforms and about the broader legal and socio-economic implications of data sharing and algorithmic decision making. I believe this is only way to ensure that the ever-growing trend of datification affecting every single aspect of our lives will have a positive impact over time.

3.3 Legal Aspects

Nikolaus Forgó (University of Vienna, AT, nikolaus.forgo@univie.ac.at)

Main reference REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)

URL https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=DE

The presentation gave an overview on governance issues using legal means in the development of the web. Specific attention was given to data protection and data security law in Europe as it is enshrined in the General Data Protection Regulation, a law that became directly applicable in all EU-member states in May 2018. The history of data protection, leading back to the eighties of the 20th century shows that data protection as always about conflicting fundamental values that are articulated in fundamental rights. The main argument made was that rules that are "clear" enough to regulate technical developments on the web require (some kind of) basic understanding about how to outbalance conflicting fundamental rights. As long as these discussions are not led and not decided, data protection law risks to be too late too abstract and too unclear to produce the certainty Europe needs. The examples taken to illustrate this argument were mainly taken from European research projects in the domain of ICT for health.

3.4 Social experiments in the Web: The case of Bibsonomy – technical, social and legal implications

Andreas Hotho (University of Wuerzburg, DE, hotho@informatik.uni-wuerzburg.de)

It's more than ten years now, that the transition of the Web from a static to a user-driven and dynamic Web has started. In 2006, we have initiated a social experiment in the form of BibSonomy, a research platform to publicly share, manage and exchange bookmarks and publication. Users annotate or describe these items by tags which enable their efficient retrieval and support the management of large document collections. The collected data results in a lightweight semantic structure contributed to by millions of users and forms a valuable source of knowledge which can be exploited in many different research and application scenarios. For this, Data Mining and Machine Learning methods play a central role and help to extract a multitude of valuable information from the user contributed data.

The talk gives a review of the past 10 years of developing, maintaining, but also analyzing the BibSonony system. Consequently, besides addressing typical software design issues as well as standard technical development questions, this talk focuses on a set of selected research questions and their results. A concise summary of these results can be found in [1], while the latest related research on behavior analysis and representation learning can be found in [2] and [3], respectively. Additionally, the talk touches on legal aspects and ends with lessons learned. Thus, overall, this talk summarizes and illustrates the story of bringing Web Science to life.

References

1 Dominik Benz, Andreas Hotho, Robert Jäschke, Beate Krause, Folke Mitzlaff, Christoph Schmitz, and Gerd Stumme. The social bookmark and publication management system bibsonomy. *The VLDB Journal*, 19(6):849–875, dec 2010.

2 Stephan Doerfel, Daniel Zoller, Philipp Singer, Thomas Niebler, Andreas Hotho, and Markus Strohmaier. What users actually do in a social tagging system: A study of user behavior in bibsonomy. *ACM Transactions on the Web*, 10(2):14:1–14:32, May 2016.

3 Thomas Niebler, Luzian Hahn, and Andreas Hotho. Learning word embeddings from tagging data: A methodological comparison. In *Proceedings of the LWDA*, 2017.

3.5 What can Web Science give to industry

Paolo Parigi (Airbnb and Stanford University – San Francisco, US, pparigi@stanford.edu)

Ever since Thomas Edison's Menlo Park laboratory or Bell Labs, modern companies have recognized the importance of research for the development of new products. Until recently, most of the research done in industrial research facilities focused on building new physical objects. Exceptions existed but there were indeed exceptions. Against this background, companies like Airbnb, Uber, Lyft, Facebook, Linkdin, Tinder are investing in research that focuses on learning about people's beliefs and behavior. This shift reflects the fact the products of these companies uses technology in order to facilitate interactions (in person or electronically). In turn, interactions create systems with emergent properties that require the expertise of Web scientists to study. I will illustrate this point using examples taken from my work and that of my collaborators in and outside of academia.

3.6 Web Science: What Next?

Ricardo Baeza-Yates (NTENT & Northeastern Univ., USA, rbaeza@acm.org)

I participated in a Web Science panel held at ECML/PKDD 2009 in Bled, Slovenia. This was my first encounter with this name, just one year after its inception. Then I had many questions: Do we need a new science? How much knowledge do you cover? Must a new science have something new? How hard it is to start? Science of an abstraction? (concepts, tools, and applications) Science of an object? Science in the name? The last two are the same problems that Computer Science has. Today, many of these questions are still valid and I do not know all the answers.

So, does Web Science covers all possible applications of the Web or we should just focus on the Web as communication channel plus the technology behind it? I believe the answer is the second one, which includes web search and data mining, web advertising, web user experience understanding, online social networks, etc. However, this does not include the applications to other fields such as e-health, e-science, e-learning, etc. Even computer networking might be out of the scope as Internet is the infrastructure that allows the Web to exist, but is not really part of the Web.

The Web conceptually could be considered as having three parts: content (including everything such as links), interaction (as result of the people using it), and incentives (the information market that drives the interaction with the content). This implies that Web Science should be important for data scientists, economists, sociologists, ethnographers, web designers, and many kind of software developers.

In the long list of web research problems, I am really concerned about the vicious cycle of bias in the Web [1], understanding users and the content generation process (most users do not contribute [3, 5]), the limits of small web data [2], and the use of machine learning to take fair decisions. The last two implies understanding the trade-offs between quality and data volume [4] as well as understanding how prediction errors are distributed in each problem space (usually this is not analyzed today). This naturally leads to design of ethical codes and accountable algorithms that can be explained and audited.

However, the future can be more complicated if we leave the solution of the societal problems to politicians and lawyers. So, we need to step up and not worry about discrimination and ethics when is already too late.

References

1 Ricardo Baeza-Yates. Bias on the Web. *Communications of ACM* 61(6), pp. 54–61, 2018.
2 Ricardo Baeza-Yates. BIG, small or Right Data: Which is the proper focus? https://www.kdnuggets.com/2018/10/big-small-right-data.html.
3 Ricardo Baeza-Yates, Diego Sáez-Trumper Wisdom of the Crowd or Wisdom of a Few?: An Analysis of Users' Content Generation. In *ACM Hypertext & Social Media*, pp. 69-74, Guzelyurt, TRNC, Cyprus, Sep 2015.
4 Ricardo Baeza-Yates, Zeinab Liaghat. Quality-efficiency trade-offs in machine learning for text processing. In *IEEE BigData 2017*, Boston, USA, Dec 2017.
5 Lorena Recalde, David Nettleton, Ricardo Baeza-Yates, Ludovico Boratto. Detection of Trending Topic Communities: Bridging Content Creators and Distributors. In *Hypertext 2017*, pp. 205–213, Prague, Check Republic, Jul 2017.

3.7 AI and Society

Wendy Hall (University of Southampton, GB, wh@ecs.soton.ac.uk)

Why is there so much hype around AI now? The concept of AI has been in the minds of science fiction writers for centuries. The idea of computers with intelligence started in the UK with Alan Turing's seminal paper Computing Machinery and Intelligence, published in 1950. The term AI was coined in the US in the 1950's and has been exciting researchers all over the world ever since but up until now it's progress has been one of exciting leaps forward followed by research funding blights, often called the AI winters. The current leap forward is driven by new developments in deep learning, high performance computing, the internet and the availability of vast amounts of data and on which algorithms can be trained. The current leap forward is also referred to as the Fourth Industrial Revolution and maybe as deeply profound for the future of society as previous industrial revolution. Only time will tell.

Without doubt in this early part of the 21st century, the leading countries in terms of the scale of AI research they support are the US and China, but because of the potential impact of AI on all our futures, other governments are actively considering strategies and policies for adopting and adapting to AI over the coming years. To build on the strengths the UK has in AI and to ensure the UK stays at the forefront of the AI revolution, Jerome Pesenti and I were asked by the UK government in March 2017 to undertake an AI review focussing on job creation and the growth of AI as an industry sector in the UK. This review was published in October 2017.

The review proved timely and was well received. Money was allocated in the UK government budget in November 2017 and the same month it was announced that AI would be one of the key sectors identified for support in the government's Industrial Strategy. There was funding also in the budget for the establishment of a new Centre for Data Ethics and Innovation in the UK, which was established in 2018. The detail of the £1 Billion AI sector deal was published at the end of April 2018 and at the same time the government established the new Government Office for AI.

The recommendations in the review were largely organised under four major themes – data, skills, leadership and adoption.The Office for AI has already started the consultation work for the development of data trusts, which were highlighted in the AI review as mechanisms to support the exchange of data between government, businesses – large and small, and university research labs, to support innovation in and around AI. During the review consultations, companies told us that their biggest problem was getting access to the data they needed to train algorithms, and this has to happen in safe and secure, ethically sound legal frameworks to enable innovation to flourish in a way that is good for society. The role of the new Centre for Data Ethics and Innovation will be crucial in taking this forward in collaboration with the OAI, it's implementation partners, and the AI Council.

Implementing the recommendation under skills has also started. UK Research and Innovation (UKRI) launched a call in February 2018 for up to 20 new AI Centres of Doctoral Training to start in October 2019, each of which will train at least 10 students per year for 5 years. The Alan Turing Institute – now the national institute for AI – will manage a commensurate AI fellowship programme, and there will also be programmes to establish industry funded AI and Machine Learning MSc's in UK universities.

However, when we were writing the review we were very conscious that to fill the AI skills gap and to ensure that AI is not only about technology but is also about how AI will

impact society and how society will utilise AI, we included a stream of work to facilitate the development of "conversion" courses that will take students from all disciplines (STEM and non-STEM) as well as students who are re-skilling and give them the expertise they need to work in the AI sector (not necessarily as ML programmers). Running alongside all this will be a campaign to increase diversity in the AI workforce. It is really important that the AI based products and services that are going to become increasingly important in every aspect of our lives in the future are not dominated by any one gender, age group, ethnicity, or culture and are accessible to all.

The UK is not the only government that has developed an AI strategy over the last year and it is interesting to see how the different strategies are panning out. However it will be many years before we can really evaluate what impact government intervention in AI has around the world given the potentially seismic political and global shifts that are bound to come over the next few years. How much and how fast AI is going to become a reality in our lives rather than just science fiction is hard to tell. There is a lot of hype at the moment, but one thing is for sure it will become a matter of global competition.

I would argue however, that it is just as important that we collaborate internationally as compete. AI has the potential to solve or help manage the biggest challenges that society faces in the 21st century but if we pool resources (data, research results, expertise etc.) we could achieve a lot more a lot faster and still enable our companies to compete internationally to sell the products and services that are produced as a result.

AI also has the potential to do a great deal of harm. We must keep at the front of our minds the sociotechnical impact of AI to ensure that we develop AI technologies that are first and foremost for the good of society. I believe the first ten years of Web Science were just the preparation for the real challenges that lie ahead in terms of building a world in which AI plays an ever more significant role in our increasingly interconnected lives.

3.8 Web Science, Artificial Intelligence and Intelligence Augmentation

Fabien Gandon (Inria, Université Côte d'Azur – Sophia Antipolis, FR, fabien.gandon@inria.fr)

Main reference Gandon, Fabien: "Web Science, Artificial Intelligence and Intelligence Augmentation", Journal of Seminar Documentation, 1:8, pp. 34–78.

This abstract paper summarizes some challenges and opportunities at the intersection of Web Science, Artificial Intelligence and Intelligence Augmentation.

Intelligent approaches to follow and support Web evolution

Initially, the Web was essentially perceived as a huge distributed library of linked pages, a worldwide documentary space for humans. In the mid-90s, with wikis and forums, the Web was re-opened in read-write mode and this paved the way to numerous new social media applications. The Web is now a space where three billion users interact with billions of pages and numerous software. In parallel, extensions of the Web were developed and deployed to make it more and more machine friendly supporting the publication and consumption by software agents of worldwide linked data published on a semantic Web. As a result of all its evolutions, the Web became a collaborative space for natural and artificial intelligence. This raises the problem of supporting these worldwide interactions and forming these hybrid communities. In my talk I presented some of the opportunities and challenges for Web Science in building this evolution of a Web toward a universal space linking all kinds of intelligence.

AI in classical tasks and problems of the Web

A first set of challenges can be directly identified from the classical tasks and problems we encounter on the Web e.g. help us search, browse, contribute, 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. We could imagine Web farms for Web AIs hosting autonomous agent that would study, monitor and report on the Web. 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.

The special relation of AI and data(sets) on the Web

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). Artificial Intelligence can be used to assist Web Scientists and vice-versa. 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. 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.

Benevolent AIs for a resilient Web

The two previous ideas could be generalized to the goal of designing benevolent AIs for the Web. 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.

AI to help us humans scale and face humanity on the Web

With the advent of the Web, human individuals also face humanity in all its scale and diversity. Web scientists could design AIs to help humans face humanity on the Web and help us scale to the world-wide web scale. 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.

A variety of AIs to absorb the varieties of the Web

The force of these AIs could also be in their multiplicity and interactions. The law of requisite variety of W.R. Ashby says that "variety absorbs variety" and in our case a diversity of AIs

could be a good way to address the many types of diversity we find of the Web (content, users, contexts, tasks, usages, resources, etc.). In fact more than AI, it is maybe distributed AI that has a rendezvous with the Web and its sciences [6]. 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 and its methods. 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.

Explore and expand all the forms of intelligence on the Web

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 [5] 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. These 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. The multidisciplinary domains of Web Science could also be leveraged to identify other ways of simulating, reproducing or engaging intelligence including emotional intelligence, communication skills, imagination, etc.

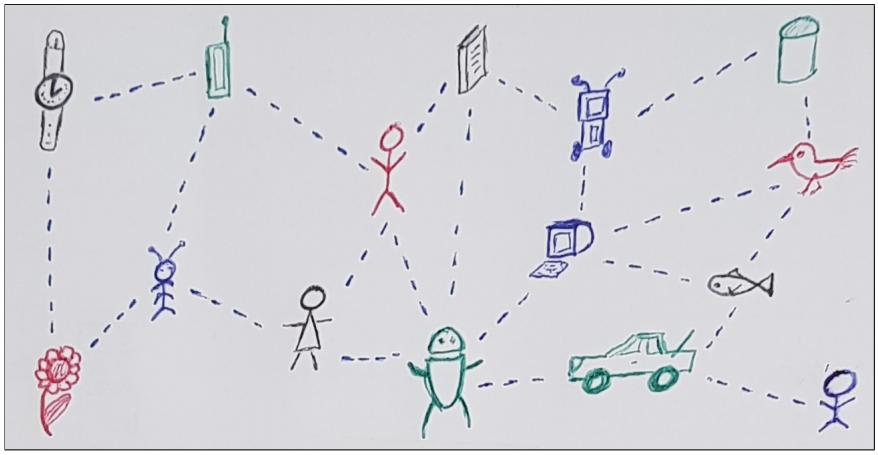
Studying and building the hybrid societies of the Web

Such an evolution as the one described in the previous sections would finally lead Web Science to consider the challenge of studying and designing hybrid societies of natural intelligence and artificial intelligence on the Web. This study would have to include different forms of natural intelligence (e.g. people, connected animals, connected plants) and different forms of artificial intelligence (reasoning, learning, inducing, etc.). The challenge will also be to study their interactions with the resources of the Web (linked pages, linked data, connected objects, etc.) forming the environment of these forms of intelligences. Web Science will have to face the problem of this massive interaction design with the Web and everything it links [7] 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 [4]. Studying and designing these hybrid societies, from swarms to complex societies with their normative rules, their social constructs, their governance, etc. will be a highly challenging and multidisciplinary task.

Towards a Web linking all forms of intelligence

In Web Science, we should build our research program as a joint effort between Web Science and two research fields born in the 50s: "AI" for Artificial Intelligence [2] and "IA"for Intelligence Amplification [3] and Intelligence Augmentation [1].

To conclude this abstract in one sentence, I would say that a Web Science research agenda must account for the fact that the long term potential of the Web is to augment and link all forms of intelligence.



References

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

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

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

4 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 S. Hammoudi, J. Cordeiro, L.A. Maciaszek, and J. Filipe, editors, 5h International Conference, ICEIS 2013, volume 190 of Lecture Notes in Business Information Processing, pages 3–15, Angers, France, July 2013. Springer. https://hal.inria.fr/hal-01059273.

5 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, 10th International Conference, WEBIST 2014, volume 226 of Web Information Systems and Technologies, Barcelona, Spain, April 2014. Springer International Publishing. https://hal.inria.fr/hal-01223236,doi:10.1007/978-3-319-27030-2.

6 Fabien Gandon. *Distributed Artificial Intelligence And Knowledge Management: Ontologies And Multi-Agent Systems For A Corporate Semantic Web*. PhD Thesis, INRIA, Université Nice Sophia Antipolis, November 2002. https://tel.archives-ouvertes.fr/tel-00378201

7 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. https://hal.inria.fr/hal-01560180,doi:10.1038/nbt0609-508

3.9 New ethics for the web and for the web scientist?

Katharina E. Kinder-Kurlanda (GESIS – Leibniz Institute for the Social Sciences, Cologne, DE, katharina.kinder-kurlanda@gesis.org)

Main reference Katharina E. Kinder-Kurlanda, Katrin Weller, Wolfgang Zenk-Möltgen, Jürgen Pfeffer, Fred Morstatter: "Archiving information from geotagged tweets to promote reproducibility and comparability in social media research." Big Data & Society 4(2), 2017.
URL https://doi.org/10.1177/2053951717736336

The talk looked at new challenges for ethically reflective web science research, and in particular at the challenges of changing data landscapes for epistemology and ethics. There is apparent a lack of transparency and validity, often linked to new arrangements for data access. Proprietary data 'owners' play and an important, changing and often unclear role. Questions that arise are what knowledge we can gain – and should or should not gain – from various types of data and how the validity of statements can be checked if research processes and properties of data used are mostly opaque and inaccessible for scientific peer review. It is argued that Web Science requires more innovation in research documentation and in approaches to public private partnership arrangements, such as trusted third party models that meet both researchers' and companies' interests. Trusted research support infrastructures could play an important role in this as they offer expertise in both data management and long-term preservation. A challenge here lies in the linking of different types of support infrastructure institutions, both established and emerging, and in avoiding the building of a new library (where the data scientists go) next to an existing one (where the theory books are). Haraway's concept of response-ability was used to argue for a research stance where we as web scientists are a) careful of methodology and proactive about interdisciplinarity b) create spaces for reflection and deliberation c) innovate in the ways to trigger critical thinking about ethics, epistemology and theory in teaching so as to d) not close ourselves off from research subjects, rejecting 'research at a distance'.

3.10 Why formalising fairness won't fix (algorithmic) discrimination *(reloaded)*

Bettina Berendt (KU Leuven, BE, bettina.berendt@cs.kuleuven.be)

The last ten years have seen large developments in analyses of and approaches to mitigating discrimination, bias and unfairness related to (semi-)automated decision making. In this talk, I will argue why a focus on "solving" this problem by formalising fairness and "fixing the algorithms" used in decision making is too narrow. Starting from the classical Myrdal analysis of discrimination's cumulative causation, I argue that a) the computationally grounded concept of "non-discrimination" or "fairness" neglects core aspects of the real-life goals, which rest on human agency and empowerment, b) algorithms don't discriminate, people do (and algorithms help), and c) to mitigate algorithm-related or any other discrimination, people are key (and computer science including, but importantly also beyond, algorithms can help). In addition, d) while this is a deeply ethical problem, addressing algorithm-related discrimination should not be left to "ethical approaches" alone, whether by scientists, developers, companies,

or others, but also be informed by, rely on, and cooperate with enforcement by laws and regulations.

The slides of this talk are available at https://people.cs.kuleuven.be/~bettina.berendt/Talks/berendt_2018_06_28.pdf.

3.11 World Wide Weapons: Project Maven, Google and Web Ethics

Guglielmo Tamburrini (Universita' di Napoli Federico II – Napoli, IT, tamburrini@unina.it)

In 2010 Tim Berners-Lee warned us that "The Web, as we know it, is being threatened... by governments – totalitarian and democratic alike – monitoring people's online habits, endangering important human rights". This threat is enhanced today by the growing convergence of AI, big data, and Web analytics in the design, development and deployment of increasingly autonomous weapons systems that are endowed with both monitoring and kinetic military capabilities. The Web Science community has to address these threats by reinforcing the ethical pillar of Web Science. This can be achieved by moving from reactive to proactive ethical issues identification and ethical policy development, as well as by expanding both reflective and scientific work on protecting and promoting human rights in the light of web integration with AI, big data, Internet of things and cyber-physical systems.

A specific paradigmatic example of this technological convergence and the ethical problems that it gives rise to is discussed in this talk. This is the Project Maven, undertaken in 2017 by the US Department of Defence (DoD) in cooperation with Google and other Web service actors. The project aims to achieve military target selection from drone video footage and to integrate eventually this capability into Gorgon Head and other powerful surveillance systems.

The organization of the project Maven is used to illustrate the fact that that the so-called military-industrial complex must adapt to the circumstance that today AI development is chiefly happening in the commercial sector. Significant implications of this fact for action towards the protection of fundamental human rights will be emphasized too. Indeed, commercial companies like Google are a complex multi-actor entity, whose employees have manifested deep concern for the protection of human rights and the peaceful character of their company's activities in the face of their company's involvement in Project Maven. In particular, this talk will concentrate on a contextual analysis of the ethical concerns manifested by over 3,000 Google's employees in a letter of April 4th, 2018 addressed to the company's CEO. These are issues of trust for billions of users all over the world, the company's moral reputation, and technologically possible, albeit maleficent uses of the developed technological tools beyond their declared aims. Moreover, in agreement with President Eisenhower's concerns about the lack of democratic transparency and accountability of what he called the military-industrial complex, researchers supporting Google's employees letter pointed out that Google has moved into this sort of military work without subjecting itself to public debate or deliberation, either domestically or internationally.

Similar ethical issues will be discussed in connection with security and military uses of the Amazon's face recognition system , 'Amazon Rekognition' (https://aws.amazon.com/rekognition) , which integrates video footage processing with searches in databases containing tens of millions of faces. These various ethical debates about the embedding of AI and

Web analytics techniques into military applications will be analyzed in connection with ongoing ethical debates about autonomous weapons systems and their underpinnings in both deontological and consequentialist normative ethical theorizing [1].

On the whole, these are urgent and emerging ethical issues for the Web Science research community to address. This community may play a crucial role in promoting ethical dialogue about surveillance and military uses of AI, big data, and Web analytics integration. In particular, one should promote ethical dialogue and understanding between Web scientists working on different sides of «Web experiments» and with different cultural backgrounds, much as the Pugwash conferences have been doing with the international community of nuclear physicists in the wake of the 1955 Einstein-Russell manifesto.

References

1 D. Amoroso; G. Tamburrini, *The ethical and legal case against autonomy in weapons systems*, Global Jurist, 17:3, 1–20, 2017.

4 Working Groups

4.1 Innovative methods for Web Science (Visualization group)

Susan Halford (University of Southampton, GB, Susan.Halford@soton.ac.uk)
Lynda Hardman (CWI, Amsterdam and Utrecht University, NL, Lynda.Hardman@cwi.nl)
Katharina Kinder-Kurlanda (GESIS, Cologne, DE, katharina.kinder-kurlanda@gesis.org)
Claudia Müller-Birn (Freie Universität Berlin, Berlin, DE, clmb@inf.fu-berlin.de)
Paolo Parigi (Airbnb and Stanford University, San Francisco, US, pparigi@stanford.edu)
Elena Simperl (University of Southampton, GB, e.simperl@soton.ac.uk)

This working group focused on novel research methods in Web Science to address the following challenges:

- How can we unlock the promise of interdisciplinary research inherent in the vision of Web Science?
- How do we link qualitative and quantitative research results in a more meaningful way?
- How can Web scientists engage with (proprietary) data owners on equal terms?
- How do we address issues around data availability and quality?

4.1.1 Discussed Problems

Three themes were discussed as potential solutions to approach the previous questions:

- Visualization
- Online experiments
- Participatory

4.1.1.1 Visualization can facilitate and empower interdisciplinarity

Data visualisation can help Web scientists from different disciplines work with data more effectively. It creates new ways to engage with data, helping people 'interview' it to make sense of it. It enables a productive dialogue between data, method and theory, for example

in the context of abductive reasoning. Data is often complex either by virtue of its content or quality, or because it is structured or formatted in a particular way. Visualisation helps manage some of this complexity, assisting with hypothesis formulation iteratively.

Data visualisation can also help facilitate collaborations – computer and social scientists could work together to bring data and theory into dialogue from the very beginning. They could look at the same data, ask questions, and present the outcomes in multiple ways in an abductive process. Tool support and inspiration could be drawn from the fields of visual analytics and visual data mining, where questions are derived from visual, more accessible data representations to lead to new hypotheses.

Visualization can also be a way to represent research outcomes to other audiences – academics from other fields, industry, government and the public. It helps communicate complex data and results effectively and as such it has to be an integral part of any research effort, considered and planned from the beginning and not towards the end of a project when preliminary results become available.

The group also touched upon a series of additional topics, which are listed here as they provide important research questions the Web science community could consider:

- Could we map interviewing methods using in social sciences (e.g. expert, semi-structured, open) to data exploration?
- Could we map such methods to computational techniques?
- How would one annotate visualisations of qualitative data with qualitative information?
- What would be the best ways to engage with other relevant disciplines e.g., art, design?

4.1.1.2 Online experimentation and relationships with data owners

This theme looked at how Web scientists could do valid, ethical research, especially but not only in the social sciences and when using Web data. Such efforts aim to understand and bring together core social research questions with engineering methods. The group identified three sub-themes:

- Getting access and creating new sources of data for Web science research;
- Engaging in a critical analysis of ethics, methods and effects; and
- Developing insights and understanding of industry practice

4.1.1.3 Participatory methods

In this theme the group explored the use of participatory methods as a means to

- Learn about people's views and opinions and engage with different groups and communities to develop a broader understanding of the Web and its future
- Improve the quality of critical datasets
- Approach ethical dilemmas

The group acknowledged the importance of a participatory approach to take into account multiple perspectives and encourage dialogue between them. Insights from participatory methods shape research questions and solutions. These methods are very diverse, from citizen (social) science and harnessing new and emerging forms of data (e.g., from social media) to online deliberation methods, citizens' assemblies and the use of AI techniques (for example, to enhance knowledge and understanding of the Web and extending dialogue).

4.1.2 Possible Approaches

Innovation in methods requires:

- new models for accessing and sharing data;
- better mechanisms and tools to share and reuse algorithms within academia
- building Web science data archives, data management tools (for documenting research processes throughout a project's lifetime) and infrastructures;
- novel models and frameworks to collaborate with industry, especially on data sharing;
- actively seeking collaboration across domains and being open and interested in including new disciplines in Web science.

4.2 Working Group on Values

Bettina Berendt (KU Leuven, BE)
Eirini Ntoutsi (Leibniz Universität Hannover, DE)
Evaggelia Pitoura (University of Ioannina, GR)
Steffen Staab (Universität Koblenz-Landau, DE)
Guglielmo Tamburrini (University of Naples, IT)
Pinelopi Troullinou (The Open University – Milton Keynes, GB)

This working group focused on the values of the Web Scienc community which stem from and reflect the vision of the Web Science community in accordance with human rights. It is important to acknowledge the distinctive role of Web Science Community (as stated in previous documentation) in the promotion of social goals. Therefore, the values need to reflect this. In response to the call of Sir Tim Berners-Lee to fix the Web, it is now the time to design and develop a training process for web scientists' reflectivity on ethical values. It is critical for Web Science to have/reflect on such values due to the huge impact of the Web in our society.

The discussion in this working group was triggered by the following questions:

- What is the Web we want (what are our values to be reflected on the Web)?
- How do we position ourselves in ethical and societal debates? We cannot be neutral, we need to take a stance adopting the human rights.
- What is the role of the web scientist?
- How do we make these things explicit?
- How should the values and the training process of web scientists' reflectivity on these values become part of a Web Science curriculum?

With these questions in mind, the discussion was organized according to the following topics:

- Current and foreseen future value-related problems in the Web.
- The role of technology, especially of AI as an amplifier of existing problems and as a creator of new value-related problems
- Vision of the Web Science Community on values and ethics in the Web
- A code of ethics for Web Scientists

4.2.1 Discussed Problems

4.2.1.1 Part 1: The value-related problems of the current/future Web

The group discussed a large variety of value-related problems in the Web from online harassment, misinformation, polarization/extremism, attempts to stir the public media attention using "distractions" to the normalization of such sort of behavior under your real name and the rising problem of misconducting behaviour by AI (e.g., Microsoft's Tay bot).

4.2.1.2 Part 2: The role of technology in amplifying or even triggering value-related problems

User-generated content and Web data in general are among the main enablers for AI, as one of the most successful forms of AI nowadays, machine learning (ML), relies heavily on data. Therefore, whatever biases exist in the data are also reflected in the result of these algorithms or even worse, they are amplified as ML aims to maximize predictive performance rather than ensure fairness. Such implications have been already reported for web search engines (e.g., ranking outputs problems), online services (e.g., Amazon same day prime case), online advertisements etc.

4.2.1.3 Part 3: What are our values? What is the Web we want to have?

When we speak about "values", we often think of "the protection of human rights". We endorse the human rights of the Universal Declaration of Human Rights, and their expression in the fundamental rights of our respective constitutions and similar charters. This seemingly simple and easy-to-agree-on statement already shows two key problems that a value-centric ethics for Web Science faces. First, there is only one human/fundamental right that is generally agreed upon to be non-negotiable: human dignity. All the other human rights can be, and regularly are, balanced and traded off against other rights. Second, even the conceptual movement from human rights to fundamental rights show that what is considered "basic" may depend on the jurisdiction. Thus, human rights are not enough as a guideline.

A second layer, and one that is often closer to "implementations" and more amenable to enforcement, are laws. Laws may embody fundamental rights and also already be the product of a societal balancing between different such rights. (An example is the EU's General Data Protection Regulation.) A challenge for the Web and Web Science is that the laws of different jurisdictions may be incompatible with one another; a reason for optimism is the existence of international laws (many of which are not well-known in the community). On the whole, we observe the need for more awareness of which laws exist and are applicable, and for informed discussions of what legal compliance means (and what value orientation over and above legal compliance is).

There are other values too that play a large role for Web Science, such as universal access to information (see Fabien Gandon's talk, described in Section 3.8). More debates on these values (as well as their possible limitations, see for example [**?**],are needed.

Finally, values need to be identified and discussed at different levels: that of the individual researcher (or practitioner), that of platforms, and that of the Web as a whole.

4.2.1.4 Part 4: What can we do?

Agreeing on codes of ethics is a necessary first step. The AoIR Guidelines [1] are an excellent starting point in that they were developed by and meant for *internet* researchers. However, Web Science not only does research on the Web (as a part of the internet), it also and

centrally is about *designing* that very Web. Therefore, we need methodologies to embed ethical concerns/values in design, maintenance, and revision processes. Examples are various forms of (a) value-based/sensitive design (as propagated, for example, in the IEEE Ethically Aligned Design Guidelines [2], of (b) participatory design (for an older but very comprehensive survey, see [3]), and reflective-design [4].

In addition to the initial design, affirmative/corrective actions must be provided for. Here, the question arises based on which values and laws, and "armed" with which techniques, these can be effected. For example, can copyright be used to prevent revenge porn? Finally, design thinking and methodology must provide for newly emerging situations and sudden emergencies, in which reaction speed becomes crucial.

Another field of action is the raising of awareness. This includes awareness about which fields are regulated by law. For example, the activities of bots that produce public speech content (such as Wikipedia bots) are subject to laws on public speech.

Web Science must also move beyond design and awareness-raising. We need to carry out comparative studies between countries and risk assessment exercises. We need to preview and be proactive, trying to predict things that will happen.

There are many phenomena on the Web that can only be understood properly when more disciplines are involved in these studies. These include various "human sciences" such as psychology and anthropology. Web-related phenomena that could be studied better in such interdisciplinary settings include the differences between online and offline behavior and the lack of empathy over the spatial and other distances in the Web/internet, as well as the – possibly mediating – effects of cultural proximity and distance (such as having more empathy with one's compatriots located in China than with the Chinese themselves).

4.2.2 Raising awareness – a case study on building a new social network system

To exercise our own awareness on problems, we followed a reflective exercise on building a social network system to conduct an ethical analysis (similar to Potter's box of reasoning). The goal of the exercise is to generate awareness on societal and ethical concerns that are not pre-fixed but emanate from any stage of the project activities (see d'Aquin et. al., 2018)

Our analysis went as follows:

1. Identify the functional requirements of the social network system
2. Identify the stakeholders of the network system
3. Identify the values of the Web Science community and of the different stakeholders (examine conflicting interests)
4. Make a decision on each functional requirement reflecting on the above. For example,(1) the growth of the network as a functional requirement (2) who are the stakeholders of the specific network system e.g., owner, user, non-user, web scientist, advertisers,marketeers etc.) (3) what are the values of each stakeholder and what are the potential conflicts of interest e.g., owner seeks for profit – includes everybody, web science community fights against hate speech circulation – excludes extremists. (4) decide the criteria or not on growth.
5. Reflect on contingent critical situations (e.g., shutting down Microsoft's Tay bot after it became racist).

4.2.3 Conclusions

As the Web plays a tremendous role in all aspects of our life and as each and every one of us is affected by the Web but also impacting the Web (and consequently, the society) via his/her actions, we all agreed that it is now more urgent than ever for the Web Science community to take actions to safeguard our values and promote societal good.

We outline such actions below:

- Raising awareness of the Web Science values and the role and responsibilities of each and every involved entity, from the web scientist to the platforms and the Web as a whole.
- Agree on a code of ethics for Web Scientists, possibly by drawing on existing guidelines and combining/extending them to the specifics of Web Science. Particularly relevant guidelines include: AoIR [1], IEEE [2].
- Integrate values into all stages of the knowledge discovery process, from data collection (e.g., is the dataset used for training representative of the broader population?), to preprocessing (e.g., can I use sensitive attributes or proxies to sensitive attributes in the model?), analysis (e.g., is my model also learning for the minority groups?), evaluation (e.g., is my testing set coming from the same distribution as the training set?) and interpretation of the results (e.g., how the end user personal biases/ preferences affect the interpretation?) as well as adopt model lifecycle management to ensure the model is performing well beyond its predictive performance (e.g., is the deployed model "rejecting" more and more minority instances over the course of its operation?).

References

1 AOIR (Association of Internet Researchers), Ethical Decision-Making and Internet Research: Recommendations from the AoIR Ethics Working Committee (Version 2.0). (2012) https://aoir.org/reports/ethics2.pdf

2 IEEE (2016). Ethically Aligned Design. A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems. Version 1 for public discussion. http://standards.ieee.org/develop/indconn/ec/ead_v1.pdf

3 Muller, M.J. & Druin, A. (2003). Participatory Design: The Third Space in HCI. IBM Technical Report. http://www.watson.ibm.com/cambridge/Technical_Reports/2010/TR2010.10%20Participatory%20Design%20The%20Third%20Space%20in%20HCI.pdf. Shorter version in J. A. Jacko & A. Sears (Eds.), *The Human-computer Interaction Handbook* (pp. 1051/1068). Hillsdale, NJ: Lawrene Erlbaum Associates.

4 Sengers, P., Boehner, K., David, S., & Kaye, J. (2005). Reflective Design. In *Proceedings of the 4th Decennial Conference on Critical Computing: Between Sense and Sensibility* (pp. 49–58). New York: ACM.

4.3 Working Group on Web Science and Artificial Intelligence

Fabien Gandon (Inria, Université Côte d'Azur, I3S, CNRS – Sophia Antipolis, FR, fabien.gandon@inria.fr)
Oshani Seneviratne (Rensselaer Polytechnic Institute – Troy, New York, USA, senevo@rpi.edu)
Noshir S. Contractor (Northwestern University – Evanston, US)
David De Roure (University of Oxford, GB)
Kemal A. Delic (Hewlett Packard – Grenoble, FR)
Wendy Hall (University of Southampton, GB)
Andreas Hotho (Universität Würzburg, DE)

> " 'True, and yet the pattern is fixed. Revolt, suppression, revolt, suppression – and within a century Earth will be virtually wiped out as a populated world. So the sociologists say.'
> Baley stirred uneasily. One didn't question sociologists and their computers."
> —Isaac Asimov, The Naked Sun (1957)

4.3.1 Bootstrapping the debate: members of the group and their opening statements

The working group started with a round table where every participant shared some opening statements to start the discussion. We report them here following in the alphabetical order of the participants' family name:

- *Noshir S. Contractor (Northwestern University – Evanston, US)* proposed to consider the very concrete effect of AI and the Web on work and the work force and the possible futures these two domains are drawing for work. He proposed to consider the role of AI can play in human-agent teaming and personal assistance to team members: "imagine every manager with Watson in his pocket".
- *David De Roure (University of Oxford, GB)* distinguished up front two subjects: (1) using AI as a method in conducting Web Science and (2) Web Science studying a Web that includes AI. Coming from the symbolic AI school, he reminded us that AI is also about knowledge representation, and that the whole area of the Semantic Web is currently not much discussed in Web Science. He also noted that the subject may also be discussed under many different names "knowledge graphs" and linked data. He expressed a personal interest for a Web Science studying human-agent collectives and social machines including AIs.
- *Kemal A. Delic (Hewlett Packard – Grenoble, FR)* raised the challenge of hyperscale systems and terabytes of Web data and the role of Web Science in the study of the complexity of these hyperscale systems.
- *Fabien Gandon (INRIA Sophia Antipolis, FR)* started by recalling the two meanings of the word "Web" : (1) a standardized software architecture (2) the actual hypermedia we weave worldwide. He believes that both aspects must be part of the Web Science research topics and both raise different research questions with respect to the relation of AI and the Web. As a first set a question he identified classical Web problems and tasks that can benefit from AI including: indexing, searching, browsing, etc. He then identified the special case of data on the Web insisting on both directions e.g. 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. He then moved to the challenge of inventing 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). He insisted on the fact that the domain of distributed AI (e.g. multi-agent systems, blackboards) could be instrumental in bridging AI and the decentralized Web.

- *Wendy Hall (University of Southampton, GB)* reminded us that from the beginning, Web Science was meant to be more than the study of the protocols of the Web and that one of the possible names initially considered was "network science" but was already taken. She insisted that Web Science is by nature interdisciplinary and socio-technical on a large scale. She recalled that AI is already used on the Web and will be used hugely as the Web evolves. She also pointed that we needed to talk about the future of the Internet as part of the future of Web Science.
- *Andreas Hotho (Universität Würzburg, DE)* recalled that AI is a broad area already and that this gets even worse when trying to relate it to Web Science. He stressed the importance for him of the specific relation between Machine Learning and Semantic Web to support Web science, including the application to Natural Language Processing (NLP) and Semantic Web mining. He also insisted on the key role of data and their management in the relation between AI and Web Science.
- *Oshani Seneviratne (Rensselaer Polytechnic, US)* raised the challenge of encouraging the benevolent usages of the Web and AI and preventing undesirable uses and effects. This was further specialized as a need to consider how AI can be used to perform good science research that is fair, can be reproduced easily, and the required characteristics for that such as explainable AI and transparent algorithms.

4.3.2 Propositions and reactions: brainwriting pool on Web Science and AI convergence

As a second step, we did the exercise of "Brainwriting Pool"[1] or "Consequences" using some topics mentioned during the first round-table and new topics proposed by the participants to initialize the pool.

This resulted in the following suggestions of research areas for Web Science:

- The role of Web data science to help solve problems of datasets of AI (bias, etc), to automate scrutiny on the Web and generate reports for Web scientists.
- The use of AI techniques to avoid privacy implications and issues.
- The fostering of Web exploration by using AI
- The support of reproducible research and dataset sharing as a core feature in Web Science research ("Web science AIs should be citable in Web Science papers")
- The need for "laws of Web AI" (cf Asimov) because a Web AI is at least as dangerous as a robot. The need for ethical principles of Web Science research and normative systems for hybrid societies.
- The convergence of digital, physical and artificially generated Web, and the need to use provenance to understand how content is generated
- The possibility of having AIs involved in the management of our Web Science community (e.g. propose programs, topics, etc.)

[1] https://www.mycoted.com/Brainwriting

- The design of Human-AI interaction and communication; Human-Human, Avatar-Avatar, and AI-AI interactions on the Web
- The goal of regulating subversive Human/AI activity in hybrid solutions
- The provision of an infrastructure for workflows and data as key drivers of Web Science
- The question of how adding AI to the Web re-shapes the future of work on the Web and how the demands of the future of work shapes AI on the Web.

4.3.3 Converging on three research directions

The last sessions of the group were dedicated to build a synthesis of the discussions. Three main research directions were drafted to the group identified research questions:

1. **Knowledge Infrastructure and governance of Web**: *extending the Web observatory vision*
   a. AI infrastructure to study the Web: Using the power of web scale to better understand the Web evolution phenomena.
   b. AI to detect and counter-attack some undesirable network effects: starting by defining the Web we want, how can AI Help? e.g. consider the "Giant attractors" of the Web either with the global view (crawling) to maintain metrics and then intervene or agents with simple rules pushing emergent behaviors such as an agent purposefully posting links to other platforms to foster linking, decentralizing.
   c. Workflows and data are key drivers of Web Science;
   d. 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.);
   e. Explainable AI and linked to accountability
2. **AI in the relation between Web Science and data sets/data lakes (big web science)**: *AI for data science on the Web and vice-versa*
   a. Data science to help solve problems of data sets of AI (bias, etc.) : automate scrutiny on the Web and generate reports for Web scientists, and vice versa AI simulations avoiding privacy implications and issues; Web exploration powered by AI; possible simulation and synthetic data;
   b. Reproducible research and data set sharing should be a core feature in Web Science research and the need to use provenance to understand how content is generated;
   c. Data sharing architectures / data trust archives/reserves, met searching, search across sources; Web of archives and search across that Web;
   d. Reproducibility and secondary use of the data sets
3. **Designing and studying Intelligence forms and Hybrid Web Societies**: *designing artificial Web intelligence*
   a. Different AI forms: machine learning, knowledge representation and reasoning, etc.;
   b. The "laws of Web AI" (cf Asimov) since a Web AI is at least as dangerous as a robot. Ethical principles (of Web Science research)?; normative systems for hybrid societies Human-AI interaction and communication;
   c. Human-Human, Avatar-Avatar, and AI-AI interactions; Regulating subversive Human/AI activity in hybrid solutions; Convergence of digital, physical and artificially-generated Web; How adding AI to the Web, re-shapes the future of work on the Web and how the demands of the future of work shapes AI on the Web
   d. Involve designers, interaction designers, HCI, interaction with AIs;
   e. AI to observe AI, AI watchdogs checking on other AIs.

The group also identified a number of general considerations and transversal topics:

- The manifesto could include a lexicon/glossary of preferred terms e.g. “ethically designed agents” instead of “ethical agent”;
- Alternatively the manifesto could select and explain examples of ambiguous terms and expressions to help interdisciplinary interactions;
- Should Web AI be driven by the individual values of the different communities that we represent or do we need a new set of values for the Web?

5 Panel Discussions

5.1 Closing the Loop: a panel discussion moderated by Susan

In this final session, an interdisciplinary panel of speakers was asked to reflect on the future for Web Science. The panel was comprised of:

- Dr Rob Ackland, the Australian National University, Australia
- Dr Kemal Delic, Industry and Academia
- Prof. Dave de Roure, Oxford University, UK
- Dr Oshani Seneviratne, Rennesler Polytechnic Institute, USA
- Prof. Gugliemlo Tamburinni, University of Naples Federico II, Italy

The discussion included the following points:

- the importance of maintaining methodological breadth in Web Science, to include both critical and applied methodologies;
- the importance of looking ahead, to emerging technologies and the changing formation of the online interactions, e.g. through blockchain and the internet of things;
- the importance of a ‘level playing field’ for data sharing (where at present this is largely limited to bi-lateral relations between major companies and a small number of ivy league Universities; the importance of developing new trusted mechanisms for data sharing, e.g. data trust.

5.1.1 Statements

5.1.1.1 David De Roure

I have recently been studying the broader landscape of social data science and computational social science, and considering how Web Science sits in this interdisciplinary ecosystem. The answer, not surprisngly but very importantly, is that Web Science is about the Web, i.e. a unique socially-constituted system with at least half of humanity participating. Surely this demands study in its own right: Web as an object, an evolving arefact.

During the seminar I have appreciated the nuanced discussions about ethics and normative aspects, which must be part of the study–I am pleased that this is part of the Web Science mission, and that our conversations have moved on from the early strapline “ensuring the social benefit of the Web”.

My new insights gained this week are about the various relationships between AI and the Web: AI in the everyday practice of the Web Scientist, AI as automated web-scale Web Scientist, and Web Science studying a Web wih AI inside. We need all three going forward, but I have found these distinctions useful.

Going forward I think it would be useful to discuss creativity in the Web, which is an aspect not widely addressed in Web Science. It is also topical in the context of AI, as we address creative computing but also computational creativity. There would be value in bringing humanists into our discussions.

5.1.1.2 Oshani Seneviratne

I would like to bring to attention an emerging field waiting to be explored by web scientists. Applying Web Science research methodologies for understanding the blockchain ecosystem is a subject that has fascinated me recently. Just like the Web in the early 90's, blockchain technologies are now going through a 10-20 year maturity phase. There are many parallels between the Web and the blockchain. They are both disruptive technologies that has gained massive user adoption through open architectures that promote a giant connected component, and they are both hard to quantify because of their massive scale. Similar to the early Web days, we are starting to see the rise of many applications developed using 'smart contracts' deployed on the blockchain. Some example applications include: Sapien (a democratized social news platform), Steemit (Reddit equivalent), SOLA (a decentralized Social Network with over 700,000 active users), and Indorse (a LinkedIn equivalent, where AI bots can even evaluate your skills). We are starting to see lot of public data generated through 'social' interactions on the blockchain using such applications, even though the original usecase for the blockchain was cryptocurrency.

Therefore, very much like the Web, we will start seeing the need for the next generation of scientists informed by methodologies that are similar to the ones used in Web Science to study blockchain based ecosystems. However, there will also be some challenges, as gathering the data from truly decentralized ecosystems such as those powered by the blockchain will be more challenging. But on the other hand, data sharing using the blockchain will lead to much more transparent and accountable research practises due to the immutable ledger used for data sharing transactions, which will in turn promote good data stewardship. In conclusion, it is my belief that Web Science researchers should pay some attention to this emerging area of research of blockchain based ecosystems; it is a treasure trove in unchartered territory waiting to be explored!

5.1.1.3 Kemal A. Delic

We are living in the age of 'BIG systems': Big Data, Big Infrastructures and Big Algorithms are omnipresent, always on and serving billions of users dispersed around globe. At very abstract level, this represent immensely complex cyber-physical systems which I would call 'fabrics'. This resembles a huge living organism characterized by the large diversity, high dynamics and constantly evolving behavior. Social scientists explored and tried to explain wide variety of Web phenomena during the last decade, but a new playing field is waiting for the scientific breakthroughs, deeper insights and surprising discoveries. Torrents of data flow around high-speed networks, keeping data centers around globe busy and quickly consuming space and filling up data repositories for analytic usage later. Present reality is the peta-scale world, whilst recent advances here will be heralding the exa-scale range. Those data volumes will require special algorithms – and most likely will be AI based – to enable the rise of an entire new kind of instrument – Exascale Scientific Computing (ESC). This will also represent the next big challenge for academic, industrial and government communities. It is envisaged that such kind of systems will become reality in 2020-2030-time horizon. Ultimately, this will be an extraordinary opportunity for the power of Web to explore and analyze itself for scientific, industrial and commercial purposes. Just as the invention of the microscope and the telescope have changed the course of humanity, I am strongly convinced that a web instrument will provide the foundations for the next chapter of Web Science history – evolving into the exciting, intricate and mysterious sphere of the Science of Complex Hyperscale Systems.

Participants

- Robert Ackland
Australian National University – Canberra, AU
- Ricardo A. Baeza-Yates
NTENT – Carlsbad, US & Northeastern Univ., US
- Bettina Berendt
KU Leuven, BE
- Noshir S. Contractor
Northwestern University – Evanston, US
- David De Roure
University of Oxford, GB
- Kemal A. Delic
Hewlett Packard – Grenoble, FR
- Nikolaus Forgó
Universität Wien, AT
- Fabien Gandon
INRIA Sophia Antipolis, FR
- Susan Halford
University of Southampton, GB
- Wendy Hall
University of Southampton, GB
- Lynda Hardman
CWI – Amsterdam and Utrecht University, NL
- Andreas Hotho
Universität Würzburg, DE
- Katharina E. Kinder-Kurlanda
GESIS – Köln, DE
- Claudia Müller-Birn
FU Berlin, DE
- Wolfgang Nejdl
Leibniz Universität Hannover, DE
- Eirini Ntoutsi
Leibniz Universität Hannover, DE
- Paolo Parigi
Airbnb and Stanford University – San Francisco, US
- Evaggelia Pitoura
University of Ioannina, GR
- Oshani Seneviratne
Rensselaer Polytechnic Institute – Troy, US
- Elena Simperl
University of Southampton, GB
- Steffen Staab
Universität Koblenz-Landau, DE
- Guglielmo Tamburrini
University of Naples Federico II, IT
- Pinelopi Troullinou
The Open University – Milton Keynes, GB