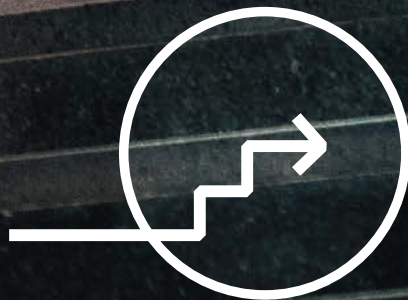


# THE ETHICS OF DIGITALISATION

---

From Principles to Practices



A PROJECT OF THE GLOBAL NETWORK  
OF INTERNET AND SOCIETY RESEARCH CENTERS (NOC)



# THE ETHICS OF DIGITALISATION

---

From Principles to Practices

A PROJECT OF THE GLOBAL NETWORK  
OF INTERNET AND SOCIETY RESEARCH CENTERS (NOC)

**FEDERAL PRESIDENT FRANK-WALTER STEINMEIER**  
**at the conference to launch the international research project Ethics of the Digital Transformation at Schloss Bellevue on 17 August 2020**

A warm welcome to Schloss Bellevue! That welcome is addressed to all of you here in this room, but of course it also includes the many people who are watching online. I'm delighted that you have all joined us!

We are aware that our lives, our interaction, our communication have acquired a new, digital dimension, not only since the COVID-19 pandemic has had the world in its grip. The pandemic, however, is the reason why today only a small group of us are present here in this room, but in fact many more guests and discussion participants are with us online. But what situation could demonstrate more clearly and urgently the issue we want to focus on today? That issue is the development of digital space.

The workplace, the classroom, the theatre, the concert hall, and indeed even parliament have moved to this digital space to avoid the virus. And all of us who have moved there with them are wondering: what are the conditions like? Are digital spaces secure and reliable? Is our privacy, is our data protected from outside interference? What rules apply, and do people respect them? We recall data scandals and Cambridge Analytica, we follow the debate on digital technology and its role in foreign policy, the disputes surrounding Huawei and TikTok.

The questions concerning how to handle the spread of digital technology have not dwindled at all over the past months and years. And now the pandemic is showing us even more clearly how closely we are connected with one another through trade and technology. The algorithm revolution, the massive consequences of digital communication constitute a global challenge. No state in the world can escape it, no state could ever be in a position to cope with it single-handedly.

That is why we need to engage in dialogue, to ask ourselves what rules exist in digital space, and what rules we want to impose on ourselves. Are we a global internet community, or are we still American, Chinese, European when we are online? What problems concern us? What can we expect from one another? And where is there common ground that we can build on? We need to ask ourselves these questions if we want to enjoy peace and prosperity in a connected world.

Two years ago, in 2018, I travelled to California and to China to trace the path of the digital revolution. On the one hand, Silicon Valley—the pioneers of the liberal, globalised data economy, whose products are used by billions of people, whose innovative potential has changed our lives and whose goal is to generate economic profit with mountains of data that are increasing by the day. On the other hand, Guangzhou and Beijing—state capitalism with huge digital ambitions, with its own internet, an almost completely separate, state-controlled system that is growing at incredible speed and renews itself on an almost daily basis—and that always has to bow to the central need for control and the pressure of surveillance from the party apparatus. And when

















## PREFACE

Digital technologies have fundamentally changed the ways in which we communicate and collaborate with each other, how entrepreneurs and businesses operate and innovate, how people express themselves and engage with the knowledge ecosystem, and how governments build systems and structures for their citizens, to enable interaction. Today, digitalisation plays a major role in almost all areas of our lives.

At their best, digital technologies can facilitate meaningful engagement among individuals, enable businesses to develop more equitable processes, support education and learning during COVID lockdowns, or help reflect on key spaces and what we as humans expect from them. At their worst, digital technologies exacerbate inequalities, amplify surveillance concerns, and strengthen existing power structures and asymmetries. As we grapple with the opportunities and challenges of technological development and deployment, it is crucial to understand the technologies themselves but more importantly to also develop a nuanced—including a global, multi-sectoral, and interdisciplinary—understanding of the underlying and overarching ethical dilemmas and priorities.

Recognizing that ethics are strongly influenced by regional, historical and cultural characteristics, they are nevertheless very useful normative tools to help shape digitalisation. Digitalisation is a process that changes entire societies, and societies are also held together by ethical principles. These change over time yet still provide societies with a foundation on which to live together and build visions of the future. However, in addition to guiding and limiting digitalisation through ethical principles, these principles must also be adapted to the new conditions of a digitized society.

Led by the Alexander von Humboldt Institute for Internet and Society (HIIG), the Berkman Klein Center at Harvard University (BKC), and the Digital Asia Hub (DAH)—under the patronage of the German Federal President Frank-Walter Steinmeier and supported by the Stiftung Mercator—and in collaboration with the Global Network of Internet & Society Centers (NoC), the “The Ethics of Digitalisation: From Principles to Practice” project tackles some of the most pressing ethical challenges with the aim of advancing dialogue and action at the intersection of science, politics, digital economy, and civil society.

From August 2020 to October 2021, NoC research teams led a series of four research sprints, four clinics, and one multi-stakeholder dialogue. The researchers illustrated the promise that an ethically sensitive approach to digitalisation can offer in the context of automated content governance and ad delivery. They illuminated the potential of a community-led approach to the use of AI in cities. They proved how digital self-determination can be an empowering approach for individuals and communities, how control over personal data can be leveraged for a greater willingness to participate in civic life and increase well-being. They emphasized the importance and feasibility of improving fairness in targeted job advertising. By focusing on real-world use cases, the researchers examined how to translate AI ethics and governance principles into practice in an

educational setting and developed systems that support a sustainable ecosystem of responsive and empowered stakeholders. They showed how digital education and literacy were essential to master today's challenges, especially in a pandemic. They provided examples of the potential of digital sovereignty for Africa. Last but not least, they developed strategies to make AI explainable and created best-practice models for explanations. Together, the outputs paint a picture of how digitalisation can be oriented towards the people.

We are proud to see the lasting impact of the project in academia, policy, and civil society, and to see the results of the project's sprints and clinics being used in practice. For instance, the formats for democratic reconnection of AI ("public sector AI") have already led to an optimization of the use of AI in school management in a European capital. The educational sector in general can also use the findings on the optimal design of digital learning spaces, due to the focus this interdisciplinary part of the project had on practical applications. Legal scholars and practitioners can benefit from the report of the XAI clinic, which provides answers to the question on what the GDPR actually requires when it comes to the explanation of automated decisions. Besides the broad impact of the results, the innovative research formats introduced and tested in the project (sprints and clinics), have proven to be very effective and are already being used in other projects.

We are deeply grateful to all contributors, especially the partnering NoCs as well as the many colleagues and experts who contributed their valuable time and expertise to the sprints and clinics. We would also like to thank the Stiftung Mercator for funding the project and making this exciting journey possible. A special incentive for all participants and a great honor is the patronage of the Federal President Frank-Walter Steinmeier for the project, for which we express our sincere gratitude.

Reimagining and reshaping the future takes time and the road to a digital society based on ethical principles is still a very long one. There is an infinite number of forks our path may take. We hope that this report may serve as a signpost at some of these forks and with it, as well as with all other outputs from the project, we were able to contribute to a more ethically-led digitalisation.



Sandra Cortesi  
Director of Youth and Media, BKC



Malavika Jayaram  
Executive Director, DAH



Wolfgang Schulz  
Research Director, HIIG



**CONTENTS**

ABOUT THE PROJECT .....16

AI AND CONTENT MODERATION .....19

DIGITAL ETHICS IN TIMES OF CRISIS:  
COVID-19 AND ACCESS TO EDUCATION AND LEARNING SPACES .....23

INCREASING FAIRNESS IN TARGETED ADVERTISING:  
THE RISK OF GENDER STEREOTYPING BY JOB AD ALGORITHMS. .... 29

DIGITAL SELF-DETERMINATION .....33

TOWARD AN AFRICAN NARRATIVE ON DIGITAL SOVEREIGNTY .....37

CHALLENGES AND OPPORTUNITIES OF PUBLIC SECTOR AI POLICY. .... 41

EXPLAINABLE AI .....45

CITIES, DIGITALISATION, AND ETHICS .....49

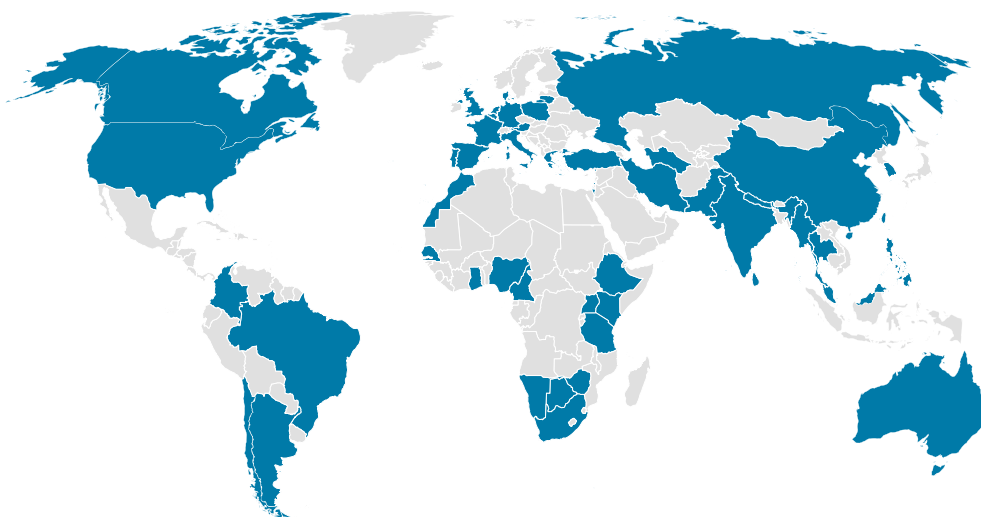
STRATEGIES FOR AN ETHICS OF DIGITALISATION. ....55

IMPRINT .....61

## ABOUT THE PROJECT

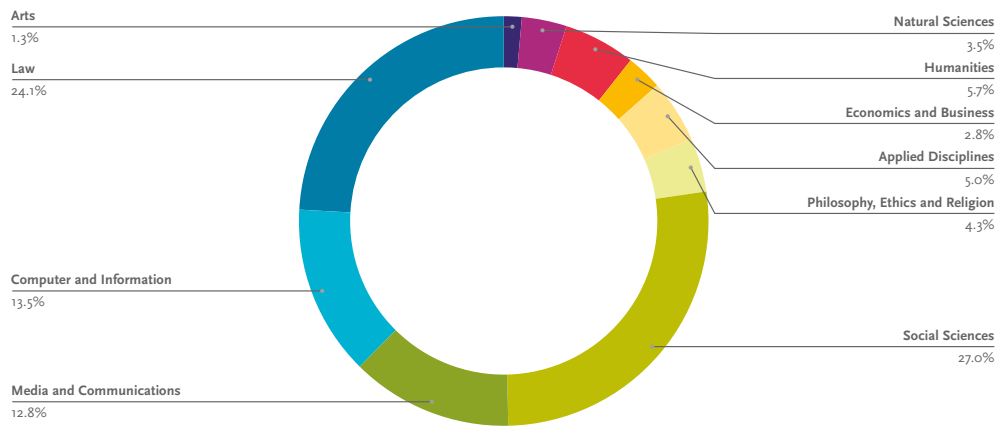
The internet is an almost infinite space full of conflicting interests—of states, individuals, and platforms, each of which are pursuing their own goals based on national or vested interests. Not all of these actors accord equal importance to preserving values and ensuring functioning societies. That is why, since its inception, the Network of Centers (NoC) has made it its mission to generate scientific knowledge in the field of digitalisation. Over the years, our researchers have found that ethical standards have not yet been established in all areas of this process. Under the patronage of German Federal President Frank-Walter Steinmeier and with financial support from the Stiftung Mercator, partnering NoCs have therefore joined forces to work on the international research project “The Ethics of Digitalisation: From Principles to Practices” in order to advance and implement ethical principles and practices in the digital space. It has piloted innovative research formats, research sprints and clinics, which have enabled interdisciplinary scientific work on application- and practice-oriented questions. The project aimed to develop groundbreaking and innovative answers to challenges in the tension between ethics and digitalisation and achieve outputs of high societal relevance and impact.

NoC research institutes cooperated with and led interdisciplinary teams of a total of 151 fellows from 51 countries spanning all continents:





The fellows had various academic backgrounds, ranging from law, sociology, and economics to computer science, political science, and philosophy:



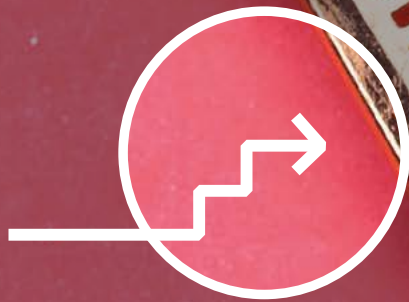
The main partners of the project include the Alexander von Humboldt Institut for Internet and Society (HIIG), the Berkman Klein Center at Harvard University (BKC), the Digital Asia Hub (DAH), and the Leibniz Institute for Media Research | Hans-Bredow-Institut (HBI). The project promotes an active exchange at the interface of science, politics, and society and intends to contribute to a global dialogue on an ethics of digitalisation.



GLOBAL NETWORK OF INTERNET AND SOCIETY RESEARCH CENTERS







RESEARCH SPRINT

# AI and content moderation

ALEXANDER VON HUMBOLDT INSTITUTE  
FOR INTERNET AND SOCIETY

AUGUST – OCTOBER 2020

## AI AND CONTENT MODERATION

In response to increasing public pressure to tackle hate speech and other challenging content, platform companies have turned to algorithmic content moderation systems. These automated tools promise to be more effective and efficient at identifying potentially illegal or unwanted material. But algorithmic content moderation also raises many questions, all of which eschew simple answers. Where is the line between hate speech and freedom of expression, and how to automate this on a global scale? Should platforms scale the use of AI tools for illegal online speech like terrorism promotion, or also for regular content governance? Are platforms' algorithms over-enforcing against legitimate speech, or are they failing to limit hateful content on their sites? And how can policymakers ensure an adequate level of transparency and accountability in platforms' algorithmic content moderation processes?

These were just some of the issues that drove the virtual research sprint on AI and content moderation, which took place virtually over the course of ten weeks from August until October 2020. In line with the project's interdisciplinary approach, the HIIG research team brought together thirteen fellows working in nine different countries across seven different time zones, whose academic expertise ranged from law and public policy to data science and digital ethics.

The fellows formed working groups to address key challenges arising from the use of automation and machine learning in content moderation. They received support by a group of mentors and also had the chance to engage with industry perspectives by meeting representatives from Facebook and Google.

In intense and thought-provoking discussions, the fellows constantly pushed the boundaries of the research sprint's format with their motivation and intellectual curiosity. Starting from the premise that algorithmic content moderation is here to stay, the fellows identified glaring gaps in our knowledge of how platform companies automate content moderation processes. Moreover, they recognized that highly imperfect machines pose grave risks to fundamental rights, particularly freedom of expression. Against this background, the working groups produced policy briefs, making recommendations on how to address these challenges across the following key areas:

*Meaningful transparency obligations:* To overcome the current information gap, the fellows propose wide-ranging measures to establish a multi-level transparency regime, facilitating evidence-based platform regulation and society-wide debate about how algorithmic content moderation systems should be designed.

*Effective appeal mechanisms:* Given a lack of redress against automated enforcement decisions, the fellows recommend imposing binding and enforceable obligations on platforms to provide users with effective appeal mechanisms. The proposals also recommend establishing an independent Ombudsperson with powers to supervise and evaluate platforms' algorithmic content moderation practices.

*Principle-based algorithmic auditing:* Lastly, the fellows identify algorithmic audits as the most promising mechanism for monitoring the risks associated with the use of AI in content moderation. To ensure carefully crafted legal mandates, the fellows recommend the four guiding principles of independence, access, publicity, and resources.

## FELLOWS

**Hannah Bloch-Wehba**, Texas A&M University, USA  
**Josh Cowls**, University of Oxford’s Internet Institute, UK  
**Philipp Darius**, Hertie School, Germany  
**Angelica Fernandez**, University of Luxembourg, Luxembourg  
**Valentina Golunova**, University of Maastricht, Netherlands  
**Aline Iramina**, University of Glasgow, UK, and University of Brasilia, Brazil  
**Sunimal Mendis**, Tilburg University, Netherlands  
**David Morar**, George Washington University, USA  
**Dominiquo Santistevan**, University of Chicago, USA  
**Charlotte Spencer-Smith**, University of Salzburg, Austria  
**Wayne Wei Wang**, University of Hong Kong, Hong Kong SAR  
**Wai Yan**, Koe Koe Tech, Myanmar

## OUTPUTS

### Policy Paper

Making Audits Meaningful—Overseeing the Use of AI in Content Moderation

*Hannah Bloch-Wehba, Angelica Fernandez, David Morar*

🌐 <https://graphite.page/policy-brief-audits/>

### Policy Paper

Disclosure Rules for Algorithmic Content Moderation—A Call for a Multi-Level Transparency Regime for Social Media Platforms

*Aline Iramina, Charlotte Spencer Smith, Wai Yan*

🌐 <https://graphite.page/policy-brief-blackbox/>

### Policy Paper

Freedom of Expression in the Digital Public Sphere—Strategies for Bridging Information and Accountability Gaps in Algorithmic Content Moderation

*Josh Cowls, Philipp Darius, Valentina Golunova, Sunimal Mendis, Erich Prem, Dominiquo Santistevan, Wayne Wei Wang*

🌐 <https://graphite.page/policy-brief-values>

## FELLOW IMPRESSIONS

“It was such a rewarding experience to work with a multicultural and interdisciplinary team of researchers. I learned a lot from the HIIG team and the other fellows. As someone new to academia who has worked for many years in the government as a regulator, to be able to participate in such an important project involving Ethics of AI, and contribute to the discussions now as a researcher, was extremely insightful.”

Aline Iramina

“The sprint helped me bridge the analytical gaps between theory and practice about platform governance. I was impressed to see a team of scientists, lawyers, and engineers making a reform-focused argument for public goods. As a lawyer, I realized that ethical directives must be translated in a way that enables the public to participate in open regulation, governance, and accountability for digital technologies.”

Wayne Wei Wang

“The highlight of my sprint experience was the energetic atmosphere of interdisciplinary collaboration. We each brought our backgrounds and expertise every week to facilitate an exchange of ideas that was creative, generative, and respectful of each other’s differences of opinions. The Internet policy landscape would be dramatically improved if interdisciplinary collaboration of this kind were the rule rather than the exception.”

Hannah Bloch-Wehba



RESEARCH SPRINT

# Digital ethics in times of crisis: Covid-19 and access to education and learning spaces

BERKMAN KLEIN CENTER FOR INTERNET & SOCIETY

OCTOBER – DECEMBER 2020

## **DIGITAL ETHICS IN TIMES OF CRISIS: COVID-19 AND ACCESS TO EDUCATION AND LEARNING SPACES**

Young people and adult learners around the globe have been affected by the impact of the COVID-19 pandemic on access to education—both in terms of educational resources, and learning spaces such as schools, campuses, museums, studios, clubhouses, afterschool, and maker-spaces. This has had unforeseen consequences for their economic futures, lives, and well-being. At this moment, digital technologies highlight both the opportunities and possible long-lasting challenges that will have profound ethical implications for decades to come. At its best, digital technology can be used during COVID lockdowns to promote and support learning across spheres and barriers. At its worst, digital technologies create new inequalities between digital haves and have-nots and amplify surveillance concerns.

The Berkman Klein Center hosted a ten-week research sprint, convening a global cohort of approximately forty student participants from twenty-one different countries over five continents. The research sprint examined the ethical, human rights, and societal aspects of digital transformation, with an emphasis on education and learning at a moment of unprecedented crisis.

The goal of the research sprint was to engage students and experts from the Global Network of Internet and Society Centers (NoC), and expert stakeholders, to create a map of the relevant issues and corresponding questions that policy-makers around the globe need to address to harness the benefits of digital technologies while avoiding some of the possible downsides during the current crisis, and as we collectively attempt to prepare better for the next crisis. As an experimental educational program, our intention was to both explore this topic in depth, while also creating a truly 'global classroom' where students from all around the world—many of whom, under normal circumstances, may not have been able to participate in such a program—could engage difficult ethical questions and other questions of digital transformation among one another, as well as with practitioners, scholars, designers, policy-makers, and industry leaders.



The output was a result of an iterative co-creation process among student participants, program staff, and experts, and represents a concise synthesis of each of the program's anchor sessions and associated themes. Some of the key findings that the report explores include inequities in access to digital technologies and the skills to use them; privacy, surveillance, and safety concerns related to education and learning; and the importance of cultivating learners' social and emotional development and well-being. For example, one student group created recommendations for key ethical, human rights, and societal aspects of digital transformation, with an emphasis on education and learning. Research Associate Alexa Hasse highlighted this final point about well-being during a final event for the program: "Given that for many of us COVID has impacted so many facets of our day-to-day lives, it may be helpful to think about well-being in a way that spans beyond just physical or mental health, to include, for instance, learning experiences, living conditions, and social interactions. And in terms of learning experiences during the sprint, we talked about how schools serve not only educational functions, but also provide environments and services that can promote well-being, ranging from meeting basic needs like food through school lunches, to providing rich social and emotional learning opportunities."

## FELLOWS

**Hamdalat Alabi**, Carnegie Mellon University Africa, Nigeria  
**Valerie Albrecht**, Danube University Krems, Austria  
**Mudasir Amin**, Jamia Millia Islamia, India  
**Sara Bubenik**, Boston University, USA  
**Daniel Calarco de Oliveira**, International Youth Watch, Brazil  
**Bernardo Caycedo**, Pontificia Universidad Javeriana and Researcher at Centro de Internet y Sociedad de la Universidad del Rosario (ISUR), Colombia  
**Sidharth Chauhan**, Harvard Law School, USA  
**Phoebe Chua**, University of California, Irvine, USA  
**Tomas Dodds**, University of Amsterdam and University of Leiden, Netherlands  
**Elora Raad Fernandes**, Rio de Janeiro State University, Brazil  
**Martin Fertmann**, University of Hamburg and Leibniz Institute for Media Research | Hans-Bredow-Institut, Germany  
**Dilrukshi Gamage**, University of Moratuwa and Diversity Collective Lanka, Sri Lanka  
**Sakshi Ghai**, University of Cambridge, UK  
**Tomasz Hollanek**, University of Cambridge, UK  
**Milan Ismangil**, Chinese University of Hong Kong, Hong Kong SAR  
**Catherine Keegan**, University College London, UK  
**Daum Kim**, Keio University and Ethnic Neighborhoods, Japan  
**Swathi Krishnaraja**, Weizenbaum Institute for Networked Society, Germany  
**Benedict Lang**, ETHOS Lab, ITU Copenhagen, Denmark  
**Enze Liu**, Fudan University, China  
**Fang-ying Riva Lo**, Asia University, Taiwan  
**Sharu Luo**, National Tsing Hua University Institute of Law for Science and Technology, Taiwan  
**Maya Malik**, McGill University School of Social Work, Canada  
**Sri Ranjani Mukundan**, NUS Singapore, Singapore  
**Arnel F. Murga**, The Asia Foundation, USA  
**Musa Ndahi**, Nasarawa State University, Nigeria  
**Sarah Nizamani**, Institute of Business Administration, Pakistan  
**David Otoo-Arthur**, University of the Witwatersrand, Johannesburg, South Africa, and Presbyterian Women's College of Education, Aburi, Ghana  
**Sachini Perera**, King's College London, UK  
**Atandra Ray**, Charles University in Prague, Czech Republic  
**Alexis Shore**, Boston University, USA  
**Vince Straub**, Oxford Internet Institute, UK  
**Sadaf Taimur**, The University of Tokyo, Japan  
**Santiago Uribe**, Nordic Centre for Internet and Society and BI-Norwegian Business School, Norway  
**Laura Garcia Vargas**, University of Ottawa Centre for Law, Technology and Society, Canada  
**Clara Wang**, Peking University, China  
**Janis Wong**, University of St Andrews and Open Data Institute, UK  
**Jingyi Yu**, Fudan University, China

## OUTPUTS

### Playbook

Participants in the Ethics of Digitalisation Research Sprint: Digital Ethics in Times of Crisis: COVID-19 and Access to Education and Learning Spaces

📄 <https://cyber.harvard.edu/sites/default/files/2021-02/Digital%20Ethics%20In%20Times%20of%20Crisis%20Report.pdf>

### Video

Participants in the Ethics of Digitalisation Research Sprint: Digital Ethics in Times of Crisis: COVID-19 & Access to Education and Learning Spaces

📺 <https://cyber.harvard.edu/story/2021-02/video-re-imaginings-covid-19-and-access-education-and-learning-spaces>

## FELLOW IMPRESSIONS

“I explored the need to invest in low-tech/no-tech initiatives for marginalized communities, as this would minimize the impact on learning due to COVID-19. I was able to relate it to my country, India, and realized how similar or related our issues were! I was able to share my knowledge about the best practices employed around the world with my community.”

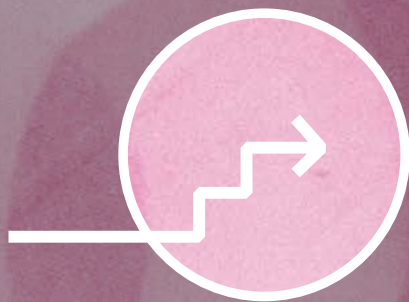
**Sidharth Chauhan**

“The highlight of the sprint was the unique collaboration opportunities with emerging scholars from around the world. There was so much energy within the interdisciplinary approaches that fueled creative outputs. My biggest takeaway from the sprint experience was learning the importance of international perspectives and how to prioritize those in my own research and career.”

**Alexis Shore**

“I really enjoyed being part of global and interdisciplinary teams because it allows our work to be more representative of different socio-geographic contexts. I was able to extend my cross-collaboration opportunities and continue to speak with those that I participated in the research sprint with.”

**Janis Wong**



CLINIC

# Increasing fairness in targeted advertising: The risk of gender stereotyping by job ad algorithms

ALEXANDER VON HUMBOLDT INSTITUTE  
FOR INTERNET AND SOCIETY

FEBRUARY 2021

## **INCREASING FAIRNESS IN TARGETED ADVERTISING: THE RISK OF GENDER STEREOTYPING BY JOB AD ALGORITHMS**

Who gets to see what on the internet? And who decides why? These are among the most crucial questions regarding online communication spaces, and they especially apply to job advertising online. Targeted advertising on online platforms offers advertisers the chance to deliver ads to carefully selected audiences. However, optimizing job ads for relevance also carries risks, from problematic gender stereotyping to potential algorithmic discrimination.

The virtual clinic examined the ethical implications of targeted advertising, bringing together twelve fellows from six continents and eight disciplines engaged in an interdisciplinary solution-oriented process facilitated by a project team at the HIIG. The fellows also had the chance to learn from and engage with a number of leading experts on targeted advertising who joined the clinic for thought-provoking spark sessions. The objective of the clinic was to produce actionable outputs that contribute to improving fairness in targeted job advertising. The fellows developed three sets of guidelines covering the whole targeted advertising spectrum. While the guidelines provide concrete recommendations for platform companies and online advertisers, they are also of interest to policymakers.

The first set of guidelines focuses on ad targeting by advertisers. This stage of the targeted advertising process involves creating the ad, selecting the target audience, and choosing a bidding strategy. In light of the variety of targeting options, researchers have voiced concerns about potentially discriminatory targeting choices that may exclude marginalized user groups from receiving, for example, job or housing ads, increasing marginalization in a “Matthew effect” of accumulated disadvantage. Although discrimination based on certain protected categories, such as gender or race, is prohibited in many jurisdictions, and despite platforms such as Google and Facebook restricting sensitive targeting features in sectors like employment and housing, problems persist due to problematic proxy categories (like language or location). The fellows address these challenges by calling for a legality by default approach to ad targeting and a feedback loop informing advertisers about potentially discriminatory outcomes of their ad campaigns.

The second set of guidelines centers on ad delivery by platforms, which mainly consists of auctioning ads and optimizing them for relevance. Research has shown that ad delivery can still be skewed along gender lines, even when advertisers are careful not to exclude any kind of user group from their ad campaign. This can be partially explained by market effects. Younger women, for instance, are more likely to engage with ads, and are therefore more expensive in ad auctions. Platforms also optimize for relevance based on past user behavior, so gender stereotyping is likely to happen with respect to historically male or female dominated employment sectors. To address this, the fellows developed a user-centered approach in their guidelines, enabling users to be in charge of their own advertising profiles.

The third set of guidelines addresses how ads are displayed to users. Currently, users can not usually look behind the scenes of targeted advertising and understand why they see certain ads and not others. Existing transparency initiatives by platforms still fall short of providing users with meaningful transparency. The proposed Digital Services Act imposes online advertising transparency obligations on online platforms, but these provisions have yet to become law. The fellows propose an avatar-solution in their guidelines—a user-friendly, gamified tool to visually communicate the information collected by the platform and the attributes used to target the user with job ads.

## FELLOWS

**Ezgi Eren**, The University of Edinburgh, UK  
**Lukas Hondrich**, AlgorithmWatch, Germany  
**Linus Huang**, University of Hong Kong, Hong Kong SAR  
**Basileal Imana**, University of Southern California, USA  
**Joanne Kuai**, Karlstad University, Sweden  
**Marcela Mattiuzzo**, University of São Paulo, Brazil  
**Sylvie Rzepka**, University of Potsdam, Germany  
**Marie-Therese Sekwenz**, University of Vienna, Austria  
**Zora Siebert**, Heinrich Böll Foundation, Germany  
**Sarah Stapel**, University of Amsterdam, Netherlands  
**Ana Pop Stefanija**, Vrije Universiteit Brussels, Belgium  
**Franka Weckner**, University of Heidelberg, Germany

## OUTPUT

### Clinic report

Increasing Fairness in Targeted Advertising—The Risk of Gender Stereotyping by Job Ad Algorithms

*Nadine Birner, Shlomi Hod, Matthias C. Kettemann, Alexander Pirang, and Friederike Stock (Eds.)*

🌐 <https://graphite.page/fair-targeted-ads/>

## FELLOW IMPRESSIONS

“The Virtual Clinic was quite stimulating and enjoyable. It was an eye-opening experience to meet and brainstorm with wonderful colleagues from various disciplines. Through the clinic, I understood that an enabling mindset and interdisciplinary collaboration lead to realistic, practical solutions to the pressing issues in the field of targeted job ads. I am truly happy to have been a part of it.”

**Ezgi Eren**

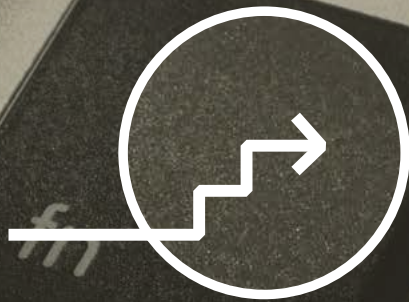
“The biggest highlight for me was how the interdisciplinary nature of the collaboration helped me think beyond my technical lens and hear new perspectives from others regarding solutions to bias in targeted advertising. Each person brought a unique perspective and useful insights that contributed to the final outcome of the clinic.”

**Basileal Imana**

“The only way to build truly empowering technologies for individuals is to merge the knowledge of all included parties. This Clinic taught me that real solutions emerge when an interdisciplinary team of bright minds works together. And that’s my favorite way of collaborating to save the future! I am so grateful for this experience!”

**Ana Pop Stefaniija**





RESEARCH SPRINT

# Digital self-determination

DIGITAL ASIA HUB

MARCH – MAY 2021

## DIGITAL SELF-DETERMINATION

As the world becomes increasingly digitally networked, there is a need for a deeper examination and understanding of important dimensions of self-determination, from control over personal data to self-expression, participation in civic life and the digital economy, relationship-building, and our health and well-being.

From March to May 2021, the Digital Asia Hub and the Berkman Klein Center for Internet and Society, in collaboration with partners of the Global Network of Internet and Society Centers (Noc), launched a research sprint focused on digital self-determination. The virtual program convened twenty-five student participants from twenty-one countries over six continents. Through engaging sessions with expert speakers and ongoing projects, participants engaged in critical dialogue on how to define and understand the problems and potential of digital self-determination.

Over the course of the sprint, two outputs were identified that were reflective of the current debate over digital self-determination and related concepts. The first output was to create a repository of learning artifacts that explore and consider digital self-determination from different perspectives, such as governments, small businesses, NGOs, communities, marginalized identities, and individuals, and place it in a space where it can be both widely accessed and used, and added to. The repository was placed on Wikiversity, Wikimedia's open educational resource platform. The second output was focused on creating the inaugural entry on digital self-determination on Wikipedia. Both outputs were intended to create opportunities to start a discussion about the future of self-determination in the digital age, and the design of these outputs was guided by the need to be open to others to add to this discussion.

A newsletter was launched for the research sprint to share findings in real-time with a broader audience. This newsletter was shared for the duration of the sprint and provided highlights and participants' artifacts that were later put into the open repository.

The resources created by the participants, in collaboration with experts and mentors, serve as a living repository to inform and support thematic networks and future discourse on the theory and practice of digital self-determination from a diverse set of interdisciplinary and global perspectives.

## FELLOWS

**Karolina Alama-Maruta**, Queen Mary University of London, UK

**Kawsar Ali**, Macquarie University, Australia

**Rachid Benharrouse**, University of Mohamed V, Morocco

**Hei Yin Chan**, University of North Carolina at Chapel Hill, USA

**Ana Margarida Coelho**, Catholic University of Portugal, Portugal

**Leonid Demidov**, Turkmen National Institute of Foreign Languages, Turkmenistan

**Maria Francesca De Tullio**, University of Antwerp and University of Naples, Italy

**Alexandra Giannopoulou**, University of Amsterdam, Netherlands  
**Tomás Guarna**, Massachusetts Institute of Technology, USA  
**Martyna Kalvaitytė**, Sciences Po Paris, France  
**İdil Kula**, Middle East Technical University, Turkey  
**Zachary Marcone**, Yenching Academy of Peking University, China  
**Derguene Mbaye**, Université Cheikh Anta Diop de Dakar, Senegal  
**Hillary McLauchlin**, University of Oxford, Oxford Internet Institute, UK  
**Samreen Mushtaq**, Ashoka University, India  
**Areej Mawasi**, Arizona State University, USA  
**Narayanamoorthy Nanditha**, York University, UK  
**Carmen Ng**, Technical University of Munich, Germany  
**Oluwatimilehin Olagunju**, University of Lagos, Nigeria  
**Temitayo Olofinlua**, University of Ibadan, Nigeria  
**Mary Rhauline Torres**, Harvard Law School, USA  
**Jean-Baptiste Scherrer**, Panthéon-Sorbonne University, France  
**Eraldo Souza Dos Santos**, Panthéon-Sorbonne University, France  
**Christian Thönnies**, Max Planck Institute for the Study of Crime, Germany  
**Constanza Vidal Bustamante**, Harvard University, USA

## OUTPUTS

### Wikipedia page

Participants in the Ethics of Digitalisation Research Sprint: Digital Self-Determination

🌐 [https://en.wikipedia.org/wiki/Digital\\_self-determination](https://en.wikipedia.org/wiki/Digital_self-determination)

### Syllabus on Digital Self-Determination

Participants in the Ethics of Digitalisation Research Sprint: Digital self-determination:

A living syllabus

🌐 [https://en.wikiversity.org/wiki/Digital\\_self-determination](https://en.wikiversity.org/wiki/Digital_self-determination)

### Video

Participants in the Ethics of Digitalisation Research Sprint: Digital self-determination research sprint showcase

🌐 <https://cyber.harvard.edu/story/2021-06/video-digital-self-determination-research-sprint-showcase>

## FELLOW IMPRESSIONS

“Having done much of my graduate coursework in isolation over the course of the past year, I found the Research Sprint’s collaborative nature and discussions to be incredibly refreshing. The Sprint illustrated what remote learning is at its best—bringing together students and experts from across the globe to share perspectives and work towards a common project. I’m feeling all the more energized as I return to my dissertation and continue to think about the many questions highlighted during the Sprint.”

Hillary McLauchlin

“We have to peel beneath the gender and classism that exists in our different communities. I am quite inquisitive about how this works in Nigeria: What does it mean to have access to the internet as an empowered Nigerian woman? What are the threats? What are the opportunities? What are the challenges?”

Temitayo Olofinlua

“The Sprint is composed of an amazing group of people from all over the world with so many diverse backgrounds. There are students studying psychology, communication, political science, social science, and culture. The diversity of language, culture, and studies makes the task of understanding digital self-determination challenging, because we all have different contexts. But there’s also beauty in it because it allows us to find connections, confront ourselves with different, sometimes opposing perspectives, and accept and respect that.”

Mary Rhauline Torres



RESEARCH SPRINT

# Toward an African narrative on digital sovereignty

UNIVERSITY OF JOHANNESBURG

JUNE – JULY 2021

## TOWARD AN AFRICAN NARRATIVE ON DIGITAL SOVEREIGNTY

The debate on the digital economy is heating up. Many questions abound. Will robots displace jobs? Is there a new colonialism of data? Will large platform companies push out traditional production and sales in countries or can all businesses prosper? And if so, how can we create the right balance between the greater use of digital technologies, and the threats of data extraction and commodification, the rising costs of innovation, and digital surveillance? This research sprint focused on digital sovereignty in Africa to find solutions to help realize the national and individual interests of citizens in the digital economy across Africa, and assist African countries in leveraging their own unique advantages.

The South African Research Chair in Industrial Development at the University of Johannesburg hosted an eight-week sprint and brought together twenty-five international fellows; a mix of academics and practitioners from fourteen African countries, across a wide range of disciplines and focus areas. The fellows were invited to empathize with users, technologists, and policy makers, exploring important questions and solutions, including technology tools, from an African perspective, articulating an internal African vision for development in the digital age.

Particularly in the global South, there is little clarity on the term digital sovereignty and its application. This research sprint, the first of its kind in Africa, was designed to focus on what digital sovereignty could mean in Africa, with the intent of extracting the key elements of a pan-African narrative on digital sovereignty. Key issues considered in the sprint include:

- In the digital economy, how, and to what end, can citizens and states reassert control? Is digital sovereignty a useful concept, and if it is, what could be the meaning and import of digital sovereignty in the global South, and specifically in Africa?
- Can there be economic autonomy and a break away from technological dependence without political autonomy on the one hand, and data infrastructure and data control on the other?
- Can a collective capacity for states, individuals, and communities to engage in technological development be created, and if so, how?
- Do current developments in Africa reflect or build towards a sovereign, pan-African vision for development and economic independence in the digital age?
- In such a vision, how can data extraction, data use, and data re-use foster the creation of competitive advantage, innovation, and technological learning, enabling local businesses, creating jobs, and promoting structural change in Africa?
- What sort of relationships between citizens and states could enable such a developmental model?
- How can we frame a new discourse that factors in development as a central component of the data economy, taking into account the different starting points of countries as they enter and engage with data?

In grappling with these broader framing questions, the sprint addressed linguistic and cultural heterogeneity in the internet world, and the need for a homegrown narrative on privacy, informed consent, and data protection in Africa.

The outputs engage with a range of topics on digital sovereignty in Africa and offer a rich set of perspectives. They have a high degree of policy relevance and provide fresh insights into key issues relating to digital economies, digital transformation, and data access in governance in Africa. The study is also relevant to developing countries more broadly. We hope that these contributions will stimulate further research in this field, while also having value for policymakers and other stakeholders. Ultimately, the volume seeks to contribute to developing a distinctly African narrative on the topic of digital sovereignty—a topic that is likely to become increasingly important in years to come.

## FELLOWS

**Halefom Abraha**, University of Malta, Malta  
**Benjamin Akinmoyeje**, Namibia University of Science and Technology, Namibia  
**Peace Amuge**, Women of Uganda Network, Uganda  
**Michael Asiedu**, University of St. Gallen, Switzerland  
**Ayça Atabey**, University of Edinburgh, UK  
**Olusesan Michael Awoleye**, Obademi Awolowo University, Nigeria  
**Odilile Ayodele**, Independent, South Africa  
**Ngwinui Azenui**, Denison University, USA  
**Bridget Baokye**, Tony Blair Institute for Global Change, UK  
**Blaise Bayuo**, Tony Blair Institute for Global Change, UK  
**Ibtissam Chafia**, Mohammed 6 Polytechnic University/OCP Group, Morocco  
**Adio-Adet Dinika**, University of Bremen, Germany  
**Winnie Kamau**, Talk Africa, Kenya  
**Animata Kidiera**, Gaston Berger University, Senegal  
**Tarirayi Machiwenyika-Mukabeta**, Bindura University of Science Education, Zimbabwe  
**Julius Mboizi**, Harvard Law School, USA  
**Peter Mmbando**, Southern Africa Youth Forum, Tanzania  
**Oarabile Mudongo**, Research ICT Africa, University of the Witwatersrand, South Africa  
**Sylvia Mutua**, Communication University of China, China  
**Jacqueline Mwangi**, Harvard Law School, USA  
**Lydia Namugabo**, University of South Africa, South Africa  
**Fatih Obafemi**, Future Proof Intelligence, Nigeria  
**Emma Ruiters**, Genesis Analytics, South Africa  
**Bendjedid Rachad Sanoussi**, Internet Society/KNUST, Benin  
**Sadrag Shihomeka**, University of Namibia, Namibia

## OUTPUT

### Study

Digital Sovereignty: African Perspectives

*Padmashree Gehl Sampath, Fiona Tregenna (Eds)*

🌐 <https://www.hiig.de/en/project/the-ethics-of-digitalisation/>

### Blog Series

Outcomes of the Virtual Research Sprint

*Padmashree Gehl Sampath, Fiona Tregenna (Eds)*

🌐 <https://digitalsovereigntyafrika.wordpress.com/>

### Wikiversity Syllabus

Digital Sovereignty in Africa: An Open Syllabus

*Participants in the Ethics of Digitalisation Research Sprint: Toward an African Narrative on Digital Sovereignty*

🌐 [https://en.wikiversity.org/wiki/Digital\\_Sovereignty\\_in\\_Africa:\\_An\\_Open\\_Syllabus](https://en.wikiversity.org/wiki/Digital_Sovereignty_in_Africa:_An_Open_Syllabus)

## FELLOW IMPRESSIONS

“It was a nice experience to join colleagues and other professionals from different and related backgrounds in the clinic on Ethics of Digitalisation. The Sprint was an eye opener towards the concern of digital sovereignty and has set the thought on what Africa needs to put in place to avert digital colonization in this era of the 4IR.”

Olusesan Michael Awoleye

“The sprint improved my awareness of digital sovereignty. In addition, it provoked me to think about the implications of digital sovereignty on the transformation of African economies. The diverse background of scholars in my team, other fellows, and speakers, made for rich, dynamic discussions and lasting connections. My main take-away was that we cannot discuss digital sovereignty without talking about digitalisation.”

Ngwinui Azenui

“The opportunity to engage in conversations around fundamental issues of Africa's digital sovereignty with an interdisciplinary team was priceless. I went away with the conviction that to strengthen digital sovereignty, Africa needs to transition from being just a consumer to also becoming a creator in the digital economy. It should be a two-way digital street, give and take.”

Fatih Obafemi





CLINIC

# Challenges and opportunities of public sector AI policy

BERKMAN KLEIN CENTER FOR INTERNET & SOCIETY

JULY 2021

## CHALLENGES AND OPPORTUNITIES OF PUBLIC SECTOR AI POLICY

The Berkman Klein Center, in collaboration with the City of Helsinki's Education Division and the AI-transparency company Saidot, hosted a virtual research clinic on AI policy to study the ethical governance of artificial intelligence-enhanced technologies deployed to support learning, student wellbeing, and retention in Helsinki's vocational schools. The clinic convened a dozen early-career scholars to examine a real-world use case of AI in the public sector.

During the first portion of the month-long program, the clinic explored questions related to the creation of an inclusive, participatory, and sustainable strategy for stakeholder engagement, throughout the design, development, deployment, and assessment stages of the new technology. During the second half, the student cohort examined the viability and appropriateness of different human oversight mechanisms. In both instances, the primary goal was to create resources to help the City of Helsinki (and similarly situated municipalities) navigate thorny AI governance challenges related to participatory design and human oversight.

Divided into two working groups, the first group produced four distinct outputs: a human oversight model to enhance cooperation across the City of Helsinki's technical, operational, and governance layers; a translational matrix for different stakeholders based on ethical and regulatory requirements in the European Union; and a wireframe, alongside explanatory documentation, for an accountability web-portal to facilitate public participation, transparency, and the work of individuals tasked with the human oversight of AI tools. The group consolidated their background research, justification for each resource, and implementation recommendations into a policy playbook.

The second group adapted an existing method, deployed in Catalonia by Cobo Lab, to the City of Helsinki's requirements and the specifics of the AI technology. In an extensive playbook, the students present the essential, recommended, and contingent elements of a four-phase participatory process for the introduction of AI technologies. The resource is a detailed, step-by-step guide to integrating participatory and accountability elements into the design, development, and deployment process of public-sector technology.

The first group's efforts drew heavily from AI-specific ethical guidelines and practices, while the second group adapted a more holistic approach to a use case involving AI. Ultimately, both approaches emphasized that the deployment of AI in municipal services should be a continuous exercise, with constant input from stakeholders, technical teams, and the general public, as new challenges and concerns arise. By focusing on real-world use cases, the program examined how to translate AI ethics and governance principles into practice in an educational setting. These systems support a sustainable ecosystem of responsive and empowered stakeholders.

## FELLOWS

**Yung Au**, University of Oxford, UK

**Nagadivya Balasubramaniam**, Aalto University, Finland

**Bruna de Castro e Silva**, Tampere University, Finland  
**Karolina Maria Drobotowicz**, Aalto University, Finland  
**Erika Ly**, Australian National University, Australia  
**G.R. Marvez**, Massachusetts Institute of Technology, USA  
**Franziska Poszler**, Technical University of Munich, Germany  
**Kaivalya Rawal**, Harvard University, USA  
**Giulia Schneider**, Sant'Anna School of Advanced Studies, Italy  
**Vera Vidal Rougier**, Open University of Catalonia, Spain  
**Thomas Vogl**, University of Oxford, UK  
**Marta Ziosi**, University of Oxford, UK

## OUTPUTS

*Various Outputs by Brunna de Castro e Silva, Erika Ly, Franziska Poszler, G. R. Marvez, Nagadivya Balasubramaniam, Thomas Vogl:*

### **Human Oversight Model**

Translational Model for Human Oversight Measures

🔗 [https://cyber.harvard.edu/sites/default/files/2021-08/FINAL\\_Comprehensive%20Human%20Oversight%20Translational%20model.pdf](https://cyber.harvard.edu/sites/default/files/2021-08/FINAL_Comprehensive%20Human%20Oversight%20Translational%20model.pdf)

### **Translational Matrix**

AI Ethics Systemic Translational Matrix for AI and Learning Analytics at Vocational Education and Training (VET) in Helsinki

🔗 [https://cyber.harvard.edu/sites/default/files/2021-08/FINAL\\_Ethics%20Systemic%20Translational.pdf](https://cyber.harvard.edu/sites/default/files/2021-08/FINAL_Ethics%20Systemic%20Translational.pdf)

### **Wireframe**

Helsinki City of EdTech

🔗 [https://cyber.harvard.edu/sites/default/files/2021-08/20210811\\_Mockups\\_BKC\\_Helsinki.pdf](https://cyber.harvard.edu/sites/default/files/2021-08/20210811_Mockups_BKC_Helsinki.pdf)

### **Policy Playbook**

Municipal Stakeholder Engagement Strategies for Learning Analytics and AI in Education: Participatory Design, Accountability and Oversight Mechanisms

🔗 [https://cyber.harvard.edu/sites/default/files/2021-08/AI%20Policy%20Research%20Clinic%20-%20Policy%20Paper\\_wg2.pdf](https://cyber.harvard.edu/sites/default/files/2021-08/AI%20Policy%20Research%20Clinic%20-%20Policy%20Paper_wg2.pdf)

### **Playbook**

The Playbook on Participation and Accountability in City Challenges

*Karolina Drobotowicz, Vera Vidal, Marta Ziosi, Yung Au, Kaivalya Rawal, Giulia Schneider*

🔗 [https://cyber.harvard.edu/sites/default/files/2021-08/Playbook\\_participation\\_AI\\_wg2.pdf](https://cyber.harvard.edu/sites/default/files/2021-08/Playbook_participation_AI_wg2.pdf)

## FELLOW IMPRESSIONS

“The highlight of the clinic experience was applying my academic knowledge and expertise to a real-life policy challenge. The need to adapt one’s critical thinking and problem-solving skills to tangible constraints and variables such as time, resources, and best interests, in the most actionable way, was an immeasurably valuable exercise.”

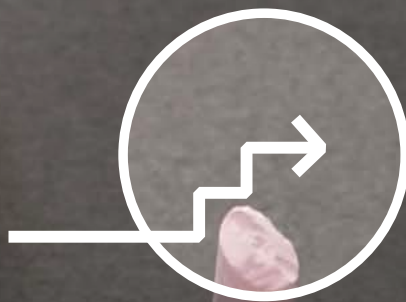
**Bruna de Castro e Silva**

“The cultural and experiential diversity of researchers during the AI Policy Research Clinic was both a challenge and a treat! Learning from others and overcoming language problems to provide a sensible response to the real-life challenge was a very satisfactory task. The clinic also offered me a great lesson in what the challenges and attitudes can be in public sector AI teams; I appreciated a lot that we could work with the people who have a real impact on society.”

**Karolina Drobotowicz**

“Developing ethical AI, especially in municipal settings, can involve a tradeoff between participatory design and swift implementation. For example, if AI is deployed in the education system, citizens are mandated to engage with it in order to engage with public education. It becomes ever more important to include affected individuals in the design of this new technology.”

**Franziska Poszler**



CLINIC

# Explainable AI

ALEXANDER VON HUMBOLDT INSTITUTE  
FOR INTERNET AND SOCIETY

SEPTEMBER 2021

## EXPLAINABLE AI

The opacity of machine-learning algorithms and the functioning of (deep) neural networks make it difficult to adequately explain how AI systems reach results ('black box phenomenon'). Calls for more insights into how automated decisions are made have grown increasingly loud over the past couple of years. The solution seems clear: We need enough understanding of automated decision-making processes to be able to provide the reasons for a decision to those touched by that same decision, and in a way they understand. Explainability is the necessary first step in a row of conditions which lead to a decision being perceived as legitimate; decisions which can be justified are perceived as legitimate. However, only what is questioned is justified, and only what is understood is questioned. And to be understood, a decision has to be explained. Thus, explainability is a precondition for a decision to be perceived as legitimate (justifiability). Given these circumstances, it is difficult to ensure that the power of AI can be harnessed for good. It is also difficult to obtain explanations of how decisions are reached, despite this being a requirement under European laws such as the GDPR.

The on-site clinic on the topic of "Explainable AI" convened twelve expert participants with interdisciplinary backgrounds for a four-day journey, focusing on how to provide meaningful explanations of AI-based decisions from a legal, design, and technical perspective. The paper the participants produced asked three questions: Who needs to understand what in a given scenario? What should explanations look like to be meaningful to affected users? What do we know about the systems in place to provide convincing explanations?

The outcomes are intended to advance the debate among legal scholars and help developers and designers understand legal obligations when developing or implementing an ADM system. These are some of the key findings:

From a legal perspective, the explanation has to enable the user to appeal the decision made by the ADM system. "The logic" can be understood as "the structure and sequence of the data processing". This need not necessarily include a complete disclosure of the entire technical functioning of the ADM system. Since the explanation is intended to balance the power of the ADM developer with that of the user, this balance has to be at the center of the explanation. The GDPR focuses on individual rather than collective rights. This is the subject of many discussions among scholars. However, the interpretation of the GDPR as protecting mainly individual rights is just the minimum requirement for an explanation. Any explanation that goes further and also has the protection of collective rights in mind will be compliant with the GDPR as long as the individual's rights are also protected. Therefore, we recommend putting the individual at the center of the explanation as a first step to comply with the GDPR.

With regard to the question "What should explanations look like?", clinic participants argued that XAI is more than just a technical output. According to their view, XAI has to be understood as a complex communication process between human actors and cannot be merely evaluated in terms of technical accuracy. Evaluation of the communication process should be accompanied by evaluation of the ADM system's technical performance.

Furthermore, transparency at the input level is a core requirement to mitigate potential bias, as post-hoc interpretations are widely perceived as too problematic to tackle the root cause. The focus should therefore shift to transparency in the underlying rationale, design, and development process.

Finally, participants demonstrated that a gap exists between how developers and legal experts define explanations. Understanding “the logic” of such diverse systems requires action from different actors and at numerous stages, from the conception to the deployment of AI. Documenting the input data is part of the “logic involved” from a technical perspective. Explanation is not easy, and methods to explain the explanation often involve using additional approximate models with potentially lower accuracy. Therefore, participants argued, the overall XAI process should involve direct and indirect stakeholders from the very beginning.

## FELLOWS

**Hadi Asghari**, Alexander von Humboldt Institute for Internet and Society, Germany

**Johannes Baeck**, Continental, Germany

**Aljoscha Burchardt**, German Research Center for Artificial Intelligence, Germany

**Judith Faßbender**, Alexander von Humboldt Institute for Internet and Society, Germany

**Nils Feldhus**, German Research Center for Artificial Intelligence, Germany

**Freya Hewett**, Alexander von Humboldt Institute for Internet and Society, Germany

**Vincent Hofmann**, Leibniz Institute for Media Research I Hans-Bredow-Institut, Germany

**Matthias Kettemann**, Leibniz Institute for Media Research I Hans-Bredow-Institut, Germany

**Wolfgang Schulz**, Leibniz Institute for Media Research I Hans-Bredow-Institut, Germany

**Judith Simon**, Universität Hamburg, Germany

**Jakob Stolberg-Larsen**, Alexander von Humboldt Institute for Internet and Society, Germany

**Theresa Züger**, Alexander von Humboldt Institute for Internet and Society, Germany

## OUTPUT

### Report

What to Explain When we Cannot Easily Explain?—An Interdisciplinary Primer on XAI and Meaningful Information in Automated Decision-Making

*Hadi Ashgari, Aljoscha Burchardt, Daniela Dicks, Judith Faßbender, Nils Feldhus,*

*Freya Hewett, Vincent Hofmann\*, Matthias C. Kettemann, Wolfgang Schulz, Judith Simon,*

*Jakob Stolberg-Larsen, Theresa Züger*

🌐 <https://www.hiig.de/en/project/the-ethics-of-digitalisation/>

## FELLOW IMPRESSIONS

“The XAI clinic was a unique research experience; three days in a beautiful secluded area, a dozen interdisciplinary researchers, one topic. The presentations by all the groups and the practitioners were quite good, leading to lively discussions. I personally learned a lot about XAI. And as a collective, we wrote 20+ pages, the basis for our report. A big thank you to the organizers and participants.”

Hadi Asghari

“It was valuable to have an interdisciplinary forum where assumptions about how the same concepts are seen from other perspectives could be tested, adapted accordingly, and integrated into your work again. It was a welcome realization for me that interdisciplinarity lends itself especially to intensive formats, like the clinic, and is more like a rewarding hike than a sprint on the running track.”

Judith Faßbender

“The Explainable AI Clinic organized by the HIIG was a delightful and fruitful experience. We had plenty of opportunities and a great setting near Bad Belzig to exchange our thoughts on Explainable AI, coming from very different backgrounds. This enabled us to see connections between law, design, and software engineering that existing publications have not fully covered yet.”

Nils Feldhus





CLINIC

# Cities, digitalisation, and ethics

DIGITAL ASIA HUB

OCTOBER 2021

## **CITIES, DIGITALISATION, AND ETHICS**

Cities around the world are on a mission to get smart, building ‘city brains’ made up of networked operating systems collecting streams of data from sensors and smart-phones for automated and ‘intelligent’ decision making. Governments and technology firms weave narratives of efficiency, convenience, innovation, and sustainability, in pushing their visions of the future of cities. This phenomenon is especially prominent in Asia, home to many of the world’s largest urban centers. India plans to build 100 smart cities, Singapore is building a ‘Smart Nation’, Seoul has been building itself into a smart city since the early 2000s, and over 800 Chinese cities and towns have introduced Smart City pilots.

This research clinic examined the ethics of digitization, with a focus on cities as key sites of enquiry. Participants were encouraged to bring a critical lens to challenges related to the proliferation of digital and networked technologies, exploring the multiple imaginaries and narratives behind the so-called “Smart City” and discovering entry points to embed ethical principles by (re)design. With participants from many places across the globe, the clinic is inspired by a spirit of interdisciplinary collaboration, mutual learning, and open exchange.

Students had the opportunity to use their skills and diverse academic backgrounds to inform perspectives on building and designing networked cities, while developing vital translational skills for different audiences and interdisciplinary problem-solving. As part of the final outputs of the clinic, the cohort participated in a scenario, tasked to design ‘city bid’ books to build digital cities of the future. Divided into two working groups, the cohort selected Bangkok, Thailand, and Tarawa, Republic of Kiribati, as their sites to articulate a more ethical and inclusive digital city. The cohort selected these two regions because they represent key ethical challenges facing urban planners, such as climate change and rising sea levels, equitable access, rising inequality and digital divide, and personal privacy. They worked alongside a number of experts and mentors to think through key issues, and presented their city bids to a panel of senior experts at the end of the clinic.

Working group A looked at the city of Bangkok, and the areas where interventions can be made to shape Bangkok into an “ethical city” by the year 2030. The inquiry began by examining what is meant by “ethical”, inspired by such principles as the United Nations Global Compact, the New Urban Agenda (UN-Habitat III initiative), and the Sustainable Development Goals (SDGs). Through an examination of various frameworks currently in place in the city, the group highlighted specific areas of concern, how current frameworks already in place can be improved upon, and other recommendations and considerations to help develop new frameworks for the city to become an “ethical city” in 2030.

Working group B focused on South Tarawa in Kiribati as a case study for the ethics of digitalisation in the city. Its unique position as an urban settlement differentiates it from traditional understandings of “the city.” Not only does South Tarawa highlight an area of the world and a population that is often excluded from discussions around global connectivity and the digital economy, it also provides a space to reimagine what technological systems can be built and what new models of governance can grow out of it. Rather than a traditional report style output, the group chose to design the project as a play on tarot and card games, choosing to highlight how issues of participation and future-thinking are linked. The future of a city exists in a collective imagination, but who is in that collective, and who gets to shape it? These are some of the questions that helped the group realize that the most important ethical questions in discussions around digitalisation are those regarding autonomy and the right to self-determination. The question of a game raises the question of players. The tone of this project changes dramatically based on who is playing; for external consultants with no experience on the ground to be “playing” such a “game” feels like “helicoptering” or gamifying the incredibly consequential issues of digital development. Yet these cards could also be given to the people of Kiribati, transforming the game into an important participatory design tool.

Both these outputs are starting points towards the incubation of more interdisciplinary approaches to critiquing Smart City projects and urbanization in Asia, and imagining new approaches that bring ethics embedded in local, non-western values and contexts.

**FELLOWS**

**Aashiyana Adhikari**, Asian Institute of Technology, Thailand

**Line Algoed**, Vrije Universiteit Brussels, Belgium

**Ayça Atabey**, Edinburgh University, UK

**Dicky Danny Willem Bettay**, Australian National University's School of Cybernetics, Australia

**Wei Quan Chua**, Lee Kuan Yew School of Public Policy, Singapore

**Haohan Hu**, University of Hong Kong, Hong Kong SAR

**Elisabeth Huth**, Columbia University, USA

**Bruno Idini**, Sciences Po, Paris, France

**Rhea Hua Jiang**, Harvard University, USA

**Zeynep Ülkü Kahveci**, Istanbul Bilgi University, Turkey

**Abhibhu Kitikamdhorn**, Chulalongkorn University, Thailand

**James Liu**, Australian National University's School of Cybernetics, Australia

**Kuansong Zhuang**, University of Illinois at Chicago, USA

## OUTPUTS

### **Clinic Report and Reflections**

Bangkok: An Ethical City in 2030

*Abhibhu Kitikamdhorn, Wei Quan Chua, Danny Willem Bettay, Aashiyana Adhikari, Elisabeth Huth, Ayça Atabey*

[https://www.hiig.de/wp-content/uploads/2021/11/DAH\\_clinic\\_Cities\\_Digitalization\\_Ethics\\_Group\\_A\\_Bangkok\\_An\\_Ethical\\_City\\_in\\_2030.pdf](https://www.hiig.de/wp-content/uploads/2021/11/DAH_clinic_Cities_Digitalization_Ethics_Group_A_Bangkok_An_Ethical_City_in_2030.pdf)

### **Report, Game, and Reflections**

The Ethics of Digitization in the South Pacific—Projected Futures in South Tarawa, Kiribati

*Bruno Idini, Haohan Hu, James Liu, Kuansong Zhuang, Line Algoed, Rhea Hua Jiang, Zeynep Kahveci*

[https://www.hiig.de/wp-content/uploads/2021/11/DAH\\_clinic\\_Cities\\_Digitalization\\_Ethics\\_Group\\_B\\_Final\\_Presentation.pdf](https://www.hiig.de/wp-content/uploads/2021/11/DAH_clinic_Cities_Digitalization_Ethics_Group_B_Final_Presentation.pdf)

## FELLOW IMPRESSIONS

“For me, there were two. First, the exposure to the diverse nature of interdisciplinary work. Second, the professional and collegiate network that was built during the clinic, which has continued to provide me with various avenues to assist with my own research, future collaboration opportunities, and personal friendships across geographical boundaries that I believe would not have otherwise been possible.”

Danny Bettay

“The people were the best experience of this Research Clinic. We had such a diverse group of people—not just in terms of background but also perspectives—which made for knowledge-rich, empathetic discourse on ethical urban and digital development. Our sharing of personal stories also contextualized heavy discussions, e.g., on the history of traffic lights, while developing camaraderie among the cohort.”

Ryan Chua

“The highlight of the sprint was the final day when we could see the output of 4 intense weeks of learning, planning, and creating. During our team meetings, I received feedback from teammates with different backgrounds. My biggest take-away was that laws that look perfect on paper might be impracticable, and an ethical city should be supported by laws that address the needs of all residents.”

Zeynep Kahveci



MULTI STAKEHOLDER DIALOGUE

# Strategies for an ethics of digitalisation

ALEXANDER VON HUMBOLDT INSTITUTE  
FOR INTERNET AND SOCIETY

OCTOBER 2021

## STRATEGIES FOR AN ETHICS OF DIGITALISATION

### FRAMING DIGITALISATION

The internet is a vast operational domain. States, individuals, and platforms pursue independent goals based on national or self-interest. Creating value and ensuring sustainable civil societies does not possess the same meaning for each of these stakeholders. Ensuring that ethics play an important role in digitalisation has been a key focal point of the international “Ethics of Digitalisation” project initiated by the Network of Centers (NoC). At the stakeholder dialogue, we went over the progress in the field of ethics of digitalisation, further brainstormed how to turn our visions into practice, and discussed how digitalisation can be made to work for everyone. We were eager to share some of the insights and innovative research methods applied within the project with the German platform governance stakeholder community and therefore invited around thirty representatives and experts from science, civil society, politics, and public administration fields to the HIIG. The aim of this exchange was to take stock of progress in the field of ethics of digitalisation and define goals for an ethical digitalisation, along with the necessary requirements to reach them.

### VISIONS FOR AN “ETHICS OF DIGITALISATION”

Where do we picture ourselves in 2040, after twenty more years of digitalisation? During the first session, participants shared their visions for an ethics of digitalisation. One particular vision that all could agree on: human rights are key! An ethical digitalisation should put human rights at the core of the entire process. This vision pairs well with that of a digital world which serves the public interest, and aims to eliminate technology-fueled polarization. However, this requires an understanding of what is discussed when it comes to ethics of digitalisation. We are currently experiencing a worldwide boom on artificial intelligence (AI) guidelines, which are attempting to define and dictate the boundaries of AI. Although these guidelines tend to frequently use the same terms, there is an absence of a common understanding of complex and foundational terms like “AI”. Therefore, one of the visions for how we picture a digitalized world in twenty years, serves as a basis for a common understanding of the key terms of digitalisation. The current issue we are facing when it comes to defining these terms could stem from a separate issue: the goal of digitalisation remains unclear. Comparing it to the climate crisis, the primary goal of all activities is to promote environmental protections and habits, in order to maintain a livable and healthy atmosphere on earth for all of its inhabitants. This common understanding of a primary goal is hard to find in the discourse on an ethics-driven digitalisation.



## CONDITIONS FOR SUCCESS

Ethical problems of digitalisation are often only discussed within a small bubble. Although prominently placed in most of the parties' policy agendas, digitalisation was a rarely discussed topic during the German federal election campaigns in 2021. In order to change this pattern of avoiding important and necessary discussions on digitalisation, primary stakeholders like members of the scientific community and civil societies could be tasked with combating this challenge. The scientific community must deliver the foundation for a technical, legal, and communicative perspective on questions of human-machine interaction, and define the terms used in the context of digitalisation. When it came to identifying important questions to ask, one idea received highly positive feedback: approaching digitalisation from a utopian or dystopian perspective. This was the approach with the HIIG project "twentyforty", where participating researchers published stories about what could potentially occur up until 2040 with respect to digitalisation. Apart from allowing the researchers to engage with the project in a creative way, the method proved helpful in allowing us to consider what we should focus on, in order to apply the positive scenarios in the story to the real world. A different, yet similar method was also used to inspire participants to change the way they think. Rather than only focusing on positive scenarios and what a "human-centered digitalisation" could entail, participants were asked to discuss the potential characteristics and outcomes of a "machine-centered digitalisation".

Civil society was also identified as a primary stakeholder, which could play an important role in defining an ethical digitalisation. Civil society can contribute to the cause by sharing results acquired from scientific research with the general public in a language that everyone can understand, while also addressing issues that other stakeholders may notice, in an effort to advance the broader discussion on an ethical digitalisation.

## WHAT'S NEXT?

The concluding session focused on the necessary steps to implement the visions of ethics in digitalisation. The ideas presented covered all stakeholders of digitalisation. Regulators should introduce a mandatory ethical assessment for all products related to digitalisation. Research formats like twentyforty should be widely introduced, in order to advance a common understanding of key terms in digitalisation. Civil societies should introduce formats such as "Digital Speed-Dating", where individuals could meet with experts on digitalisation to help advance the public discourse. Increased cooperation between different organizations like [digitalezivilgesellschaft.org](https://www.digitalezivilgesellschaft.org) could help enrich their work. Ministries should be held accountable to maintain consistent communication with representatives from civil societies. However, this accountability should not only apply to ministries. All participants agreed that a diverse group of stakeholders should come together more often to discuss the issues of digitalisation from different perspectives.

## PARTICIPANTS

**Viktoria Aygül**, The European Centre for Minority Issues, Germany  
**Dr. Thomas Bagger**, Office of the Federal President, Germany  
**Prof. Dr. Christoph Bieber**, Center for Advanced Internet Studies, Germany  
**Lajla Fetic**, Bertelsmann Stiftung, Germany  
**Dr. Frauke Gerlach**, Grimme-Institut, Germany  
**Dr. Isabella Hermann**, Independent Expert on Science Fiction, Germany  
**Vincent Hofmann**, Alexander von Humboldt Institute for Internet and Society, Germany  
**Carla Hustedt**, Stiftung Mercator, Germany  
**Adrian Kopps**, Alexander von Humboldt Institute for Internet and Society, Germany  
**Vérane Meyer**, Heinrich-Böll-Stiftung, Germany  
**Dr. Markus Oermann**, Office of the Federal President, Germany  
**Ann Cathrin Riedel**, LOAD e.V., Germany  
**Johannes Röder**, RWTH Aachen University, Germany  
**Victoria Guijarro Santos**, WWU University of Münster, Germany  
**Jan Schallaböck**, iRights.Lab, Germany  
**Francesca Schmidt**, Federal Agency for Civic Education, Germany  
**Thilo Scholle**, Federal Ministry of Labor and Social Affairs, Germany  
**Prof. Dr. Wolfgang Schulz**, Alexander von Humboldt Institute for Internet and Society, Germany  
**Marie-Therese Sekwenz**, Institute for Information Systems and Society, Austria  
**Zora Siebert**, Heinrich-Böll-Stiftung, Germany  
**Fabian Stein**, Initiative D21, Germany  
**Dr. Thorsten Thiel**, Berlin Social Science Center (WZB), Germany  
**Dr. Kyriaki Topidi**, The European Centre for Minority Issues, Germany  
**Raphael von Aulock**, Alfred Landecker Foundation, Germany  
**Tobias Wangermann**, Konrad-Adenauer-Stiftung, Germany

## PARTICIPANT IMPRESSIONS

“For me as a PhD student it was especially exciting to experience the open and equal atmosphere at the event. Everyone could speak their mind and bring in their experience without any hierarchy. For example, in my working group we came up with an idea that inspired me: If you aim for a certain goal, think about the opposite of that goal and what you need to do to reach this counter goal.”

Vincent Hofmann

“For me, the Stakeholder Dialogue represented a unique opportunity for cross-sectoral exchange. It was very insightful to hear the different perspectives from government, academia and civil society about how to best approach the ethical challenges of digitalisation and to experience the mutual curiosity. I left inspired and enriched by the great discussions that took place and one input that stood out to me was the presentation from the “twentyforty—Utopias for a digital society“ project.”

Adrian Kopps

“For me, the stakeholder dialogue was shaped above all by many individual talks and group discussions with talented people, all of whom have different perspectives on digital ethics. Together, it was possible to develop concrete demands for a human rights-centered and participatory digitalisation.”

Johannes Röder




## IMPRINT

Report of the international research project “The Ethics of Digitalisation: From Principles to Practices”, a joint initiative of the Global Network of Internet and Society Research Centers (NoC).

### PUBLICATION

June 2022

### PUBLISHER

Alexander von Humboldt Institute for Internet and Society gGmbH  
Französische Str. 9  
10117 Berlin  
 [www.hiig.de](http://www.hiig.de)

### RESEARCH LEADS

Wolfgang Schulz  
Sandra Cortesi  
Urs Gasser  
Malavika Jayaram  
Matthias C. Kettemann  
Vincent Hofmann  
Alexander Pirang  
Padmashree Gehl Sampath  
Fiona Tregenna  
Dev Lewis

### PROJECT COORDINATION

Nadine Birner  
Hanna-Sophie Bollmann  
Friederike Busch

### LAYOUT

Martina Kogler  
Larissa Wunderlich

### EDITING

Translabor Berlin

### PICTURE CREDITS

istockphoto.com | 3DStock, da-kuk, adaask





